An Unbalanced Optimal Transport Problem with a Growth Constraint

by

Yuqing Dai

Department of Mathematics
Duke University

Defense Date: April 2, 2024

Approved:

James Nolen, Supervisor

Thomas P Witelski

Jian-Guo Liu

Jianfeng Lu

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Mathematics
in the Graduate School of Duke University
2024

ABSTRACT

An Unbalanced Optimal Transport Problem with a Growth Constraint

by

Yuqing Dai

Department of Mathematics
Duke University

Defense Date: April 2, 2024

Approved:

James Nolen, Supervisor

Thomas P Witelski

Jian-Guo Liu

Jianfeng Lu

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Mathematics
in the Graduate School of Duke University
2024

Abstract

In this paper, we introduce several unbalanced optimal transport problems between two Radon measures with different total masses. Initially, we explore a generalization of the Benamou-Brenier problem, incorporating a growth constraint to accommodate a non-decreasing total mass during transportation. This leads to the formulation of a modified Hellinger-Kantorovich (HK) problem, denoted as $\mathbf{mHK}$. Our investigation reveals quasi-metric properties of this novel problem and characterizes it within a cone setting through a newly defined quasi-cone metric, resulting in an equivalent formulation denoted by $\mathbf{mHK}_{\mathfrak{C}}$. This formulation simplifies the demonstration of the existence of optimal solutions and facilitates explicit calculations for transport problems between two Dirac measures.

A significant advancement in our work is the construction of a dual problem for $\mathbf{mHK}_{\mathfrak{C}}$, a topic previously unexplored. We confirm the duality and identify optimality conditions for transport plans, successfully deriving a one-to-one (Monge) map under certain regularity conditions for the initial measure. Furthermore, we propose a dynamic formulation for $\mathbf{mHK}_{\mathfrak{C}}$, focusing on minimization over dynamic plans involving absolutely continuous curves between cone points. This approach not only projects a dynamic plan onto an absolutely continuous curve between initial and target measures but also establishes a close relationship with solutions to continuity equations.

Motivated by dynamic models of biological growth, our study extends to practical applications, providing an equivalent convex formulation of $\mathbf{mHK}$ and developing numerical schemes based on the Douglas-Rachford algorithm and the Alternating Direction Method of Multipliers algorithm. We apply these schemes to synthetic data, demonstrating the utility of our theoretical findings.

Contents

List of Figures

Acknowledgements

I want to show my deepest gratitude to my advisor, James Nolen, for his exceptional technical and emotional support throughout my research journey. His organizational skills have been paramount, as evidenced by his commitment to regular meetings and ensuring that my work remained on the right track. His patience and willingness to understand my thought process have been especially noteworthy, allowing me to explore my research interests deeply. Moreover, his guidance in steering me toward the correct path and his diligent provision of relevant literature has been indispensable. Completing my thesis would have been an impossible task without his support.

I also sincerely thank the Duke Mathematics Department for its financial support in my research.

1. Introduction

The field of optimal transport theory focuses on determining the most efficient way to relocate resources and minimize the transportation costs associated with moving from one distribution to another. This theory was first introduced by the French mathematician Gaspard Monge in 1781. Monge was interested in devising the most efficient strategy for relocating a pile of sand from a construction site to another location, laying the groundwork for what is known today as the Monge problem. This problem requires that the total mass at the starting and target locations are equivalent. Subsequently, Leonid [Kan39], a Soviet mathematician, economist, and Nobel Prize winner, expanded upon Monge's initial formulation, introducing greater flexibility into the solutions of these transport problems. Unlike the Monge problem, which restricts mass transportation from one specific location to another, Kantorovich's problem permits the mass distribution from a single location to multiple destinations. Furthermore, Kantorovich introduced a dual problem, a maximization problem constrained by the Kantorovich problem. This dual problem facilitates the investigation of the optimality conditions necessary for identifying the optimal solution to Kantorovich's formulation.

A significant advancement in classical optimal transport theory is the development of the Wasserstein Distance within a metric space, essentially applying Kantorovich's problem with a specific cost function. This distance measures the discrepancy between two probability distributions and imbues the resulting space with several metric properties, such as separability, compactness, completeness, and the existence of geodesics. Computationally, the Wasserstein distance finds numerous applications in fields like image matching, computer vision, and shape recognition within image processing.

The landmark study by [BB00] introduced a dynamic formulation characterizing a time-dependent curve linking two probability measures, governed by a continuity equation, and redefined the Wasserstein distance as the minimal kinetic energy required for the transport. Additionally, their work proposed an efficient numerical approach for approximating these optimal solutions. Comprehensive overviews of balanced optimal transport theory have been adeptly provided by [Vil03], [Vil09], and [ABS21], offering deep insights into the subject's complexities and applications.

The theory of unbalanced optimal transport extends the classical framework of transporting resources between two probability distributions to non-negative Radon measures, eliminating the necessity for the total mass to be identical at both the origin and destination. This advancement addresses the challenge of measuring distances between distributions with different total masses. Early efforts to adapt the Wasserstein distance for this broader context and to investigate the metric properties of these novel problem formulations were undertaken in studies by [FG10], [PR12], [PR13]. Concurrently, a variation of the Benamou-Brenier problem was independently introduced by [LMS16], [Chi+18a], and [KMV16]. These researchers not only demonstrated that their modified formulation defines a metric between two non-negative Radon measures but also approached the derivation of this metric from unique perspectives, employing various techniques.

In their groundbreaking work, [KMV16] extended the traditional Benamou-Brenier problem by incorporating a non-conservative continuity equation with an additional reaction term, facilitating changes in mass. This modification conceptualizes the problem as minimizing the Lagrangian action associated with the total energy, encapsulated by the sum of kinetic energy and the potential energy attributed to the growth and decay of the transported mass. According to [KMV16], this newly defined distance carries significant implications, offering physical interpretations for the movement of charged particles as well as biological contexts, particularly in the fitness-driven dispersal of organisms. Echoing the innovative approach of [KMV16], [Chi+18a] similarly applied the continuity equation with a reaction term but introduced the Fisher-Rao Riemannian metric, originally proposed in [[Rao45]], to quantify the impact of the reaction term. In their seminal contribution, [Chi+18a] formulated a metric representing the minimization of interpolation between the Wasserstein and Fisher-Rao metrics. Moreover, they developed a numerical framework for approximating these optimal solutions, further applying their methodology to conduct numerical experiments on images.

[LMS16] constructed the generalized Benamou-Brenier problem as the infimum of the weighted combination of the Kantorovich-Wasserstein distance and Hellinger distance. In this context, the Wasserstein distance quantifies the displacement of the distribution, whereas the Hellinger distance captures the creation or destruction of mass within the distribution. To bridge the gap between this dynamic challenge and a static framework, [LMS16] alongside [LMS18] introduced the Hellinger-Kantorovich distance. This innovative approach involves lifting the measures to a cone space, thereby framing it as a minimal balanced transport problem for measures residing within this augmented space. Furthermore, an alternative interpretation of the Hellinger-Kantorovich distance emerges through the Entropy-Transport problem. This formulation is characterized by the objective of minimizing a balanced transfer problem, which is complemented by penalty functionals. This development not only broadens the theoretical landscape of unbalanced optimal transport but also enhances the toolkit available for addressing problems where mass changes are an intrinsic part of the system's dynamics.

With an extensive background in optimal transport problems and their applications in biological growth, such as the study of non-reversible tumor growth outlined in [[MC20]], my Ph.D. thesis advisor, Professor James Nolen, and I are motivated to explore a novel problem within this domain. Our project aims to contribute to the ongoing research by introducing a growth constraint to the generalized Benamou-Brenier problem, denoted by $\mathbf{nHK}$. This modification imposes a condition where the total mass of the transported distribution is required to be non-decreasing, a constraint that mirrors the growth processes observed in biological systems.

The primary objective of our research is to rigorously establish the existence and uniqueness of optimal solutions for this newly formulated problem. To achieve that, we draw inspiration from the methodology employed by [LMS18] for addressing the Hellinger-Kantorovich problem. Our approach involves developing a new quasi-distance, defined within a cone setting and denoted by $\mathbf{nHK}_{\mathfrak{C}}$, associated with a quasi-cone metric. This new metric is designed to simplify the process of establishing minimization criteria. By demonstrating the equivalence between the $\mathbf{mHK}$ and $\mathbf{mHK}_{\mathfrak{C}}$ problems, we can confirm the existence of optimal solutions for $\mathbf{mHK}$. Moreover, leveraging our innovative approach to the dual problem, coupled with certain regularity conditions on the initial measure, we successfully identify a unique solution to the problem.

Our achievements aim to lay a solid mathematical foundation that can support further exploration and application of optimal transport theory in the context of biological growth. Furthermore, we are dedicated to developing efficient numerical schemes that can facilitate the practical application of our findings. These computational tools are crucial for translating theoretical insights into actionable knowledge that can be applied to real-world problems, such as modeling tumor growth or other biological phenomena where growth is a key factor.

Our aspiration is that this work will advance the mathematical understanding of optimal transport problems with growth constraints and open new pathways for applying these concepts to the biological field. The potential for contributing meaningful solutions to pressing challenges in biology and medicine provides a significant motivation for our research, highlighting the interdisciplinary impact that mathematical innovations can have on understanding complex biological processes.

1.1 Illustrative Examples

To elucidate the nuances of optimal transport problems, we will explore some illustrative examples. These examples will help differentiate the Monge problem, the Kantorovich problem, and the dual problem within a practical context. Consider farmers who aim to minimize transportation costs when selling apples from their farms to grocery stores, a task that involves deciding which stores to supply and in what quantities. In the scenario depicted in Figure 1.1, the farmers transport apples directly from one specific barn to a single grocery store.

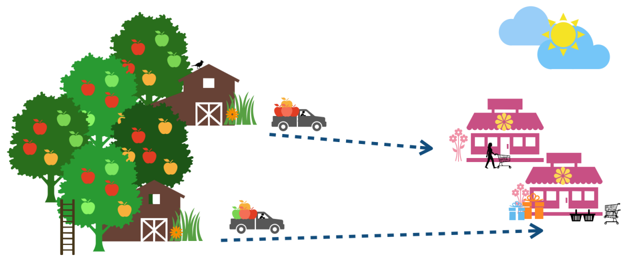


FIGURE 1.1: Monge Problem

Figure 1.2 expands on this by allowing farmers to distribute the apple load from one barn to several stores, offering a more flexible approach to transportation.

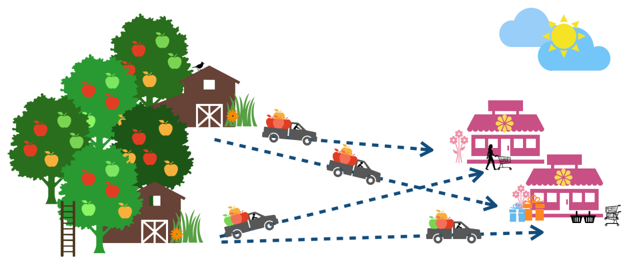


FIGURE 1.2: Kantorovich Problem

Lastly, Figure 1.3 introduces a third party: agents who collect apples from various barns to sell to grocery stores. These agents strive to minimize transportation costs to remain the preferred choice for farmers while simultaneously seeking to maximize their profits through efficient distribution strategies.

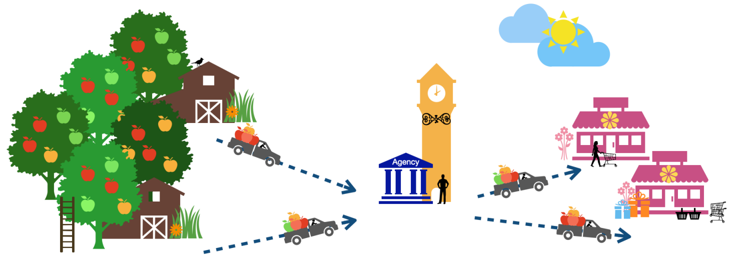


FIGURE 1.3: Dual Problem

Expanding on the apple transportation scenario, Figure 1.4 illustrates the distinction between the Kantorovich problem and the Entropy-Transport problem. In the context of balanced optimal transport, the scenario depicts a farmer delivering all 10 apples from his barn directly to the grocery store. Conversely, in the scenario of unbalanced optimal transport, the farmer opts to transport only 6 apples, incurring a penalty for the 4 apples that are not transported.

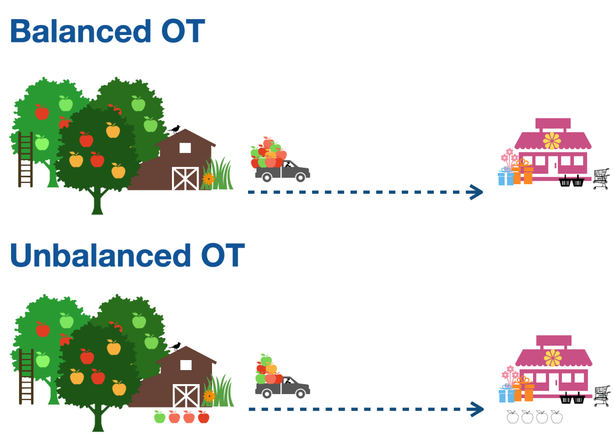


FIGURE 1.4: Comparison Between Static Formulations

Figure 1.5 illustrates the distinctions between the original Benamou-Brenier problem and its generalized version. Typically, running involves only movement without a change in mass. To optimize energy usage, an individual must modulate their running pace. However, imagine a scenario akin to Ant-Man from the Marvel comics, where a person possesses the ability to alter their physical size. In such circumstances, it could be advantageous for the individual to reduce their size for long-distance running. Throughout this transformation, the individual must achieve equilibrium between kinetic energy and the potential energy generated through the mass's creation or annihilation.

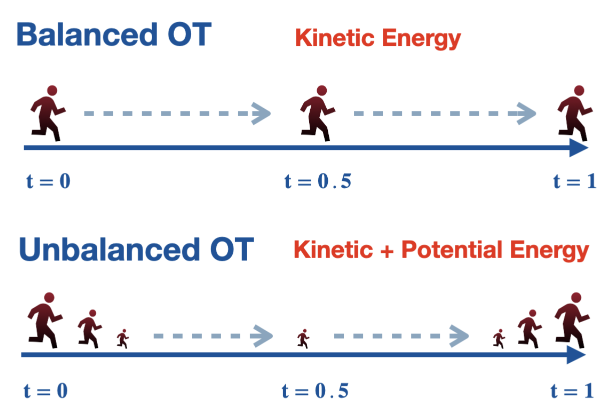


FIGURE 1.5: Benamou-Brenier Problems

Building on the example of a running man, Figure 1.6 illustrates the impact of the growth constraint on his movement. To minimize energy expenditure, the individual attempts to decrease his body size and then expand until reaching the final state. However, the growth constraint prevents the man from reducing his weight; as a result, he maintains his current state for a period before expanding to the final state.

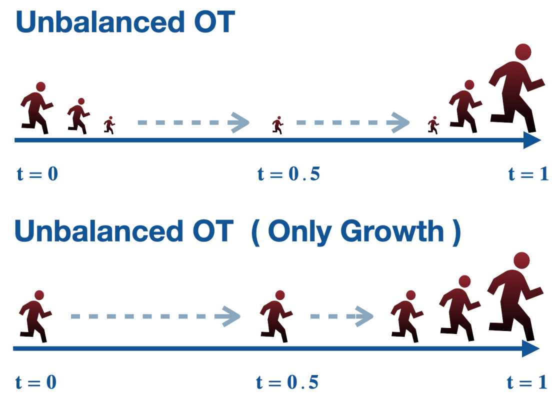


FIGURE 1.6: Constrained Problems

1.2 Organization of the Paper

This paper is structured as follows: Chapters 2 and 3 provide a comprehensive overview of the optimal transport theory pertinent to our study, while the subsequent chapters delve into our contributions. Chapter 2 is dedicated to a concise discussion of the fundamental problems in balanced optimal transport theory, including the Monge problem, Kantorovich problem, Dual problem, and a specialized form of the Kantorovich problem known as the Wasserstein Distance. This chapter also covers their characterizations, along with discussions on the existence and uniqueness of solutions. Chapter 3 shifts focus to the major problems identified in previous works [LMS16], [LMS18], such as the optimal transport problem, the Hellinger-Kantorovich problem, and the generalized Brenier-Benamou problem, setting the stage for our research.

Our novel contributions are presented starting from Chapter 4, where we introduce a modified version of the Hellinger-Kantorovich problem (referred to as the generalized Benamou-Brenier problem with a growth constraint) and its equivalent convex formulation (Benamou-Brenier functional). Chapter 5 elaborates on a variant of the modified Hellinger-Kantorovich problem within a cone setting. This innovative characterization enables us to verify the existence of an optimal solution for the modified problem. Additionally, we present a newly developed dual problem for this variant, a novel contribution not previously explored in existing literature.

In Chapter 6, we outline numerical schemes designed for approximating the solutions to these complex problems, further advancing the field of optimal transport theory. Figure 1.7, Figure 1.8, and Figure 1.9 provide an overview of the problems we will cover in this paper, where the arrow represents the relationship between the problems.

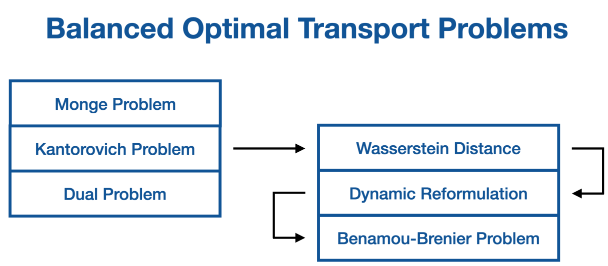


FIGURE 1.7: The Structure of Chapter 2

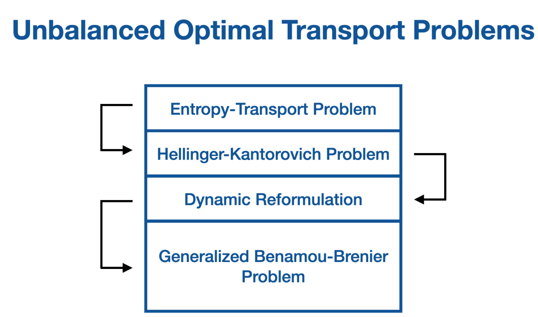


FIGURE 1.8: The Structure of Chapter 3

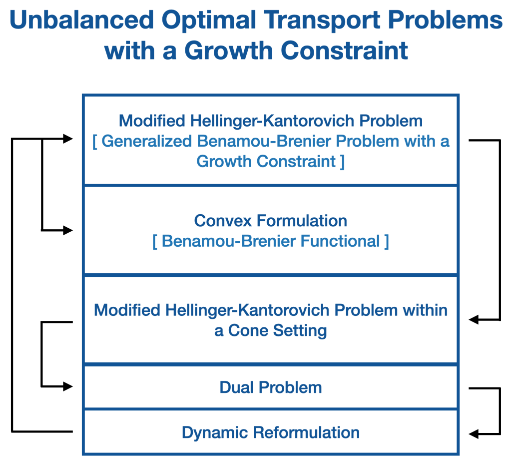


FIGURE 1.9: The Structure of Our Work

1.3 Preliminaries

Basic settings Let (Z, τ_Z) be a Hausdorff topological space and $\mathcal{B}(Z)$ be the σ-algebra of Borel sets in Z. The set of finite nonnegative Radon measures on Z is

$$\mathcal{M}(Z) := \Big\{ \sigma\text{-additive set functions } \mu : \mathcal{B}(Z) \to [0, \infty) : \\ \forall B \in \mathcal{B}(Z), \forall \varepsilon > 0, \exists K_\varepsilon \subset B \text{ compact s.t. } \mu(B \backslash K_\varepsilon) \leqslant \varepsilon \Big\} \tag{1.1}$$

We endow $\mathcal{M}(Z)$ with the narrow topology s.t. $\mu \mapsto \int_Z \varphi \, d\mu$ is lower semi-continuous for any lower semi-continuous and bounded function φ, i.e. $\varphi \in \mathrm{LSC}_b(Z)$. When (Z, τ_Z) is a Polish space, the narrow topology coincides with the weak topology induced by the duality with functions in $C_b(Z)$, i.e. $\mu_n \to \mu \in \mathcal{M}(Z)$ weakly if

$$\int_Z \varphi(z) \, d\mu_n(z) \to \int_Z \varphi(z) \, d\mu(z) \text{ for any } \varphi \in C_b(Z)$$

Notations

$\mathbb{R}_+$	non-negative real number
$\mathcal{B}(X)$	σ-algebra of Borel sets of X

$\mathcal{M}(X)$	finite non-negative Radon measures on X
$\mathcal{P}(X), \mathcal{P}_p(X)$	the set of probability measures on X (with a finite p-moment)
$C_b(X)$	continuous and bounded real functions on X
$C_c^\infty(X)$	smooth functions on X with compact support.
$\mathrm{LSC}(X), \mathrm{LSC}_b(X)$	(bounded) lower semicontinuous functions on X
$\mathrm{Lip}(X), \mathrm{Lip}_b(X)$	(bounded) Lipschitz functions on X
$C(I; X)$	continuous curves on I
$C(I; \mathcal{M}(X))$	continuous non-negative and finite Radon measures on I
$L^p(X;\mu), L^2(\mu)$	L^p space of μ-measurable functions on X (or $\mathbb{R}^d$)
$L^p(I)$	L^p space of real functions in I
$Geo(X)$	geodesics in X
$\mathrm{Lip}(\phi, A)$	Lipschitz constant of the function ϕ in the set A
$\mathrm{AC}^p(I;(X,d))$	absolutely continuous curves on I within a metric space (X,d)
$\lvert x'\rvert_d$	metric derivative of an absolutely continuous curve x within a metric space (X,d)
$\mathcal{Y}$	product space $X \times \mathbb{R}_+$
$\tilde{C}(I;\mathcal{Y}), \widetilde{AC}^p(I;\mathcal{Y})$	(absolutely) continuous curve $(x,r) \in \mathcal{Y}$ on I, see (3.27), (3.28)
$\mathfrak{C}, \mathfrak{C}^{\otimes N}, \mathfrak{o}$	(the product of) cone space and its vertex
$\mathfrak{C}[r]$	Neighborhood of radius r centered at $\mathfrak{o}$
$\mathfrak{h}_i^2, \mathrm{dil}_{\vartheta,2}$	ith homogeneous marginals and dilations, see (3.13, 3.14)
$\mathfrak{h}_t^2$	characterization of homogeneous marginals, see Section 3.4
$d_{\mathfrak{C}}, \tilde{d}_{\mathfrak{C}}$	(modified quasi-) cone distance
W_p	Wasserstein distance in $\mathcal{P}_p(X)$, see Section 2.4
ET, **LET**	(Logarithmetric) Entropy-Transport problem, see (ET)
$W_{d_{\mathfrak{C}}}$	Wasserstein distance in a Radon space, see (3.9)
HK	Hellinger-Kantorovich (HK) problem, see Subsection 3.2.3
mHK	modified HK problem, see (4.2)
$\mathbf{mHK}_{\mathfrak{C}}$	modified HK problem within a cone setting (5.25)
$W_{\tilde{d}_{\mathfrak{C}}}$	Wasserstein problem associated with quasi-cone metric $\tilde{d}_{\mathfrak{C}}$, see (5.29)
$\mathcal{B}_\kappa$	Benamou-Brenier functional, see (4.18)
$\mathcal{D}'((0,T)\times\mathbb{R}^d)$	distributional sense defined in (1.15)
$\mathcal{CE}^+, \widetilde{\mathcal{CE}}$	continuity constraints, see (4.3),(4.20)

Other notations

- $\mu \ll \mathcal{L}^d$, the measure μ is absolutely continuous w.r.t Lebsegue measure in $\mathbb{R}^d$
- $\iota_C(x) = \infty$ if $x \in C$ otherwise 0.
- δ_x is a Dirac measure with mass 1 at location x
- $spt(\mu)$ is the intersection of all closed sets that have full measure
- μ is concentrated on A if $\mu(A^c) = 0$ where A^c is the complement of A.

- (μ_t, v_t, w_t) is the abbreviation of $(\mu(t,\cdot), v(t,\cdot), w(t,\cdot))$

1.3.1 Definitions

Definition 1.2 (Polish Space, [ABS21]). A topological space (X,τ) is Polish if there exists a distance d on X inducing τ such that (X,d) is complete and separable.

Definition 1.3 ([Oxt70]). Let (X,μ) be a topological measure space. μ is non-atomic if $\mu(\{x\}) = 0$ for each $x \in X$.

Definition 1.4 (Push Forward Measure, [ABS21, p.4]). $f_\#\mu$ is the push forward measure, of a Borel function $f : Z \to Y$ on the measure $\mu \in \mathcal{M}(Z)$ such that

$$f_\#\mu(B) := \mu(f^{-1}(B)), \qquad \forall B \in \mathcal{B}(Y). \tag{1.5}$$

Proposition 1 (Change of Variable Formula, [ABS21, Proposition 1.7]). For any Borel function $f : X \to Y$ and any Borel function $\phi : Y \to [0,\infty]$, one has

$$\int_Y \phi \, d(f_\#\mu) = \int_X (\phi \circ f) \, d\mu. \tag{1.6}$$

Proposition 2 ([ABS21, Proposition 1.8]). For any Borel function $T : X \to Y$, one has $T_\#\mu = \nu$ if and only if

$$\int_Y \phi \, d\nu = \int_X (\phi \circ T) \, d\mu, \qquad \forall \phi \in C_b(Y). \tag{1.7}$$

1.3.2 Absolute Continuity

The foundational concepts of absolutely continuous curves are thoroughly discussed in the literature, particularly in Chapter 1 of the book by [AGS05, Chapter 1] and Section 8.1 of the seminar paper by [LMS18, Section 8.1]. In this subsection, we summarize the key definitions as presented in these works.

Let (Z, d_Z) be a metric space and I be a time interval in $\mathbb{R}$. Then $\gamma : I \to Z$ is an absolutely continuous curve, i.e. $\gamma \in AC\big(I;(Z,d_Z)\big)$, if there exists $g \in L^1(I)$ such that

$$d_Z(\gamma(t_0), \gamma(t_1)) \leqslant \int_{t_0}^{t_1} g(t)\, dt, \quad \forall t_0, t_1 \in I,\ t_0 < t_1. \tag{1.8}$$

If $g \in L^p(I)$ for some $p \in (1, \infty]$, then $\gamma \in AC^p\big(I; (Z, d_Z)\big)$. Its metric derivative $|\gamma'|_{d_Z}$ is the Borel function defined by

$$|\gamma'|_{d_Z}(t) := \limsup_{h \to 0} \frac{d_Z(\gamma(t+h), \gamma(t))}{|h|} \tag{1.9}$$

and by [AGS05, Theorem 1.1.2], the lim sup is a limit for $\mathcal{L}^1$-a.e. points in I.

If Z is complete and separable, then $AC^p\big(I; (Z, d_Z)\big)$ is a Borel set in the space $C(I; Z)$ endowed with the topology of uniform convergence.

1.3.3 Length Space and Geodesic Space

This subsection is a summary from the book [BBI01, Chapter 2]. Let X be a topological space and A be a class of admissible paths, which is a subset of all continuous paths in X, such that A is closed under

1. restrictions: if $\gamma : [a, b] \to X$ is an admissible path, so is $\gamma\big|_{[c,d]}$ for any $a \leqslant c \leqslant d \leqslant b$.
2. concatenations: if $\gamma|_{[a,c]}$ and $\gamma|_{[c,b]}$ are admissible paths, so is $\gamma_{[a,b]}$.
3. linear reparameterizations: given a homeomorphism $\varphi : [c, d] \to [a, b]$ of the form $\varphi(t) = \alpha t + \beta$, if $\gamma : [a, b] \to X$ is an admissible path, so is the composition $\gamma \circ \varphi(t)$.

Then a length structure on X is the class A, together with the length of path $L : A \to \mathbb{R}_+ \cup \{\infty\}$. Let the space X be associated with this length structure, we can define a metric d_L between two points $x, y \in Z$ to be the infimum of lengths of admissible paths connecting them:

$$d_L(x, y) = \inf \Big\{ L(\gamma) \,\big|\, \gamma : [a, b] \to X, \gamma \in A, \gamma(a) = x, \gamma(b) = y \Big\}. \tag{1.10}$$

Then (X, d_L) defines a length space.

Theorem 3 ([BBI01, Theorem 2.7.6])**.** Let (X, d) be a metric space and $\gamma : [a, b] \to X$ be a Lipschitz curve. Then the length of the path is

$$L(\gamma) = \int_a^b |\gamma'|_d(t)\, dt. \tag{1.11}$$

Continuing with [AGS05, Section 2.5], a curve $\gamma : [a, b] \to X$ is a shortest path if and only if $L(\gamma) = d(\gamma(a), \gamma(b))$. If X is a length space, a curve $\gamma : I \to X$ is a geodesic if for every $t \in I$, there exists an interval J containing a neighborhood of t in I such that $\gamma|_J$ is the shortest path, i.e. for every $[t_0, t_1] \subset [a, b] \subset I$,

$$d_Z(\gamma(t_0), \gamma(t_1)) = |t_1 - t_0| \cdot d_Z(\gamma(a), \gamma(b)). \qquad \text{(constant speed)}$$

Moreover, a curve $\gamma \in AC^2(I; (X, d_X))$ is a geodesic if and only if

$$\int_0^1 |\gamma'|^2_{d_Z}(t)\, dt \leqslant d_Z^2(\gamma(0), \gamma(1)). \tag{1.12}$$

We denote by $Geo(X) \subset C(I; X)$ the closed subset of all the geodesics.

Definition 1.13 (Geodesic Metric Space, [ABS21, Definition 9.12])**.** A metric space (X, d) is geodesic if for all $x, y \in X$, there exists $\gamma \in Geo(X)$ with $\gamma(0) = x$ and $\gamma(1) = y$.

In this context, the geodesic space is considered complete if every pair of points at a finite distance can be connected by a geodesic.

1.3.4 Continuity Equations

Fix $T > 0$. Define a triplet $\big(\mu(t, x), v(t, x), w(t, x)\big)$ to represent the mass distribution, velocity, and rate of growth of an object at a given time $t \in [0, T]$ and position $x \in \mathbb{R}^d$, respectively. The continuity equation with a reaction term (CE) is a partial differential equation such that

$$\partial_t \mu(t, x) + \nabla \cdot \big(v(t, x)\mu(t, x)\big) = w(t, x)\mu(t, x) \quad \text{in } (0, T) \times \mathbb{R}^d \tag{1.14}$$

in the distributional sense, denote by $\mathcal{D}'((0, T) \times \mathbb{R}^d)$, i.e., for any test functions $\varphi \in C_c^\infty((0, T) \times \mathbb{R}^d)$,

$$\int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \varphi(t, x) + v(t, x) \cdot \nabla_x \varphi(t, x) + w(t, x)\varphi(t, x)\Big)\, \mu(t, dx)\, dt = 0. \tag{1.15}$$

Informally, $v\mu$ represents the momentum at (t, x), and $w\mu$ describes the mass created or destructed at (t, x). See Appendix A for the proofs of the following statements.

Proposition 4. Let $t \mapsto \mu_t : (0,T) \to \mathcal{M}(\mathbb{R}^d)$ be a weakly continuous curve. If $t \to v_t : (0,T) \to \mathbb{R}^d$ and $t \mapsto w_t : (0,T) \to \mathbb{R}$ satisfy

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|\, d\mu_t(x)dt < \infty \quad \text{and} \quad \int_0^T \int_{\mathbb{R}^d} |w_t(x)|\, d\mu_t(x)dt < \infty, \tag{1.16}$$

then the following facts are equivalent:

1. (μ, v, w) solves (CE) in (1.14).
2. for any $\phi \in C_c^\infty(\mathbb{R}^d)$,

$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi\, d\mu_t = \int_{\mathbb{R}^d} \Big(\nabla_x \phi \cdot v_t + \phi w_t\Big) d\mu_t \tag{1.17}$$

This result is generalized from [ABS21, Proposition 16.3].

Proposition 5 (Existence and Uniqueness of Solutions to the Continuity Equations)**.** These results are generalized from [AGS05, Section 8.1] and [Man17, Section 3]. Let $t \mapsto \mu_t : (0,T) \to \mathcal{M}(\mathbb{R}^d)$ be a weakly continuous curve solving (CE) in (1.14) induced by a Borel vector $(v_t)_{t\in I}$ and a scalar function $(w_t)_{t\in I}$.

(i) If v satisfies

$$\int_0^T \Big(\sup_{\mathbb{R}^d} |v_t| + \mathrm{Lip}(v_t, \mathbb{R}^d)\Big) dt < \infty \tag{1.18}$$

and w be a Borel bounded and locally Lipschitz continuous (with respect to x) scalar function. Then there exists a unique μ_t such that (μ, v, w) solves (1.14) and μ_t is given by the explicit formula, $\mu_t := (X_t)_\#(\lambda_t \mu_0)$, where X_t and λ_t are the unique solutions of

Cauchy problem	Leibniz formula
$\begin{cases} \frac{d}{dt}X_t = v_t(X_t) \\ X_0(x) = x \end{cases}$	$\begin{cases} \frac{d}{dt}\lambda_t = w_t(X_t)\lambda_t \\ \lambda_0(x) = 1 \end{cases}$

and $t \mapsto \mu_t$ is continuous, followed by the continuity of $t \mapsto X_t$

(ii) (Approximation by regular curves) Given $p \geqslant 1$, and suppose that v and w satisfy

$$\int_0^T \int_{\mathbb{R}^d} |v_t(x)|^p\, d\mu_t(x)\, dt < \infty \quad \text{and} \quad \int_0^T \int_{\mathbb{R}^d} |w_t(x)|^p\, d\mu_t(x)\, dt < \infty. \tag{1.19}$$

Let (ρ_ε) be a family of strictly positive mollifiers (e.g. $\rho_\varepsilon(x) = \frac{1}{\sqrt{(2\pi\varepsilon^n)}} e^{-\frac{|x|^2}{2\varepsilon}}$), then

$$(\mu_t^\varepsilon, v_t^\varepsilon, w_t^\varepsilon) := \left(\mu_t * \rho_\varepsilon, \frac{(v_t \mu_t) * \rho_\varepsilon}{\mu_t^\varepsilon}, \frac{(w_t \mu_t) * \rho_\varepsilon}{\mu_t^\varepsilon} \right) \tag{1.20}$$

solves (1.14) where $t \mapsto \mu_t^\varepsilon$ is continuous and it satisfies the uniform bounds

$$\int_{\mathbb{R}^d} \left(|v_t^\varepsilon|^p + \frac{1}{4} |w_t^\varepsilon|^p \right) d\mu_t^\varepsilon \leqslant \int_{\mathbb{R}^d} \left(|v_t|^p + \frac{1}{4} |w_t|^p \right) d\mu_t. \tag{1.21}$$

Moreover,

$$\lim_{\varepsilon \to 0} \|v_t^\varepsilon\|_{L^p(d\mu_t^\varepsilon)}^p = \|v_t\|_{L^p(d\mu_t)}^p \quad \text{and} \quad \lim_{\varepsilon \to 0} \|w_t^\varepsilon\|_{L^p(\mu_t^\varepsilon)}^p = \|w_t\|_{L^p(\mu_t)}^p. \tag{1.22}$$

2. Balanced Optimal Transport

This chapter provides an overview of the classical optimal transport problem, which involves determining the minimum total cost required to transport between two probability distributions, each with a total mass of 1. We will focus on summarizing the principal problems associated with balanced transfers in optimal transport.

2.1 Monge Problem

Monge's problem revolves around computing the optimal map that minimizes the cost of transporting a specified mass distribution from one location to another. Given an initial measure $\mu \in \mathcal{P}(X)$, a target measure $\nu \in \mathcal{P}(Y)$, and a Borel cost function $c : X \times Y \to [0, \infty)$, representing the cost to transport a unit of mass from x to y, the **Monge problem** is formulated as

$$\inf \left\{ \int_X c(x, T(x))\, d\mu(x) : T : X \to Y \text{ Borel, } T_\# \mu = \nu \right\}, \tag{M}$$

where the infimum is taken over all push-forward map T, called Monge map. We denote by $\mathcal{C}_\mu(T)$ the transport cost $\int_X c(x, T(x))\, d\mu(x)$. The minimum may not be attained, for example, when μ is a Dirac measure but ν is not, in which case the Monge map does not exist.

Theorem 6 (Existence, Uniqueness in $\mathbb{R}$, [ABS21])**.** If $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and μ has no atoms (or $\mu \ll \mathcal{L}^1$), then there exists $T : \mathbb{R} \to [-\infty, \infty]$ non-decreasing pushing μ into ν and any other map S with these properties coincides with T on $spt(\mu)$, with at most countably many exceptions.

If $c(x, y) = \phi(|y - x|)$ with $\phi : [0, \infty) \to [0, \infty)$ convex and non-decreasing, and if $\mathcal{C}_\mu(T) < \infty$, then T is an optimal map. If ϕ is strictly convex, then T is the unique map.

2.2 Kantorovich Problem

Kantorovich's problem seeks to determine the optimal transport plan between two probability measures, offering a more general perspective on optimal transport issues compared to Monge's problem. While Monge's problem revolves around mapping each point from one measure to a point in another, Kantorovich's approach allows for a more flexible "splitting" of mass, facilitating transport from a point in the source measure to multiple points in the target. This problem involves not only finding the most cost-efficient way to transport mass but also ensuring feasibility in scenarios where Monge's problem does not provide solutions, thus expanding applicability to a broader array of transport issues.

Firstly, we introduce the set of transport plans from $\mu \in \mathcal{P}(X)$ to $\nu \in \mathcal{P}(Y)$, defined as

$$\Gamma(\mu, \nu) := \left\{ \pi \in \mathcal{P}(X \times Y) : (p_X)_\# \pi = \mu, (p_Y)_\# \pi = \nu \right\}, \tag{2.1}$$

where $p_X : X \times Y \to X$, $p_Y : X \times Y \to Y$ are projections maps. $\Gamma(\mu, \nu)$ is not empty, since $\mu \times \nu \in \Gamma(\mu, \nu)$. For any $A \times B \in X \times Y$, $\pi(A \times B)$ represent the mass transported from the locations in A to B. Then **Kantorovich's problem** is defined as

$$\inf \left\{ \int_{X \times Y} c(x, y)\, d\pi(x, y) : \pi \in \Gamma(\mu, \nu) \right\}. \tag{K}$$

Denote $Opt_K(\mu, \nu)$ by the set of all optimal transport plans π for the Kantorovich problem.

We also denote by $\mathcal{C}(\pi)$ the transport cost $\int_{X \times Y} c(x, y)\, d\pi$. If there exists an optimal Monge map T of $\mathcal{C}_\mu(T)$, then $\pi_T := (\mathrm{id}, T)_\# \mu$ gives a lower transportation cost than Monge problem since

$$\inf_{(K)} \leqslant \mathcal{C}(\pi_T) = \int c(x, T(x))\, d\mu(x) = \mathcal{C}_\mu(T) = \inf_{(M)}.$$

Theorem 7 (Existence, [ABS21])**.** Let X and Y be Polish spaces and $c : X \times Y \to \mathbb{R}^+ \cup \{+\infty\}$ be lower semi-continuous. Then the minimum in (K) is attained.

According to [Oxt70], in the Polish setting, the presence of Monge maps from μ to ν requires that μ is non-atomic. Building on this foundation, [Pra07] identified the optimal conditions under which the infimum of Monge's problem aligns with the minimum of Kantorovich's problem.

Theorem 8 ([Pra07])**.** Let X and Y be Polish spaces. If μ is non-atomic and $c : X \times Y \to \mathbb{R}^+ \cup \{+\infty\}$ is continuous, then

$$\inf_{(M)} = \min_{(K)}.$$

2.3 Dual Problem

The dual problem is finding a pair to maximize the cost, which always produces no higher cost than the primal problem — Kantorivich's problem. Assume that X and Y are Polish spaces. For any $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, the **Dual problem** is defined as

$$\sup_{(\phi,\psi)\in I_c} \left\{ \int_X \phi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \right\}, \tag{2.2}$$

where the supremum is taken over the set I_c where

$$I_c := \left\{ (\phi, \psi) \in \mathrm{Lip}_b(X) \times \mathrm{Lip}_b(Y) : \phi(x) + \psi(y) \leqslant c(x, y) \right\}.$$

Theorem 9 (Duality, [ABS21])**.** For any cost function $c \in \mathrm{Lip}_b(X \times Y)$,

$$\min_{(K)} = \sup_{(2.2)} \tag{duality}$$

where the supremum in (2.2) is a maximum attained by a bounded and Lipschitz pair (ϕ, ψ). More specifically, given an optimal transport plan $\hat{\pi}$ of (K) and fixing $(x_0, y_0) \in spt(\hat{\pi})$,

$$\phi(x) := \inf \left\{ c(x, y_N) - c(x_N, y_N) + c(x_N, y_{N-1}) - \ldots + c(x_1, y_0) - c(x_0, y_0) \right\}, \tag{2.3}$$

where the infimum is taken over all $N \geqslant 1$ and $\left\{(x_i, y_i)\right\}_{i=1}^{N} \in spt(\hat{\pi})$, and

$$\psi(y) := \inf \left\{ c(x, y) - \phi(x) : x \in X \right\}. \tag{2.4}$$

By establishing duality and thanks to [[Bre91], [KS84]], the uniqueness of the optimal transport plan of (K) is thereby confirmed in the following theorem.

Theorem 10 ([ABS21, Brenier, Knott-Smith]). Assume that $X = Y = \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$, $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu \ll \mathcal{L}^d$.

(i) Then the problem (K) has a unique solution $\pi = (id, T)_{\#}\mu$ where $T = \nabla\varphi$ and $\varphi : \mathbb{R}^d \to (-\infty, \infty]$ is a lower semi-continuous convex function differentiable $\mu-$almost everywhere.

(ii) If φ is convex, lower semi-continuous, differentiable μ-almost everywhere with $|\nabla\varphi| \in L^2(\mathbb{R}^d, \mu)$, then $T := \nabla\varphi$ is optimal from μ to $\nu := T_{\#}\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

(iii) if $\nu \ll \mathcal{L}^d$, denoting by $T^{\mu\to\nu}$ (resp. $T^{\nu\to\mu}$) the unique optimal transport map between μ and ν (resp. ν and μ), we get that

$$T^{\nu\to\mu} \circ T^{\mu\to\nu} = id \quad \mu\text{-a.e. in } \mathbb{R}^d, \qquad T^{\mu\to\nu} \circ T^{\nu\to\mu} \quad \nu\text{-a.e. in } \mathbb{R}^d.$$

Remark 2.5. If (ϕ, ψ) is the optimal solution of the dual problem, then φ is $\frac{1}{2}|x|^2 - \phi(x)$.

2.4 Wasserstein Distance

The optimal transport problem can be used to endow $\mathcal{P}_p(X)$ for any $1 \leqslant p < \infty$ with a natural metric structure. Many metric properties (separability, compactness, completeness, geodesic, nonbranching, lower bounds on sectional curvature) can be lifted from X to $\mathcal{P}_p(X)$. In this section, assume that (X, τ) is Polish, then there exists a distance d on X inducing τ such that (X, d) is complete and separable.

According to [ABS21], the Wasserstein distance in $\mathcal{P}_p(X)$ is a Kantorovich problem when $c(x,y) = d(x,y)^p$ for $p \in [1, \infty)$. More specifically, given $\mu \in \mathcal{P}_p(X), \nu \in \mathcal{P}_p(X)$, the **Wasserstein distance** W_p in $\mathcal{P}_p(X)$ is defined as

$$W_p^p(\mu,\nu) := \min\left\{ \int_{X\times X} d(x,y)^p \, d\pi(x,y) : \pi \in \Gamma(\mu,\nu)\right\} \tag{W_p}$$

and the extended distance W_∞ is defined as

$$W_\infty(\mu,\nu) := \inf\left\{ \|d(x,y)\|_{L^\infty(\pi)} : \pi \in \Gamma(\mu,\nu)\right\}. \tag{W_∞}$$

Remark 2.6 (Wasserstein Distance in $\mathcal{P}_2(\mathbb{R}^d)$)**.** The Euclidean space $\mathbb{R}^d$ is endowed with Euclidean distance. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\mu,\nu) = \min\left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\pi(x,y) : \pi \in \Gamma(\mu,\nu) \right\}. \tag{W_2}$$

According to [ABS21], for $p \geqslant 1$, if (X,d) is a metric space, $(\mathcal{P}_p(X), W_p)$ is a metric space. If (X,d) is a complete metric space, $(\mathcal{P}_p(X), W_p)$ is complete as well. If (X,d) is compact, then $(\mathcal{P}_p(X), W_p)$ is compact. If (X,d) is Polish, then $(\mathcal{P}_p(X), W_p)$ is separable. If (X,d) is a geodesic space, then $(\mathcal{P}_p(X), W_p)$ is geodesic as well.

2.5 Dynamic Formulation of Wasserstein Distance

The optimal transport plans of Kantorovich problem establish the origin and destination points of a transportation process. However, these plans alone do not provide a comprehensive understanding of the intricate mechanisms underlying the movement and transformation of mass during transit. By delving into the concept of absolute continuity, we can elucidate the trajectories employed for mass transfer. These trajectories manifest as curves within a geodesic space. With the pre-background in Preliminaries, leveraging an action functional $\mathcal{A}_2$, [ABS21] introduced a dynamic formulation of the Wasserstein distance in $\mathcal{P}_2(X)$. This dynamic formulation quantifies the total cost expended in transporting mass along all conceivable curves.

More formally, let $I = [0,1]$ be the time interval and $e_t : C(I;X) \to X$ be the evaluation map defined by $e_t(\gamma) := \gamma(t)$ for any $t \in I$.

Definition 2.7. The **Dynamic Formulation** of the Wasserstein distance W_2 between $\mu, \nu \in \mathcal{P}_2(X)$ is given by

$$\min\left\{ \int_{C(I;X)} \mathcal{A}_2(\gamma)\, d\eta(\gamma) : \eta \in \mathcal{P}(C(I;X)),\ (e_0)_\#\eta = \mu,\ (e_1)_\#\eta = \nu \right\} \tag{2.8}$$

where $\mathcal{A}_2 : C(I;X) \to [0,\infty]$ is the action of a curve defined as

$$\mathcal{A}_2(\gamma) := \begin{cases} \int_0^1 |\gamma'|_d^2(t)\, dt & \text{if } \gamma \in AC(I;(X,d)) \\ +\infty & \text{else} \end{cases}. \tag{2.9}$$

We denote by $OptGeo(\mu,\nu) \subset \mathcal{P}(C(I;X))$ the collection of optimal geodesic plans from μ to ν of the dynamic formulation in (2.8).

This dynamic framework is an equivalent formulation of Kantorovich's problem, stated in the following theorem.

Theorem 11 ([ABS21])**.** If (X,d) is a Polish and geodesic metric space, then the dynamic formulation in (2.8) is equivalent to the Kantorovich problem in (K).

In addition, $\eta \in OptGeo(\mu,\nu)$ if and only if η is supported in $Geo(X)$ and $(e_0,e_1)_\#\eta \in Opt_K(\mu,\nu)$, an optimal transport plan of the Kantorovich problem.

2.6 Absolutely Continuous Curves in a Probability Space

This section aims to characterize a time-independent curve $(\mu_t)_{t\in I} \in C(I;\mathcal{P}(X))$ connecting two measures $\mu_0,\mu_1 \in \mathcal{P}(X)$. Building on this, [ABS21] investigates a lifting relationship between the absolutely continuous curves in a probability space, metricized by the Wasserstein distance, and a dynamic plan within absolutely continuous curves in a metric space (X,d).

Due to [ABS21, Lecture 10], a dynamic plan $\eta \in \mathcal{P}(C(I;X))$ is a lifting of a curve $\mu_t : I \to \mathcal{P}(X)$ if $(e_t)_\#\eta = \mu_t$ for all $t \in I$. The theorem presented below demonstrates that any absolute continuous curve in a probability space metricized by Wasserstein distance induces a dynamic plan.

Theorem 12 ([Lis06, Theorem 5])**.** Let (X,d) be a complete and separable metric space. If $\mu_t \in AC^p(I;(\mathcal{P}_p(X),W_p))$ for $p>1$, then there exists $\eta \in \mathcal{P}(C(I;X))$ such that

1. η is concentrated on $AC^p(I;(X,d))$, i.e. $\eta\big(C(I;X)\backslash AC^p(I;(X,d))\big) = 0$.
2. $(e_t)_\#\eta = \mu_t$ for all $t \in I$.
3. $|\mu'|^p_{W_p}(t) = \int_{C(I;X)} |\gamma'|^p_d(t)\, d\eta(\gamma)$ for $\mathcal{L}^1-$almost every $t \in I$.

Building upon this foundation, [Lis06] extended these findings to the geodesic of $\mathcal{P}_p(X)$.

Theorem 13 ([Lis06, Theorem 6]). Let (X, d) be a complete and separable length space. A curve $(\mu_t)_{t\in I}$ is a constant speed minimizing geodesic of $\mathcal{P}_p(X)$ if and only if there exists $\eta \in \mathcal{P}(C(I; X))$ such that

1. η is concentrated on G where
$$G := \{\gamma : I \to X : \gamma \text{ is a constant speed minimizing geodesic of } X\}. \tag{2.10}$$
2. $(e_t)_\#\eta = \mu_t$ for all $t \in I$.
3. $W_p^p(\mu_0, \mu_1) = \int_{C(I;X)} d^p(\gamma(0), \gamma(1))\, d\eta(\gamma)$.

In their published book, [ABS21] more clearly outlined the relationship between the geodesic in $\mathcal{P}_2(X)$ and its lifting. They noted that necessary conditions are required to guarantee the uniqueness of the geodesic and its lifting.

Theorem 14 (Stablitiy of the Geodesic Property, [ABS21]). If (X, d) is a geodesic metric space, then $(\mathcal{P}_2(X), W_2)$ is geodesic as well. More precisely, any $\eta \in OptGeo(\mu_0, \mu_1)$ induces a constant speed geodesic $\mu_t = (e_t)_\#\eta$ in $\mathcal{P}_2(X)$ from μ_0 to μ_1. Conversely, for any $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(X))$, there exists a (possiblt nonunique) lifting $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$.

Proposition 15 ([ABS21, Proposition 10.9]). Let (X, d) be a geodesic metric space. Fix $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ and assume that there exists a unique optimal transport plan $\pi \in \Gamma(\mu_0, \mu_1)$ and that $\pi-$a.e. $(x, y) \in X \times X$ are joined by a unique constant speed geodesic $\Gamma(x, y) \in Geo(X)$. Then there exists a unique $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$, given by $\Gamma_\#\pi$, and a unique $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(X))$ connecting μ_0 and μ_1, given by $(e_t)_\#\eta$.

Corollary 16 (Uniqueness of Geodesics in $\mathcal{P}_2(\mathbb{R}^d)$, [ABS21]). Assume that $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\mu_0 \ll \mathcal{L}^d$. Then there exists a unique $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$ and a unique constant speed geodesic $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^d))$ joining μ_0 and μ_1, given by
$$\mu_t := (T_t)_\#\mu_0, \tag{2.11}$$
where $T = \nabla\phi$ is the optimal map from μ_0 to μ_1 and T_t is the interpolated transport map $T_t := (1 - t)\mathrm{id} + tT$.

2.7 Benamou-Brenier Problem

In their seminal work, [BB00] introduced an innovative framework within continuum mechanics for the Wasserstein distance W_2 in $\mathcal{P}_2(\mathbb{R}^d)$. This framework presents a dynamic perspective, offering a time-dependent interpolant $\mu(t,x)$ that transports from the initial measure μ_0 and the target measure μ_1. Additionally, it provides a velocity field $v(t,x)$ that facilitates mass movement while adhering to the continuity equation:

$$\partial_t \mu(t,x) + \nabla \cdot \big(v(t,x)\mu(t,x)\big) = 0. \tag{2.12}$$

From a computational point of view, this novel formulation brings significant advantages to numerical solutions. It introduces convexity through the incorporation of the time variable, and notably, the continuity constraint remains linear, streamlining the computational process, which may be used for applications in computational vision.

Theorem 17 (Benamou-Brenier Formula)**.** For all $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\mu_0,\mu_1) = \min\left\{ \int_0^1 \int_{\mathbb{R}^d} |v(t,x)|^2 \,\mu(t,dx)\,dt : \frac{d}{dt}\mu + \nabla\cdot(v\mu) = 0 \text{ in } (0,1)\times\mathbb{R}^d \right\} \tag{2.13}$$

where the minimization is among all curve $\mu_t : [0,1] \to \mathcal{P}_2(\mathbb{R}^d)$ continuous w.r.t the weak topology.

In their work, [ABS21] provided a fresh perspective on proving the Theorem 17, distinct from the methodology adopted by [BB00]. Moreover, they successfully established a close correspondence between absolutely continuous curves in $\mathcal{P}_2(\mathbb{R}^d)$ and solutions to the continuity equations. Fix the time interval $I = [0,1]$.

Proposition 18 ([ABS21, Proposition 17.9])**.** Let $t \mapsto \mu_t : I \to \mathcal{P}_2(\mathbb{R}^d)$ be a solution to the continuity equation in (2.12) induced by the vector field $(v_t)_{t\in I}$ and assume that $\int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \,d\mu_t(x)\,dt < \infty$. Then $\mu \in AC^2(I;(\mathcal{P}_2(\mathbb{R}^d), W_2))$ and

$$|\mu'|^2_{W_2}(t) \leqslant \int_{\mathbb{R}^d} |v_t(x)|^2 \,d\mu_t(x) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{2.14}$$

Theorem 19 ([ABS21, Thereom 17.10])**.** Given $\mu \in AC^2(I; (\mathcal{P}_2(\mathbb{R}^d), W_2))$, there exists a velocity field $(v_t)_{t\in I}$ such that μ solves the associated continuity equation in (2.12) and

$$|\mu'|^2_{W_2}(t) = \int_{\mathbb{R}^d} |v_t(x)|^2 \, d\mu_t(x) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{2.15}$$

3. Unbalanced Optimal Transport

The unbalanced optimal transport problem aims to minimize the total cost of moving an object from one location to another, accommodating variations in the initial and target masses. This scenario differs from the balanced optimal transport scenario, where the mass remains constant throughout the transportation process. The allowance for mass variation in unbalanced transport often leads to a reduction in the overall transportation cost. This chapter will delve into the unbalanced optimal transport problem between two Radon measures, allowing for different total masses. The discussion primarily draws upon the foundational work presented in [LMS16] and [LMS18], offering insights into the methodologies and results detailed within these studies.

3.1 Optimal Entropy-Transport Problems

[LMS18] developed a full theory of static formulations between non-negative and finite Radon measures in general topological spaces, which is called the Entropy-Transport problem. This problem addresses the challenge of determining optimal transport plans between two measures when their total masses are unequal. It is constructed from the Kantorovich formulation by adding the penalty of the deviation from $\mu \in \mathcal{M}(X)$ and $\nu \in \mathcal{M}(Y)$ to its corresponding marginals $\pi_X := (p_X)_\# \pi$ and $\pi_Y := (p_Y)_\# \pi$, respectively. By choosing convex entropy functions $F_i : [0, \infty) \to [0, \infty]$, the **Entropy-Transport problem** is formulated as

$$\mathbf{ET}(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) \, d\pi + \mathcal{F}_1(\pi_X|\mu) + \mathcal{F}_2(\pi_Y|\nu) : \pi \in \mathcal{M}(X \times Y) \right\}. \tag{ET}$$

The additional term in (ET) is a sum of penalizing functionals such that the entropy functional $\mathcal{F}_1$ is defined as

$$\mathcal{F}_1(\pi_X|\mu) := \int_X F_1(\sigma_1) \, d\mu + (F_1)'_\infty \pi_X^\perp(X)$$

$$\sigma_1 = \frac{d\pi_X}{d\mu}, \quad \pi_X = \sigma_1 \mu + \pi_X^\perp \quad \text{(Lebesgue decomposition)},$$

and the entropy functional $\mathcal{F}_2$ is defined as

$$\mathcal{F}_2(\pi_Y|\nu) := \int_Y F_2(\sigma_2)\, d\nu + (F_2)'_\infty \pi_Y^\perp(Y)$$

$$\sigma_2 = \frac{d\pi_Y}{d\nu}, \quad \pi_Y = \sigma_2\nu + \pi_Y^\perp \quad \text{(Lebesgue decomposition)},$$

where $(F_i)'_\infty$ are their recession constants such that $(F_i)'_\infty := \lim_{s\to\infty} \frac{F_i(s)}{s}$. The **Logarithmic Entropy-Transport problem**, denote $\mathbf{LET}$, is established by choosing a specific cost function c as

$$c(x,y) := \ell(d(x,y)) \quad \text{with } \ell(d) := \begin{cases} \log(1+\tan^2(d)) & \text{if } d \in [0, \frac{\pi}{2}) \\ +\infty & \text{if } d \geqslant \frac{\pi}{2} \end{cases}$$

and entropy functions $F_i(s) := s\log(s) - s + 1$ for $i \in \{1,2\}$. It is well-known that $(\mu,\nu) \mapsto \sqrt{\mathbf{LET}(\mu,\nu)}$ defines a distance in $\mathcal{M}(X)$, which is equivalent to Hellinger-Kantorovich Distance introduced in the next section.

3.2 Hellinger-Kantorovich Distance

From a geometric perspective, the Hellinger-Kantorovich problem is approached as the minimization of a balanced optimal transport problem between two measures within a cone space. Before presenting the definition, it is essential to first introduce the concept of a cone structure.

3.2.1 Cone Structure

The construction of cone space is cited from [BBI01, Section 3.6.2] and [LMS18, Section 7]. Let the space X be endowed with a metric d. A cone $\mathfrak{C}$ over a topological space X is the quotient space of $\mathcal{Y} := X \times \mathbb{R}_+$, i.e. $\mathfrak{C} = \mathcal{Y}/\sim$, where

$$\forall y_i = (x_i, r_i) \in \mathcal{Y}, \quad y_1 \sim y_2 \iff r_1 = r_2 = 0 \text{ or } r_1 = r_2,\, x_1 = x_2.$$

Any element in $\mathfrak{C}$ denotes as $\mathfrak{y} = [x,r]$ together with the vertex $\mathfrak{o} = [x,0]$ for all $x \in X$, where the complement of the vertex is $\mathfrak{C}_\mathfrak{o} = \mathfrak{C}\backslash\{\mathfrak{o}\}$. We endow $\mathfrak{C}$ with a topology $\tau_\mathfrak{C}$ induced by a cone metric $d_{\mathfrak{C}}$ given by

$$d^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) := r_1^2 + r_2^2 - 2r_1 r_2 \cos(d_{\frac{\pi}{2}}(x_1, x_2)) \tag{3.1}$$

$$= (r_1 - r_2)^2 + 4r_1 r_2 \sin^2(d_{\frac{\pi}{2}}(x_1, x_2)/2) \tag{3.2}$$

and $d_{\frac{\pi}{2}}(x_1, x_2) := d(x_1, x_2) \wedge \frac{\pi}{2}$. The metric $d^2_{\mathfrak{C}}([x_1, \cdot], [x_2, \cdot])$ is 2-homogenous, i.e.

$$d^2_{\mathfrak{C}}([x_1, r_1], [x_2, r_2]) = d^2_{\mathfrak{C}}([x_1, \frac{r_1}{v}], [x_2, \frac{r_2}{v}]) \cdot v^2, \quad \forall v > 0. \tag{3.3}$$

A neighborhood of the vertex $\mathfrak{o}$, shown in Figure 3.1, is

$$\{\mathfrak{y} \in \mathfrak{C} : d_{\mathfrak{C}}(\mathfrak{o}, \mathfrak{y}) < \varepsilon\} = \{[x, r] : 0 \leqslant r < \varepsilon\}, \quad \varepsilon > 0. \tag{3.4}$$

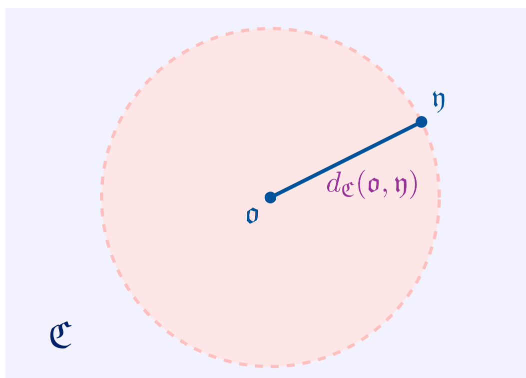


Figure 3.1: A neighborhood of the vertex $\mathfrak{o}$

Under this setting, $(\mathfrak{C}, \tau_{\mathfrak{C}})$ is a Hausdorff topological space and $(\mathfrak{C}, d_{\mathfrak{C}})$ is a complete separate metric space known as Polish space if (X, d) is. In [BBI01, Section 3.6.2], if (X, d) is a geodesic (resp. length) space, then also $(\mathfrak{C}, d_{\mathfrak{C}})$ is a geodesic (resp. length) space.

Furthermore, the product space $\mathcal{Y} = X \times \mathbb{R}_+$ can be projected to the cone space $\mathfrak{C}$ via the map $\mathsf{p} : \mathcal{Y} \to \mathfrak{C}$ where $\mathsf{p}(x, r) = [x, r]$. Conversely, we can project $\mathfrak{C}$ to $\mathcal{Y}$ by the inverse projection $\mathsf{y} : \mathfrak{C} \to \mathcal{Y}$ of p, which obtained by fixing a point $\bar{x} \in X$, then $\mathsf{y} = (\mathsf{x}, \mathsf{r})$ where

$$\mathsf{r} : \mathfrak{C} \to \mathbb{R}_+, \quad \mathsf{r}[x, r] = r, \tag{3.5}$$

$$\mathsf{x} : \mathfrak{C} \to X, \quad \mathsf{x}[x, r] = \begin{cases} x & \text{if } r > 0, \\ \bar{x} & \text{if } r = 0. \end{cases} \tag{3.6}$$

3.2.2 Wasserstein Distance and Homogeneous Marginals

This subsection will explore projecting the measures from a cone space to a regular space, with definitions from [LMS18]. Denote by $\mathfrak{C}^{\otimes N}$ the product spaces and $\mathfrak{y} = (\mathfrak{y}_i)_{i=1}^N = ([x_i, r_i])_{i=1}^N$ the element in $\mathfrak{C}^{\otimes N}$. We equip $\mathfrak{C}^{\otimes N}$ with the metric

$$d_{\mathfrak{C}}(\mathfrak{y}', \mathfrak{y}'') = \left(\sum_{i=1}^{N} d_{\mathfrak{C}}^2(\mathfrak{y}_i', \mathfrak{y}_i'') \right)^{\frac{1}{2}}, \quad \text{for } \mathfrak{y}' = (\mathfrak{y}_i')_{i=1}^N \text{ and } \mathfrak{y}'' = (\mathfrak{y}_i'')_{i=1}^N. \tag{3.7}$$

Let $p^i : \mathfrak{C}^{\otimes N} \to \mathfrak{C}$ be the projections on the i-th coordinate.

Definition 3.8 ($L2-$Kantorovich-Wasserstein distance $W_{d_{\mathfrak{C}}}$)**.** For any $\nu_1, \nu_2 \in \mathcal{M}_2(\mathfrak{C})$,

$$W_{d_{\mathfrak{C}}}^2(\nu_1, \nu_2) := \min \left\{ \int_{\mathfrak{C} \times \mathfrak{C}} d_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2) \, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) : \alpha \in \mathcal{M}(\mathfrak{C} \times \mathfrak{C}),\, p^i_{\#}\alpha = \nu_i \right\}, \tag{3.9}$$

and $W_{d_{\mathfrak{C}}}(\nu_1, \nu_2) = +\infty$ if $\nu_1(\mathfrak{C}) \neq \nu_2(\mathfrak{C})$. Denote $Opt_{W_{d_{\mathfrak{C}}}}(\mu, \nu)$ by the set of optimal transport plans for $W_{d_{\mathfrak{C}}}^2(\nu_1, \nu_2)$.

We can extend the projection maps r in (3.5) and x in (3.6) to a product space $\mathfrak{C}^{\otimes N}$ by

$$\mathsf{r}_i : \mathfrak{C}^{\otimes N} \to [0, \infty), \quad \mathsf{r}_i(\mathfrak{y}) := \mathsf{r}(\mathfrak{y}_i) = r_i, \tag{3.10}$$

$$\mathsf{x}_i : \mathfrak{C}^{\otimes N} \to X_i, \quad \mathsf{x}_i(\mathfrak{y}) := \mathsf{x}(\mathfrak{y}_i) = x_i, \tag{3.11}$$

and also the projections $\mathbf{p} = \mathsf{p}^{\otimes N} : \mathcal{Y}^{\otimes N} \to \mathfrak{C}^{\otimes N}$ and the liftings $\mathbf{y} = \mathsf{y}^{\otimes N} : \mathfrak{C}^{\otimes N} \to \mathcal{Y}^{\otimes N}$.

Consider a plan $\alpha \in \mathcal{M}(\mathfrak{C}^{\otimes N})$, we say that $\alpha \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$ if

$$\int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_2^2 \, d\alpha < \infty, \quad \text{where } |\mathfrak{y}|_2^2 := \sum_{i=1}^{N} \mathsf{r}_i(\mathfrak{y})^2 = \sum_{i=1}^{N} r_i^2. \tag{3.12}$$

Any measure $\alpha \in \mathcal{M}(\mathfrak{C}^{\otimes N})$ can be projected onto a measure in $\mathcal{M}(X_i)$, via $\mathfrak{h}_i^2\alpha \in \mathcal{M}(X_i)$, for any $i \in \{1, \ldots, N\}$, where the homogeneous marginals are defined by

$$\mathfrak{h}_i^2 : \mathcal{M}(\mathfrak{C}^{\otimes N}) \to \mathcal{M}(X_i), \quad \mathfrak{h}_i^2 := (\mathsf{x}_i)_{\#}\mathsf{r}_i^2, \quad \forall i \in \{1, \ldots, N\}, \tag{3.13}$$

and the notation $\mathfrak{h}^2 = (\mathsf{x})_{\#}\mathsf{r}^2$ is used when $N = 1$. The function r_i scales the measure α, and then the function x_i projects the cone space $\mathfrak{C}^{\otimes N}$ onto the space X_i.

If $\vartheta : \mathfrak{C}^{\otimes N} \to (0, \infty)$ is a Borel map in $L^2(\mathfrak{C}^{\otimes N}, \alpha)$, we can define a scaled transport plan by dilation, $\mathrm{dil}_{\vartheta,2}\alpha := (\mathrm{prd}_\vartheta)_\#(\vartheta^2\alpha)$, where

$$\big(\mathrm{prd}_\vartheta(\mathfrak{y})\big)_i := \mathfrak{y}_i \cdot (\vartheta(\mathfrak{y}))^{-1} = [x_i, \frac{r_i}{\vartheta(\mathfrak{y})}]. \tag{3.14}$$

For example, if $N = 2$ and $\mathfrak{y} = (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C}^{\otimes 2}$, then

$$\mathrm{prd}_\vartheta(\mathfrak{y}) = \big([x_1, \frac{r_1}{\vartheta(\mathfrak{y})}], [x_2, \frac{r_2}{\vartheta(\mathfrak{y})}]\big). \tag{3.15}$$

Thus, the homogeneous marginals is invariant w.r.t dilation, $\mathfrak{h}_i^2\big(\mathrm{dil}_{\vartheta,2}(\alpha)\big) = \mathfrak{h}_i^2\alpha$, i.e. for any $\varphi \in \mathcal{B}(X_i)$,

$$\int_{X_i} \varphi(x_i)\, d[\mathfrak{h}_i^2(\mathrm{dil}_{\vartheta,2}\alpha)] = \int_{X_i} \varphi(x_i)\, d[\mathfrak{h}_i^2\alpha]. \tag{3.16}$$

3.2.3 The Hellinger-Kantorovich problem

Definition 3.17 (The HK Problem, [LMS18])**.** Given $\mu_1, \mu_2 \in \mathcal{M}(X)$, the **Hellinger-Kantorovich problem** is defined as

$$\mathbf{HK}^2(\mu_1, \mu_2) := \inf\left\{ \int_{\mathfrak{C}\times\mathfrak{C}} d_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) : \alpha \in \mathcal{M}_2(\mathfrak{C}\times\mathfrak{C}),\ \mathfrak{h}_i^2\alpha = \mu_i \right\}, \tag{3.18}$$

where the marginal constraint $\mathfrak{h}_i^2\alpha = \mu_i$ holds w.r.t the weak topology: For any $\varphi \in C_b(X)$,

$$\int_X \varphi(x_i)\, d\mu_i(x_i) = \int_{\mathfrak{C}\times\mathfrak{C}} \varphi(x_i)\, r_i^2\, d\alpha([x_1, r_1], [x_2, r_2]). \tag{3.19}$$

We denote $Opt_{\mathbf{HK}}(\mu_1, \mu_2)$ by the set of all optimal plans α for $\mathbf{HK}^2(\mu_1, \mu_2)$. And the existence of a minimum is guaranteed in Theorem 21.

It is shown that $\mathbf{HK}$ defines a metric in $\mathcal{M}(X)$ and $(\mathcal{M}(X), \mathbf{HK})$ satisfyies many metric properties. If (X, d) is separable, then $(\mathcal{M}(X), \mathbf{HK})$ is as well. If (X, d) is complete, then $(\mathcal{M}(X), \mathbf{HK})$ is complete. Moreover, $(\mathcal{M}(X), \mathbf{HK})$ is a length (resp. geodesic) space if and only if (X, d) is a length (resp. geodesic) space.

Lemma 20 ([LMS18, Section 7.3, Section 5.2])**.** Let $\mathfrak{C} = \mathfrak{C}^{\otimes 2}$. Given any $\alpha \in \mathcal{M}_2(\mathfrak{C})$ with $\alpha(\mathfrak{C}) > 0$. If the scaling size is chosen as

$$\vartheta(\mathfrak{y}) = \frac{1}{r_*}\begin{cases} |\mathfrak{y}|_2 & \text{if } |\mathfrak{y}|_2 \neq 0, \\ 1 & \text{if } |\mathfrak{y}|_2 = 0, \end{cases} \qquad r_* := \left(\int |\mathfrak{y}|_2^2\, d\alpha + \alpha(|\mathfrak{y}|_2 = 0)\right)^{\frac{1}{2}}, \tag{3.20}$$

then we obtain a rescaled probability measure $\tilde{\alpha} := \mathrm{dil}_{\vartheta,2}\alpha$ with the same homogeneous marginals as α and concentrated on $\mathfrak{C}[r_*] := \{\mathfrak{y} = (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C} : |\mathfrak{y}|_2 \leqslant r_*\}$:

$$\tilde{\alpha} \in \mathcal{P}_2(\mathfrak{C}), \qquad \mathfrak{h}_i^2\tilde{\alpha} = \mathfrak{h}_i^2\alpha, \qquad \tilde{\alpha}(\mathfrak{C}\backslash\mathfrak{C}[r_*]) = 0.$$

Remark 3.21. Since the transport from the vertex to vertex has no physical meaning and $\tilde{d}(\mathfrak{o}, \mathfrak{o}) = 0$, then we can restrict the transport plan $\hat{\alpha} = \alpha\big|_{\mathfrak{C}\backslash(\mathfrak{o},\mathfrak{o})}$, then

$$r_* = \left(\int |\mathfrak{y}|_2^2\, d\alpha\right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{2} \mu_i(X)} := R. \tag{3.22}$$

Thus, we can obtain a rescaled measure $\tilde{\alpha} := \mathrm{dil}_{\vartheta,2}\hat{\alpha}$ such that

$$\tilde{\alpha} \in \mathcal{P}_2(\mathfrak{C}), \qquad \mathfrak{h}_i^2\tilde{\alpha} = \mathfrak{h}_i^2\alpha, \qquad \tilde{\alpha}(\mathfrak{C}\backslash\mathfrak{C}[R]) = 0.$$

where $\mathfrak{C}[R] := \{\mathfrak{y} = (\mathfrak{y}_1, \mathfrak{y}_2) \in \mathfrak{C} : |\mathfrak{y}|_2 \leqslant R\}$.

Due to Lemma 20 and that cone metric and the homogeneous marginals are invariant under the rescaling, the Hellinger-Kantorovich problem can be reformulated as

$$\mathbf{HK}^2(\mu_1, \mu_2) = \min\left\{\int d_{\mathfrak{C}}^2(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha : \alpha \in \mathcal{P}(\mathfrak{C}),\ \mathfrak{h}_i^2\alpha = \mu_i,\ \alpha(\mathfrak{C}\backslash\mathfrak{C}[R]) = 0\right\}. \tag{3.23}$$

Theorem 21 (Existence of Optimal Transport Plans for the HK Problem, [LMS18, Section 7])**.** For any $\mu_1, \mu_2 \in \mathcal{M}(X)$, the Hellinger-Kantorovich problem always admits a solution $\alpha \in \mathcal{P}(\mathfrak{C})$ concentrated on $\mathfrak{C}[R]\backslash\{(\mathfrak{o}, \mathfrak{o})\}$.

Corollary 22 (HK and Wasserstein distance on $\mathcal{P}_2(\mathfrak{C})$, [LMS18, Section 7])**.** For every $\mu_1, \mu_2 \in \mathcal{M}(X)$, we have

$$\mathbf{HK}(\mu_1, \mu_2) = \min\left\{W_{d_{\mathfrak{C}}}(\nu_1, \nu_2) : \nu_i \in \mathcal{P}_2(\mathfrak{C}),\ \mathfrak{h}^2\nu_i = \mu_i\right\}. \tag{3.24}$$

3.3 Dynamic Reformulation for the HK Distance

Let $X = \mathbb{R}^d$ be endowed with the Euclidean distance $d(x_1, x_2) = |x_1 - x_2|$. Fix a time interval $I = [0, 1]$. In this section, we will obtain a characterization of absolutely continuous curves $\mathfrak{y} : I \to \mathfrak{C}$, denoted by $\gamma(t) = [x(t), r(t)]$, where $r : I \to [0, \infty)$ and $x : I \to X$ defined as $r(t) := \mathsf{r}(\gamma(t))$ and $x(t) := \mathsf{x}(\gamma(t))$.

If $t \mapsto \gamma(t)$ is a continuous curve in $\mathfrak{C}$, then so is r in $[0, \infty)$ and x in $O_r := \{t \in I : r(t) \in (0, \infty)\}$. Thus, according to [LMS18, Section 8], any continuous curve $\gamma : I \to \mathfrak{C}$ can be lifted to $y = \mathsf{y} \circ \gamma = (x, r) : I \to \mathcal{Y}$.

3.3.1 Dynamic Interpretation for the Cone Metric

Aligned with the discussions in Subsection 1.3.3, the cone metric $d_{\mathfrak{C}}$ can be interpreted as the infimum arc length of absolutely continuous curves $(\gamma_t)_{t \in I}$ between $\mathfrak{y}_0$ and $\mathfrak{y}_1$,

$$d_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \inf \left\{ \int_0^1 |\gamma'|_{d_{\mathfrak{C}}}(t)\, dt : \gamma \in AC^2(I; (\mathfrak{C}, d_{\mathfrak{C}})),\ \gamma(i) = \mathfrak{y}_i,\ i = 0, 1 \right\}, \tag{3.25}$$

where $|\gamma'|_{d_{\mathfrak{C}}}(t)$ is the metric derivative,

$$|\gamma'|_{d_{\mathfrak{C}}}(t) = \lim_{h \to 0} \frac{d_{\mathfrak{C}}(\gamma(t+h), \gamma(t))}{|h|}. \tag{3.26}$$

Let $\mathcal{Y} := X \times \mathbb{R}_+$. [LMS18, Section 8] introduced a set of continuous curves such that

$$\widetilde{C}(I; \mathcal{Y}) := \left\{ y = (x, r) : r \in C(I; \mathbb{R}_+),\ x\big|_{O_r} \in C(O_r; X) \right\}, \tag{3.27}$$

and a set of absolutely continuous curves such that for $p \geqslant 1$,

$$\widetilde{AC}^p(I; \mathcal{Y}) := \left\{ y = (x, r) : r \in AC^p(I; \mathbb{R}_+),\ x\big|_{O_r} \in AC^p_{loc}(O_r; X),\ r|\dot{x}| \in L^p(O_r) \right\}. \tag{3.28}$$

If $y = (x, r) \in \widetilde{AC}^p(I; \mathcal{Y})$, we define a Borel vector field $y' : I \to \mathbb{R}^{d+1}$ by

$$y'(t) := \begin{cases} \big(r(t)\dot{x}(t), \dot{r}(t)\big) & \text{if } r > 0, \dot{x}, \dot{r} \text{ exists}, \\ (0, 0) & \text{else }, \end{cases} \tag{3.29}$$

and the Borel map $|y'| : I \to \mathbb{R}^+$ by

$$|y'|^2 := \begin{cases} r^2|\dot{x}|_d^2 + |\dot{r}|^2 & \text{if } r > 0, \dot{x}, \dot{r} \text{ exists}, \\ 0 & \text{else}. \end{cases} \tag{3.30}$$

Lemma 23 ([LMS18, Section 8])**.** Let $\gamma \in C(I;\mathfrak{C})$ be lifted to $y = \mathsf{y} \circ \gamma \in \widetilde{C}(I;\mathcal{Y})$. Then $\gamma = [x,r] \in AC^p(I;(\mathfrak{C}, d_{\mathfrak{C}}))$ if and only if $y = (x,r) \in \widetilde{AC}^p(I;\mathcal{Y})$ and

$$|\gamma'|_{d_{\mathfrak{C}}}(t) = |y'|(t), \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in I. \tag{3.31}$$

Corollary 24. The cone metric in (3.25) is equivalent to

$$d_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \inf_{\Gamma(\mathfrak{y}_0,\mathfrak{y}_1)} \left\{ \int_0^1 |y'|(t)\,dt \right\}, \tag{3.32}$$

where the infimum is taken over all curves $(x,r) \in \Gamma(\mathfrak{y}_0, \mathfrak{y}_1)$ defined as

$$\Gamma(\mathfrak{y}_0, \mathfrak{y}_1) := \left\{ y = (x,r) \in \widetilde{AC}^2(I;\mathcal{Y}),\ [x(i), r(i)] = \mathfrak{y}_i,\ i = 0,1 \right\}. \tag{3.33}$$

By Jensen's inequality, it is easy to verify that

$$d^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \inf_{\Gamma(\mathfrak{y}_0,\mathfrak{y}_1)} \left\{ \int_0^1 |y'|^2(t)\,dt \right\}. \tag{3.34}$$

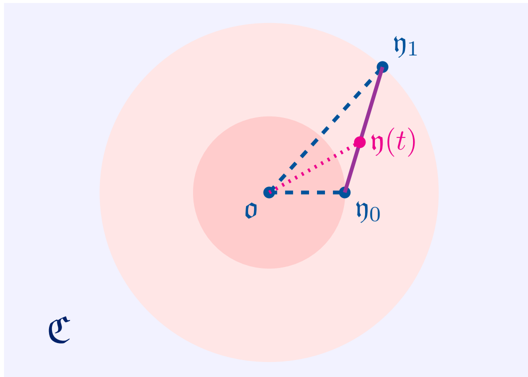
3.3.2 Optimal Curves in Cone Space

As per the definition of $d_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$, the optimal curve $([x,r])_{t\in I}$ with the shortest length between $\mathfrak{y}_0 = [x_0, r_0]$ and $\mathfrak{y}_1 = [x_1, r_1]$ is a straight line in a cone space. We are interested in the explicit formulation of the optimal curves for $d_{\mathfrak{C}}$ given the initial and the target values. Specifically, for points where $d(x_0, x_1) \leqslant \frac{\pi}{2}$, it is possible to identify a curve $\gamma = [x,r] \in AC^2(I;\mathfrak{C})$ such that the infimum is realized, that is

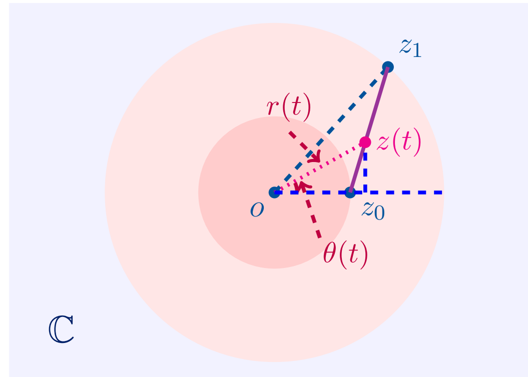
$$d^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \int_0^1 |\gamma'|^2_{d_{\mathfrak{C}}}(t)\,dt. \tag{3.35}$$

By absolute continuity, the condition for the infimum to be achieved is when $|\gamma'|_{d_{\mathfrak{C}}}(t)$ remains constant. Consequently, the optimal curve $([x,r])_{t\in I}$ is a geodesic in this context.

For simplicity in notation, we will use $\theta(t) := d(x_0, x(t))$ and $\theta_1 = d(x_0, x_1)$. According to [LMS18, Section 8], to find $x(t)$ and $r(t)$, we use the complex plane $\mathbb{C}$ and write the curve $z(t) = r(t)\exp(i\theta(t))$ connecting $z_0 = r_0 \in \mathbb{C}$ and $z_1 = r_1\exp(i\theta_1) \in \mathbb{C}$ in polar coordinates and $x(t) = x_0 + \frac{x_1 - x_0}{\theta_1} \cdot \theta(t)$.



Cone Space



Complex Space in Polar Coordinates

FIGURE 3.2: Correspondence Between the Two Spaces

The figure in 3.2 shows the correspondence between the cone space and the complex space in polar coordinates. Using the linear interpolation, we obtain

$$r(t)^2 = \big[\big((1-t)r_0 + tr_1\cos(\theta_1)\big]^2 + \big[tr_1\sin(\theta_1)\big]^2$$
$$= (1-t)^2 r_0^2 + t^2 r_1^2 + 2t(1-t)r_0 r_1\cos(\theta_1) \tag{3.36}$$
$$\cos(\theta(t)) = \frac{(1-t)r_0 + tr_1\cos(\theta_1)}{r(t)}. \tag{3.37}$$

We can verify the equality $d^2_{\mathfrak{C}}(\mathfrak{y}_0,\mathfrak{y}_1) = \int_0^1 \big(r^2|\dot{x}|^2 + |\dot{r}|^2\big)\,dt$ by computing

$$r^2|\dot{x}|^2 + |\dot{r}|^2 = \frac{\big(\dot{r}\cos(\theta) - c\big)^2}{\sin^2(\theta)} + |\dot{r}|^2 \quad \text{with } c := r_1\cos(\theta_1) - r_0$$
$$= (r_1\sin(\theta_1))^2 + c^2 = r_0^2 + r_1^2 - 2r_0 r_1\cos(\theta_1). \tag{3.38}$$

Definition 3.39. Given $\mathfrak{y}_0, \mathfrak{y}_1 \in \mathfrak{C}$, define a function $\Sigma : \mathfrak{C} \times \mathfrak{C} \to C(I;\mathfrak{C})$ mapping the cone points $\mathfrak{y}_0$ and $\mathfrak{y}_1$ to an optimal and unique time-dependent curve between them, as

$$\Sigma(\mathfrak{y}_0,\mathfrak{y}_1) := (\bar{\gamma}_t)_{t\in I} = ([\bar{x}(t), \bar{r}(t)])_{t\in I}, \tag{3.40}$$

where $\bar{x}(t) = x_0 + \frac{x_1 - x_0}{d(x_0,x_1)} \cdot \bar{\theta}(t)$ such that $\bar{r}(t), \bar{\theta}(t)$ are provided in (3.36).

Lemma 25. Given $\mathfrak{y}_0, \mathfrak{y}_1 \in \mathfrak{C}$, there exists a unique optimal curve $\bar{\gamma} = \Sigma(\mathfrak{y}_0,\mathfrak{y}_1) = [\bar{x}, \bar{r}] \in$

$C(I;\mathfrak{C})$ such that $[\bar{x}(0), \bar{r}(0)] = \mathfrak{y}_0$ and $[\bar{x}(1), \bar{r}(1)] = \mathfrak{y}_1$, and

$$d^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \int_0^1 |\bar{\gamma}'|^2_{d_{\mathfrak{C}}}(t)\, dt.$$

3.3.3 Dynamic Formulation of HK problem

Similar to the approach outlined in Subsection 2.5, we can also dynamically interpret the Hellinger-Kantorovich problem in (3.18). Assume that $(\mathfrak{C}, \tau)$ is Polish.

Definition 3.41. The **Dynamic Formulation** of the Wasserstein distance $W_{d_{\mathfrak{C}}}$ in $\mathcal{M}(\mathfrak{C})$ between $\nu_1, \nu_2 \in \mathcal{M}(\mathfrak{C})$ is defined as

$$Dyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2) := \min\left\{ \int_{C(I;\mathfrak{C})} \mathcal{A}_2(\gamma)\, d\eta(\gamma) : \eta \in \mathcal{P}(C(I;\mathfrak{C})),\ (e_0)_{\#}\eta = \nu_1,\ (e_1)_{\#}\eta = \nu_2 \right\}, \tag{3.42}$$

where $\mathcal{A}_2 : C(I;\mathfrak{C}) \to [0, \infty]$ is the action of a curve defined as

$$\mathcal{A}_2(\gamma) := \begin{cases} \int_0^1 |\gamma'|^2_{d_{\mathfrak{C}}}(t)\, dt & \text{if } \gamma \in AC(I; (\mathfrak{C}, d_{\mathfrak{C}})), \\ +\infty & \text{else.} \end{cases} \tag{3.43}$$

Denote $OptDyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$ by the set of optimal dynamic plans for $Dyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$.

Theorem 26. If (X, d) is a Polish and geodesic metric space, then

$$Dyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2) = W^2_{d_{\mathfrak{C}}}(\nu_1, \nu_2), \quad \text{where } W_{d_{\mathfrak{C}}} \text{ is defined in (3.9).} \tag{3.44}$$

Moreover, $\eta \in OptDyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$ if and only if $(e_0, e_1)_{\#}\eta \in Opt_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$ and η is supported in a set

$$\left\{ \gamma \in C(I;\mathfrak{C}) : \gamma = \Sigma(\gamma(0), \gamma(1)),\ \text{where } \Sigma \text{ defined in (3.40)} \right\}, \tag{3.45}$$

Proof. If α is an optimal transport plan of $W^2_{d_{\mathfrak{C}}}(\nu_1, \nu_2)$, then we can choose $\eta := \Sigma_{\#}\alpha \in \mathcal{P}(C(I;\mathfrak{C}))$ such that η is concentrated on a set defined in (3.45) and

$$\int \mathcal{A}_2(\gamma)\, d\eta(\gamma) = \int d^2_{\mathfrak{C}}(\gamma(0), \gamma(1))\, d\eta(\gamma) = \int_{\mathfrak{C}\times\mathfrak{C}} d^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d\alpha = W^2_{d_{\mathfrak{C}}}(\nu_1, \nu_2).$$

On the other hand, if η is an optimal dynamic plan of $Dyn_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$, let $\alpha = (e_0, e_1)_{\#}\eta$. Then α satisfies the marginal constraints $p^i_{\#}\alpha = \nu_i$ and

$$\int_{C(I;\mathfrak{C})} \mathcal{A}_2(\gamma)\, d\eta(\gamma) \geqslant \int_{C(I;\mathfrak{C})} d^2_{\mathfrak{C}}(\gamma(0), \gamma(1))\, d\eta(\gamma)$$
$$= \int_{\mathfrak{C}\times\mathfrak{C}} d^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d(e_0, e_1)_{\#}\eta \geqslant W^2_{d_{\mathfrak{C}}}(\nu_1, \nu_2).$$

The first inequality is an equality if and only if η is supported on a set in (3.45) and the second inequality is an equality if and only if $(e_0, e_1)_{\#}\eta \in Opt_{W_{d_{\mathfrak{C}}}}(\nu_1, \nu_2)$. □

Definition 3.46. With the equality in Corollary 22, we can generalize the **Dynamic Formulation** of the Hellinger-Kantorvich distance $\mathbf{HK}$ between $\mu_1, \mu_2 \in \mathcal{M}(X)$ as

$$Dyn_{\mathbf{HK}}(\mu_1, \mu_2) := \min\left\{ \int_{C(I;\mathfrak{C})} \mathcal{A}_2(\gamma)\, d\eta(\gamma) : \eta \in \mathcal{P}(C(I;\mathfrak{C})), \right.$$
$$\left. \mathfrak{h}^2 \circ (e_0)_{\#}\eta = \mu_1,\ \mathfrak{h}^2 \circ (e_1)_{\#}\eta = \mu_2 \right\}. \quad (3.47)$$

Denote $OptDyn_{\mathbf{HK}}(\mu_1, \mu_2)$ by the set of optimal dynamic plans for $Dyn_{\mathbf{HK}}(\mu_1, \mu_2)$.

Corollary 27. If (X, d) is a Polish and geodesic metric space, then

$$Dyn_{\mathbf{HK}}(\mu_1, \mu_2) = \mathbf{HK}^2(\mu_1, \mu_2), \quad \text{where } \mathbf{HK} \text{ is defined in (3.18)}. \quad (3.48)$$

Moreover, $\eta \in OptDyn_{\mathbf{HK}}(\mu_1, \mu_2)$ if and only if $(e_0, e_1)_{\#}\eta \in Opt_{\mathbf{HK}}(\mu_1, \mu_2)$ and η is supported on a set in (3.45).

3.4 Absolutely Continuous Curves in a Radon Space

Similar to Section 2.6, this section aims to characterize the evolution of a time-dependent curve $(\mu_t)_{t\in I} \in C(I; \mathcal{M}(X))$ that links two measures, μ_0 and μ_1, within $\mathcal{M}(X)$. The exploration by [LMS18] extends to examining the lifting relationship between absolutely continuous curves in a Radon space, where the metric is defined by the Hellinger-Kantorovich distance, and a dynamic plan associated with absolutely continuous curves within the metric space $(\mathfrak{C}, d_{\mathfrak{C}})$.

[LMS18] expanded upon Theorem 12, by demonstrating that any absolutely continuous curve $(\mu_t)_{t\in I}$ in $\mathcal{M}(X)$, metricised by $\mathbf{HK}$, can be written via a dynamic plan η as $\mu_t = \mathfrak{h}_t^2\eta := \mathfrak{h}^2 \circ (e_t)_{\#}\eta$, where $e_t : C(I;\mathfrak{C}) \to \mathfrak{C}$ is an evaluation map defined as $e_t(\gamma) := \gamma(t) = [x(t), r(t)]$.

Theorem 28 ([LMS18, Section 8.2])**.** Let (X, d) be a complete and separable metric space. Given a curve $\mu \in AC^p(I; (\mathcal{M}(X), \mathbf{HK}))$, $p \in [1, \infty]$ with

$$\Theta := \sqrt{\mu_0(X)} + \int_0^1 |\mu'|_{\mathbf{HK}}(t)\, dt. \tag{3.49}$$

Then there exists a curve $\nu \in AC^p(I; (\mathcal{P}_2(\mathfrak{C}), W_{d_\mathfrak{C}}))$ such that

1. ν_t is concentrated on $\mathfrak{C}[\Theta]$ for every $t \in I$.
2. $\mu_t = \mathfrak{h}^2\nu_t$ in I.
3. $|\mu'|_{\mathbf{HK}}(t) = |\alpha'|_{W_{d_\mathfrak{C}}}(t)$ for a.e. $t \in (0, 1)$.

Moreover, when $p = 2$, there exists a dynamic plan $\eta \in \mathcal{P}(AC^2(I;\mathfrak{C}))$ such that

1. $\mu_t = \mathfrak{h}_t^2\eta = \mathfrak{h}^2\nu_t$ in I where $\nu_t = (e_t)_{\#}\eta$.
2. $|\mu'|^2_{\mathbf{HK}}(t) = |\nu'|^2_{W_{d_\mathfrak{C}}}(t) = \int_{C(I;\mathfrak{C})} |\gamma'|^2_{d_\mathfrak{C}}(t)\, d\eta(\gamma)$ for $\mathcal{L}^1$-a.e. $t \in (0, 1)$.

Building upon this foundation, [LMS18] extended these findings to the geodesic of $\mathcal{M}(X)$.

Theorem 29 (Geodesic in $(\mathcal{M}(X), \mathbf{HK})$, [LMS18])**.** For every $\mu_0, \mu_1 \in \mathcal{M}(X)$,

(i) If $(\mu_t)_{t\in I}$ is a geodesic in $(\mathcal{M}(X), \mathbf{HK})$, then there exists an optimal geodesic plan $\eta \in \mathcal{P}(Geo(\mathfrak{C}))$ such that

  (a) η-a.e. curve $\mathfrak{y}$ is a geodesic in $\mathfrak{C}$.

  (b) let $\nu_t := (e_t)_{\#}\eta$, then $t \mapsto \nu_t : I \to \mathcal{P}_2(\mathfrak{C})$ is a geodesic in $(\mathcal{P}_2(\mathfrak{C}), W_{d_\mathfrak{C}})$ and ν_t is concentrated on $\mathfrak{C}[\Theta]$ with $\Theta^2 = 2(\mu_0(X) + \mathbf{HK}^2(\mu_0, \mu_1))$.

  (c) $\mu_t = \mathfrak{h}_t^2\eta = \mathfrak{h}^2\nu_t$ in I.

  (d) $(e_s, e_t)_{\#}\eta \in Opt_{\mathbf{HK}}(\mu_s, \mu_t)$ if $0 \leqslant s < t \leqslant 1$.

(ii) If (X, d) is a geodesic space, for every $\alpha \in Opt_{\mathbf{HK}}(\mu_0, \mu_1)$, there exists an optimal geodesic plan $\eta \in \mathcal{P}(Geo(\mathfrak{C}))$ such that $(e_0, e_1)_{\#}\eta = \alpha$.

3.5 Generalized Benamou-Brenier Problem

In Section 2.7, a continuum mechanics framework for the Wasserstein distance W_2 in $\mathcal{P}_2(\mathbb{R}^d)$ was established by Benamou and Brenier. This framework can be extended to accommodate the Hellinger-Kantorovich distance $\mathbf{HK}$ in $\mathcal{M}(\mathbb{R}^d)$. To address mass variation during transport, we introduce a scalar field $w(t,x)$, representing the rate of growth, which facilitates mass creation or annihilation. Coupled with a vector field $v(t,x)$, which drives the mass movement, the mass measure $\mu(t,x)$ evolves according to a continuity equation with the reaction term in (1.14). Through this transport equation, the Hellinger-Kantorovich distance can be interpreted as the minimal energy required for transportation.

Theorem 30 (Generalized Benamou-Brenier Problem, [LMS18, Section 8.5])**.** For any $\mu_0, \mu_1 \in \mathcal{M}(\mathbb{R}^d)$, the Hellinger-Kantorovich problem in (3.18) is equivalent to

$$\mathbf{HK}^2(\mu_0,\mu_1) = \min\left\{ \int_0^1 \int_{\mathbb{R}^d} \left(|v(t,x)|^2 + \frac{1}{4}|w(t,x)|^2 \right) \mu(t,dx)\, dt \right\}, \tag{3.50}$$

where the minimization is taken over a set

$$\left\{ \mu \in C(I;\mathcal{M}(\mathbb{R}^d)) : \partial_t \mu + \nabla\cdot(v\mu) = w\mu \text{ in } \mathscr{D}'((0,1)\times\mathbb{R}^d), \mu_{t=i} = \mu_i, i = 0,1 \right\}. \tag{3.51}$$

Similar to the Section 2.7, [LMS18] established a correspondence between absolutely continuous curves in $\mathcal{M}_2(\mathbb{R}^d)$ and solutions to the continuity equation with a reaction term in (1.14). Fix the time interval $I = [0,1]$.

Theorem 31 ([LMS18, Section 8.5])**.** Given $\mu \in AC^2(I;(\mathcal{M}(\mathbb{R}^d),\mathbf{HK}))$, there exists a velocity field $(v_t)_{t\in I}$ and a scalar field $(w_t)_{t\in I}$ such that (μ, v, w) solves the continuity equation in (1.14) and

$$v_t, w_t \in L^2(\mu_t) \quad \text{and} \quad \int_{\mathbb{R}^d} \left(|v_t|^2 + \frac{1}{4}|w_t|^2 \right) d\mu_t \leqslant |\mu'|^2_{\mathbf{HK}}(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{3.52}$$

Theorem 32 ([LMS18, Section 8.5])**.** If $\mu \in C(I;\mathcal{M}(\mathbb{R}^d))$ is a solution to the continuity equation with a reaction term in (1.14) induced by a velocity field $(v_t)_{t\in I}$ and a scalar field

$(w_t)_{t\in I}$ and assume that $v_t, w_t \in L^2(\mu_t)$ for $\mathcal{L}^1$-a.e $t \in (0,1)$, then $\mu \in AC^2(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{HK}))$ and

$$|\mu'|^2_{\mathbf{HK}}(t) \leqslant \int_{\mathbb{R}^d} \left(|v_t|^2 + \frac{1}{4}|w_t|^2 \right) d\mu_t \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{3.53}$$

4. Modified Hellinger-Kantorovich Problem

Shown in Theorem 30, the Hellinger-Kantorovich problem has a characterization of the Benamou-Brenier problem, which aims to determine the least amount of total energy required to transport an initial distribution to a target distribution. This energy consists of two main components: the kinetic energy from the movement of the mass and the energy associated with changes in the mass itself. During transportation, the change in mass μ, velocity v, and rate of growth w are governed by a partial differential equation known as continuity equation in (1.14), given by $\partial_t \mu + \nabla \cdot (v\mu) = w\mu$. Given this framework, the Hellinger-Kantorovich (HK) problem is formulated as in (3.50).

Let us introduce a modified variant of the HK problem, denoted as $\mathbf{mHK}^2$, which incorporates a non-negative constraint on the rate of growth denoted by $w \geqslant 0$. This constraint enforces a non-decreasing flow of mass, meaning that while the mass can either remain constant or increase during transportation, it is not allowed to decrease at any point in the process. This constraint provides a focused examination of transportation scenarios where only mass accumulation or conservation is permitted.

Definition 4.1. If $\mu_0(\mathbb{R}^d) \leqslant \mu_1(\mathbb{R}^d) < \infty$, we define the modified Hellinger Kantorovich problem as

$$\mathbf{mHK}^2(\mu_0, \mu_1) := \inf_{\mathcal{CE}^+(\mu_0,\mu_1)} \left\{ \int_0^1 \int_{\mathbb{R}^d} \left(|v(t,x)|^2 + \frac{1}{4}|w(t,x)|^2 \right) \mu(t,dx)dt \right\}, \tag{4.2}$$

where the infimum is taken over a set $\mathcal{CE}^+ := \mathcal{CE}^+_{[0,1]}$ and for any $T > 0$,

$$\begin{aligned} \mathcal{CE}^+_{[0,T]}(\mu_0, \mu_1) := \Big\{ &(\mu, v, w) : \mu \in C([0,T]; \mathcal{M}(\mathbb{R}^d)),\ v_t, w_t \in L^2(\mu_t), \\ &\partial_t \mu + \nabla \cdot (v\mu) = w\mu \text{ in } \mathscr{D}'((0,T) \times \mathbb{R}^d) \\ &w \geqslant 0 \text{ on } [0,T] \times \mathbb{R}^d,\ \mu_{t=0} = \mu_0,\ \mu_{t=T} = \mu_1 \Big\}. \end{aligned} \tag{4.3}$$

Else, if $\mu_0(\mathbb{R}^d) > \mu_1(\mathbb{R}^d)$, $\mathbf{mHK}(\mu_0, \mu_1) := +\infty$. Adding the constraint $w \geqslant 0$ makes $\mathbf{mHK}$ different from $\mathbf{HK}$ in (3.50).

Remark 4.4. Here, $\mu \in C([0,T];\mathcal{M}(\mathbb{R}^d))$ refers to that μ is continuous in the sense of weak topology induced by the duality on $\mathcal{M}(\mathbb{R}^d)$. Shown in Theorem (92), the **min** for the mHK problem is attained. Moreover, it is trivial that $\mathbf{mHK}^2 \geqslant \mathbf{HK}^2$ since the optimal solution for $\mathbf{mHK}^2$ can also serve as the solution for $\mathbf{HK}^2$.

4.1 Properties of the Modified HK Problem

Proposition 33. Suppose that there exists $R > 0$ such that $spt(\mu_0), spt(\mu_1) \subset B_R(0)$. Given $(\mu, v, w) \in \mathcal{CE}^+(\mu_0, \mu_1)$. With the imposition of the non-negative constraint $w \geqslant 0$, it is implied that the initial total mass cannot exceed the target total mass, i.e. $\mu_0(\mathbb{R}^d) \leqslant \mu_1(\mathbb{R}^d)$.

Proof. Let $\chi_R \in C_c^\infty(\mathbb{R}^d)$ be a smooth cut-off such that $0 \leqslant \chi_R \leqslant 1$, $|\nabla \chi_R| \leqslant 2$, and $\chi_R = 1$ on $B_R(0)$. By integration by parts, we obtain

$$\mu_1(\mathbb{R}^d) - \mu_0(\mathbb{R}^d) \geqslant \int_{\mathbb{R}^d} \chi_R \, d(\mu_1 - \mu_0) = \int_0^1 \left(\frac{d}{dt} \int_{\mathbb{R}^d} \chi_R \, d\mu_t \right) dt = \int_{\mathbb{R}^d} \chi_R w \, d\mu_t \geqslant 0.$$

□

Corollary 34. When $\mu_0(\mathbb{R}^d) = \mu_1(\mathbb{R}^d)$,

$$\mathbf{mHK}^2(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1), \tag{4.5}$$

where W_2 is the Wasserstein distance defined in (17). The equality holds since the rate of growth is 0, i.e. $w = 0$.

Definition 4.6. Given $T > 0$, define

$$\mathcal{J}_{[0,T]}(\mu, v, w) := T \int_0^T \left[\int_{\mathbb{R}^d} |v(t,x)|^2 + \frac{1}{4}|w(t,x)|^2 \, \mu(t, dx) \right] dt, \tag{4.7}$$

$$\tilde{\mathcal{J}}_{[0,T]}(\mu, v, w) := \left\{ \int_0^T \left[\int_{\mathbb{R}^d} |v(t,x)|^2 + \frac{1}{4}|w(t,x)|^2 \, \mu(t, dx) \right]^{\frac{1}{2}} dt \right\}^2. \tag{4.8}$$

We will use $\mathcal{J} := \mathcal{J}_{[0,1]}$.

Proposition 35.

$$\mathbf{mHK}^2(\mu_0,\mu_1) \overset{(1)}{=} \inf\left\{\mathcal{J}_{[0,T]}(\mu,v,w) : (\mu,v,w)\in\mathcal{CE}^+_{[0,T]}(\mu_0,\mu_1)\right\}, \tag{4.9}$$

$$\overset{(2)}{=} \inf\left\{\tilde{\mathcal{J}}_{[0,T]}(\mu,v,w) : (\mu,v,w)\in\mathcal{CE}^+_{[0,T]}(\mu_0,\mu_1)\right\}. \tag{4.10}$$

The first equality is the time-rescaling version.

Lemma 36. Fix $T_1, T_2 > 0$ and pick a triplet $(\tilde{\mu},\tilde{v},\tilde{w})\in\mathcal{CE}^+_{[0,T_1]}(\mu_0,\mu_1)$, we can construct a new triplet:

$$(\hat{\mu}_s,\hat{v}_s,\hat{w}_s) := (\tilde{\mu}_{ks}, k\tilde{v}_{ks}, k\tilde{w}_{ks}) \text{ with } k=\frac{T_2}{T_1} \tag{4.11}$$

such that $(\hat{\mu},\hat{v},\hat{w})\in\mathcal{CE}^+_{[0,T_2]}(\mu_0,\mu_1)$ and

$$\mathcal{J}_{[0,T_1]}(\tilde{\mu},\tilde{v},\tilde{w}) = \mathcal{J}_{[0,T_2]}(\hat{\mu},\hat{v},\hat{w}). \tag{4.12}$$

Proof of Proposition 35. Let us prove the first equality in (4.9). Pick a triplet $(\tilde{\mu},\tilde{v},\tilde{w})\in\mathcal{CE}^+_{[0,1]}(\mu_0,\mu_1)$ such that for any $\varepsilon>0$,

$$\mathbf{mHK}^2(\mu_0,\mu_1) \geqslant \mathcal{J}_{[0,1]}(\tilde{\mu},\tilde{v},\tilde{w}) - \varepsilon,$$

then by Lemma 36, we can construct a solution $(\hat{\mu},\hat{v},\hat{w})\in\mathcal{CE}^+_{[0,T]}(\mu_0,\mu_1)$ such that

$$\mathbf{mHK}^2(\mu_0,\mu_1) \geqslant \mathcal{J}_{[0,T]}(\hat{\mu},\hat{v},\hat{w}) - \varepsilon.$$

The other direction is achieved by setting $T_1 = T$ and $T_2 = 1$ in Lemma 36.

To prove the second equality in (4.10), we generalize the methodology used for the proof of [DNS09, Theorem 5.4]. By Hölder's inequality, we have $\mathcal{J}_{[0,T]} \geqslant \tilde{\mathcal{J}}_{[0,T]}$. For the other direction, pick a triplet $(\tilde{\mu},\tilde{v},\tilde{w})\in\mathcal{CE}^+_{[0,T]}(\mu_0,\mu_1)$ and given $\delta\geqslant 0$, define

$$S_\delta(s) := \int_0^s \left[\delta + \int_{\mathbb{R}^d}\left(|\tilde{v}(t,x)|^2 + \frac{1}{4}|\tilde{w}(t,x)|^2\right)\tilde{\mu}(t,dx)\right]^{\frac{1}{2}} dt, \qquad \forall s\in[0,T],$$

then $S_\delta(s)>0$ for all $s,\delta>0$. Compute the derivative of S'_δ with respect to s,

$$S'_\delta(s) = \left[\delta + \int_{\mathbb{R}^d}\left(|\tilde{v}(t,x)|^2 + \frac{1}{4}|\tilde{w}(t,x)|^2\right)\tilde{\mu}(t,dx)\right]^{\frac{1}{2}},$$

we have that $S'_\delta \geqslant S'_0 \geqslant 0$. Moreover, S_δ^{-1} exists and it is continuous since S_δ is increasing and continuous. Let $z = S_\delta(t)$, we have that $t = S_\delta^{-1}(z)$ and $dz = S'_\delta(t)\,dt$. We can construct a triplet $(\hat{\mu}, \hat{v}, \hat{w}) \in \mathcal{CE}^+_{[0,S_\delta(T)]}(\mu_0, \mu_1)$ by

$$(\hat{\mu}_z, \hat{v}_z, \hat{w}_z) := \Big(\tilde{\mu}_t,\ [S'_\delta(t)]^{-1}\tilde{v}_t,\ [S'_\delta(t)]^{-1}\tilde{w}_t\Big).$$

Substituting the triplet, we obtain that $\mathcal{J}_{[0,S_\delta(T)]}(\hat{\mu}, \hat{v}, \hat{w})$ is equivalent to

$$S_\delta(T)\int_0^T \left[\int_{\mathbb{R}^d}\left(\big|[S'_\delta(t)]^{-1}\tilde{v}(t,x)\big|^2 + \frac{1}{4}\big|[S'_\delta(t)]^{-1}\tilde{w}(t,x)\big|^2\right)\tilde{\mu}_t(t,dx)\right] S'_\delta(t)\,dt.$$

Simplifying the equation, we have

$$\begin{aligned}
&\mathcal{J}_{[0,S_\delta(T)]}(\hat{\mu}, \hat{v}, \hat{w})\\
&\qquad = S_\delta(T)\int_0^T \left[\int_{\mathbb{R}^d}\left(|\tilde{v}(t,x)|^2 + \frac{1}{4}|\tilde{w}(t,x)|^2\right)\tilde{\mu}_t(t,dx)\right][S'_\delta(t)]^{-1}\,dt\\
&\qquad \leqslant S_\delta(T)\int_0^T [S'_\delta(t)]^2\cdot[S'_\delta(t)]^{-1}\,dt\\
&\qquad = S_\delta(T)\int_0^T S'_\delta(t)\,dt = S_\delta(T)^2.
\end{aligned}$$

For any $\varepsilon > 0$, there exists a small $\delta > 0$ such that $S_\delta(T)^2 - S_0(T)^2 < \varepsilon$. Therefore,

$$\mathcal{J}_{[0,S_\delta(T)]}(\hat{\mu}, \hat{v}, \hat{w}) \leqslant S_\delta(T)^2 < S_0(T)^2 + \varepsilon = \tilde{\mathcal{J}}_{[0,T]}(\tilde{\mu}, \tilde{v}, \tilde{w}) + \varepsilon.$$

Due to the first equality in (4.9), we can conclude that

$$\begin{aligned}
\mathbf{mHK}^2(\mu_0,\mu_1) &= \inf\left\{\mathcal{J}_{[0,S_\varepsilon(T)]}(\mu,v,w) : (\mu,v,w)\in\mathcal{CE}^+_{[0,S_\varepsilon(T)]}(\mu_0,\mu_1)\right\}\\
&< \inf\left\{\tilde{\mathcal{J}}_{[0,T]}(\mu,v,w) : (\mu,v,w)\in\mathcal{CE}^+_{[0,T]}(\mu_0,\mu_1)\right\} + \varepsilon.
\end{aligned}$$

□

Proposition 37. **mHK** is a quasi-metric, i.e. **mHK** does not satisfy the symmetric property.

Proof. It is trivial that the identity property holds. Let us show the triangle inequality, i.e. for $\mu_0, \mu_1, \mu_2 \in \mathcal{M}(\mathbb{R}^d)$,

$$\mathbf{mHK}(\mu_0,\mu_2) \leqslant \mathbf{mHK}(\mu_0,\mu_1) + \mathbf{mHK}(\mu_1,\mu_2).$$

Except for when $\mu_0(\mathbb{R}^d) \leqslant \mu_1(\mathbb{R}^d) \leqslant \mu_2(\mathbb{R}^d)$, it is trivial to show since the right-hand side must be $+\infty$. Fix $T_1, T_2 > 0$. Given any $\varepsilon > 0$, let $(\tilde{\mu}^{01}, \tilde{v}^{01}, \tilde{w}^{01}) \in \mathcal{CE}^+_{[0,T_1]}(\mu_0, \mu_1)$ be a triplet such that

$$\sqrt{\tilde{\mathcal{J}}_{[0,T_1]}(\tilde{\mu}^{01}, \tilde{v}^{01}, \tilde{w}^{01})} - \mathbf{mHK}(\mu_0, \mu_1) < \varepsilon,$$

and $(\tilde{\mu}^{12}, \tilde{v}^{12}, \tilde{w}^{12}) \in \mathcal{CE}^+_{[0,T_2]}(\mu_1, \mu_2)$ be a triplet such that

$$\sqrt{\tilde{\mathcal{J}}_{[0,T_2]}(\tilde{\mu}^{12}, \tilde{v}^{12}, \tilde{w}^{12})} - \mathbf{mHK}(\mu_1, \mu_2) < \varepsilon.$$

Then we can construct a new triplet

$$(\mu_t^{02}, v_t^{02}, w_t^{02}) := \begin{cases} (\tilde{\mu}_t^{01}, \tilde{v}_t^{01}, \tilde{w}_t^{01}) & \text{if } t \in [0, T_1] \\ (\tilde{\mu}_{t-T_1}^{12}, \tilde{v}_{t-T_1}^{12}, \tilde{w}_{t-T_1}^{12}) & \text{if } t \in [T_1, T_1 + T_2] \end{cases}$$

such that $(\mu_t^{02}, v_t^{02}, w_t^{02}) \in \mathcal{CE}^+_{[0,T_1+T_2]}(\mu_0, \mu_2)$ and

$$\begin{aligned} \mathbf{mHK}(\mu_0, \mu_2) &\leqslant \sqrt{\tilde{J}_{[0,T_1+T_2]}(\mu^{02}, v^{02}, w^{02})} \\ &= \sqrt{\tilde{J}_{[0,T_1]}(\tilde{\mu}^{01}, \tilde{v}^{01}, \tilde{w}^{01})} + \sqrt{\tilde{J}_{[0,T_2]}(\tilde{\mu}^{12}, \tilde{v}^{12}, \tilde{w}^{12})} \\ &< \mathbf{mHK}(\mu_0, \mu_1) + \mathbf{mHK}(\mu_1, \mu_2) + 2\varepsilon. \end{aligned}$$

□

Definition 4.13.

$$\tilde{G}(\mu_0, \mu_1) := \begin{cases} \mathbf{mHK}(\mu_0, \mu_1) & \text{if } \mu_0(\mathbb{R}^d) \leqslant \mu_1(\mathbb{R}^d), \\ \mathbf{mHK}(\mu_1, \mu_0) & \text{otherwise.} \end{cases}$$

Proposition 38. $\tilde{G}$ is a semi-metric. It satisfies the identity and symmetry properties, but the triangle inequality

$$\tilde{G}(\mu_0, \mu_2) \leqslant \tilde{G}(\mu_0, \mu_1) + \tilde{G}(\mu_1, \mu_2)$$

might fail. A counterexample will be given in the following remark.

Remark 4.14. In the later Section 5.3.1, we can compute the $\mathbf{mHK}^2$ between the Dirac measures. Let $\mu_0 = 4\delta_{-1}$, $\mu_1 = \delta_0$, and $\mu_2 = 4\delta_1$.

$$\begin{aligned}\tilde{G}(\mu_0,\mu_2) = \mathbf{mHK}(\mu_0,\mu_2) = 4 > 2\cdot\sqrt{1+4-2\cdot 2\cdot\cos(1)} \\ = \mathbf{mHK}(\mu_0,\mu_1) + \mathbf{mHK}(\mu_1,\mu_2) \\ = \tilde{G}(\mu_1,\mu_0) + \tilde{G}(\mu_1,\mu_2).\end{aligned}$$

4.2 Benamou-Brenier Functional

The $\mathbf{mHK}^2$ problem poses challenges due to its non-convex nature with respect to the variables (μ, v, w). Furthermore, the continuity constraint, given by $\partial_t\mu + \nabla\cdot(v\mu) = w\mu$ introduces additional intricacies because of its non-linearity about the variables (μ, v, w). These properties make the development of numerical schemes more challenging. To address these issues, we drew inspiration and generalized ideas from previous works in [San15, Section 5.3.1] and [Chi+18a] for unbalanced transport.

Fix $\kappa > 0$ and let $K_\kappa := \left\{(a,b,c) \in \mathbb{R}\times\mathbb{R}^d\times\mathbb{R} : a + \frac{1}{4}(|b|^2 + \kappa^2c^2) \leqslant 0\right\}$. For $(x,y,z) \in \mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$, we can define

$$f_\kappa(x,y,z) := \sup_{(a,b,c)\in K_\kappa} (ax + b\cdot y + c\kappa^2 z). \tag{4.15}$$

Lemma 39. The function f_κ is equivalent to

$$f_\kappa(x,y,z) = \begin{cases} \frac{|y|^2+\kappa^2z^2}{x} & \text{if } x > 0, \\ 0 & \text{if } (x,|y|,z) = (0,0,0), \\ +\infty & \text{otherwise,} \end{cases} \tag{4.16}$$

and f_κ is 1-homogeneous, convex, and lower semicontinuous. Moreover, the subdifferential of f_κ is given by

$$\partial f_\kappa = \begin{cases} \left(-\frac{|y|^2+\kappa^2z^2}{x^2}, \frac{2y}{x}, \frac{2z}{x}\right) & \text{if } x > 0, \\ K & \text{if } (x,|y|,z) = (0,0,0), \\ \varnothing & \text{otherwise,} \end{cases}.$$

Definition 4.17 (Benamou-Brenier Functional)**.**

$$\mathcal{B}_\kappa(\mu,\omega,\zeta) := \sup_{(a,b,c)\in C_b([0,1]\times\mathbb{R}^d,K_\kappa)} \left\{ \int_0^1 \int_{\mathbb{R}^d} \Big(a(t,x)\,\mu(t,dx) + b(t,x)\cdot\omega(t,dx) + \kappa^2 c(t,x)\,\zeta(t,dx)\Big)\,dt \right\}. \tag{4.18}$$

Proposition 40. Suppose that $Z = [0,1]\times\mathbb{R}^d$. The functional $\mathcal{B}_\kappa$ is convex and lower semicontinuous on the space $\mathcal{M}(Z)\times\mathcal{M}^d(Z)\times\mathcal{M}(Z)$ for the weak convergence. Moreover, the following property holds:

1. $\mathcal{B}_\kappa \geqslant 0$.
2. The supremum in (4.18) can be taken over the space $L^\infty([0,1]\times\mathbb{R}^d;K_\kappa)$.
3. If $\mu,\omega,\zeta \ll \lambda$, then
$$\mathcal{B}_\kappa(\mu,\omega,\zeta) = \int_0^1\int_{\mathbb{R}^d} f_\kappa\left(\frac{\mu(t,dx)}{\lambda(t,dx)}, \frac{\omega(t,dx)}{\lambda(t,dx)}, \frac{\zeta(t,dx)}{\lambda(t,dx)}\right)\lambda(t,dx)dt.$$
4. $\mathcal{B}_\kappa < +\infty$ only if $\mu \geqslant 0$, $\omega \ll \mu$, and $\zeta \ll \mu$.
5. If $\mu \geqslant 0$, then
$$\mathcal{B}_\kappa(\mu, v\mu, \frac{1}{2}w\mu) = \int_0^1\int_{\mathbb{R}^d}\left(|v(t,x)|^2 + \frac{1}{4}w(t,x)^2\right)\mu(t,dx)dt.$$
6. Given mollifying kernels ρ_ε (e.g. $\rho_\varepsilon(x) = \frac{1}{\sqrt{(2\pi\varepsilon^d)}}e^{-\frac{|x|^2}{2\varepsilon}}$). If $\mu^\varepsilon = \mu*\rho_\varepsilon$, $\omega^\varepsilon = \omega*\rho_\varepsilon$, and $\zeta^\varepsilon = \zeta*\rho_\varepsilon$, then $\mathcal{B}_\kappa(\mu^\varepsilon,\omega^\varepsilon,\zeta^\varepsilon) \leqslant \mathcal{B}_\kappa(\mu,\omega,\zeta)$.

Proof. If we treat (ω,ζ) as a unity, the proof is the same as the proof of Proposition 5.18 in [San15]. □

Thanks to Proposition 40, we establish the convex formulation of $\mathbf{mHK}^2(\mu_0,\mu_1)$.

Corollary 41 (Convex Formulation of mHK Problem in (4.2))**.**

$$\mathbf{mHK}^2(\mu_0,\mu_1) = \inf_{\mathcal{CE}^+(\mu_0,\mu_1)} \mathcal{B}_1(\mu, v\mu, \frac{1}{2}w\mu) = \inf_{\widetilde{\mathcal{CE}}(\mu_0,\mu_1)} \mathcal{B}_1(\mu,\omega,\zeta), \tag{4.19}$$

where the constraint set $\widetilde{\mathcal{CE}}$ is defined as

$$\begin{aligned}\widetilde{\mathcal{CE}}(\mu_0,\mu_1) := \Big\{(\mu,\omega,\zeta) : \omega,\zeta \ll \mu \in C([0,1];\mathcal{M}(\mathbb{R}^d)),\\ \partial_t\mu + \nabla\cdot\omega = 2\zeta \text{ in } \mathscr{D}'((0,1)\times\mathbb{R}^d),\\ \zeta \geqslant 0 \text{ on } (t,x)\in[0,1]\times\mathbb{R}^d,\ \mu_{t=0}=\mu_0,\ \mu_{t=1}=\mu_1\Big\}.\end{aligned} \tag{4.20}$$

Remark 4.21. The Benamou-Brenier functional $\mathcal{B}_1$ is convex with respect to (μ,ω,ζ) respectively, and the continuity constraint $\partial_t\mu + \nabla\cdot\omega - 2\zeta = 0$ is linear. In Chapter 6, we will develop the algorithms to numerically solve (4.19).

5. A Variant of the mHK Problem within a Cone Setting

Recalling Corollary 22, the Hellinger Kantorovich problem is geometrically defined as a minimal balanced optimal transport problem within a cone space. This is achieved by lifting the measure from its regular space up to this cone space. Motivated by this framework, we introduce a new variant of the modified Hellinger Kantorovich, denoted as $\mathbf{mHK}^2_{\mathfrak{C}}$, which is formulated within this cone space. Furthermore, we establish that $\mathbf{mHK}^2$ is equivalent to $\mathbf{mHK}^2_{\mathfrak{C}}$. This equivalent formulation allows us to verify the existence of a minimum for the $\mathbf{mHK}^2$ problem. By leveraging this method, we successfully demonstrate that a minimizing solution for the problem is indeed attainable.

5.1 New Cone Quasi-Metric

Drawing upon the principles established for the dynamic formulation of the cone metric $d_{\mathfrak{C}}$, we can extend our approach to devise a novel distance, denoted as $\tilde{d}_{\mathfrak{C}}$, to measure the separation between two points within the cone space, but with added constraints. Specifically, we impose the constraint $\dot{r} \geqslant 0$ on any feasible curve (x, r) within $\Gamma(\mathfrak{y}_0, \mathfrak{y}_1)$ defined in (3.33), ensuring that the parameter $r(t)$ remains non-decreasing in $[0, 1]$. Furthermore, we solidify its static formulation by determining the optimal curve (x, r) that adheres to this constraint.

Let us introduce a new set of curves with a non-negative constraint,

$$\tilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1) := \Big\{ y = (x, r) \in \Gamma(\mathfrak{y}_0, \mathfrak{y}_1),\ \dot{r} \geqslant 0 \text{ a.e. if } \dot{r} \text{ exists} \Big\}. \tag{5.1}$$

Recall the definition of $\Gamma(\mathfrak{y}_0, \mathfrak{y}_1)$, we have $r \in AC^p(I; \mathbb{R}_+)$ so $\dot{r}$ exists, representing that r is non-decreasing within $[0, 1]$. Then we can define a new cone distance $\tilde{d}_{\mathfrak{C}}$ by that

Definition 5.2. Given $\mathfrak{y}_0 = [x_0, r_0] \in \mathfrak{C}$ and $\mathfrak{y}_1 = [x_1, r_1] \in \mathfrak{C}$. If $r_0 \leqslant r_1$, then

$$\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) := \inf_{\tilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)} \Big\{ \int_0^1 |y'|(t)\, dt \Big\}, \tag{5.3}$$

where $|y'|(t)$ is defined in (3.30), and if $r_0 > r_1$, then $\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) := +\infty$.

By Jensen's inequality, it is easy to verify that when $r_0 \leqslant r_1$,

$$\tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) := \inf_{\tilde{\Gamma}(\mathfrak{y}_0,\mathfrak{y}_1)} \left\{ \int_0^1 |y'|^2(t)\, dt \right\}. \tag{5.4}$$

Theorem 42. $\tilde{d}_{\mathfrak{C}}$ is a quasi-metric, i.e. $\tilde{d}_{\mathfrak{C}}$ does not satisfy the symmetric property.

Proof. It is obvious that the identity holds and symmetry does not hold. Now, let us prove that triangle inequality holds, i.e. for any $\mathfrak{y}_0, \mathfrak{y}_1, \mathfrak{y}_2 \in \mathfrak{C}$,

$$\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_2) \leqslant \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) + \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2).$$

Except for when $r_0 \leqslant r_1 \leqslant r_2$, it is trivial to show since the right-hand side must be ∞. Pick two curves $(x^{01}, r^{01}) \in \tilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)$ and $(x^{12}, r^{12}) \in \tilde{\Gamma}(\mathfrak{y}_1, \mathfrak{y}_2)$ such that for any $\varepsilon > 0$,

$$\int_0^1 |(r^{01}\dot{x}^{01}, \dot{r}^{01})|\, dt < \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) + \varepsilon \quad \text{and} \quad \int_0^1 |(r^{12}\dot{x}^{12}, \dot{r}^{12})|\, dt < \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) + \varepsilon,$$

then we can construct

$$\big[x^{02}(t), r^{02}(t)\big] := \begin{cases} \sqrt{2} \cdot \big[x^{01}(2t), r^{01}(2t)\big] & \text{for } t \in [0, \frac{1}{2}], \\ \sqrt{2} \cdot \big[x^{12}(2t-1), r^{12}(2t-1)\big] & \text{for } t \in (\frac{1}{2}, 1]. \end{cases}$$

such that $(x^{02}, r^{02}) \in \tilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_2)$. Therefore,

$$\begin{aligned} \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_2) &\leqslant \int_0^1 |(r^{02}\dot{x}^{02}, \dot{r}^{02})|(t)\, dt \\ &= 2\int_0^{\frac{1}{2}} |(r^{01}\dot{x}^{01}, \dot{r}^{01})|(2t)\, dt + 2\int_{\frac{1}{2}}^1 |(r^{12}\dot{x}^{12}, \dot{r}^{12})|(2t-1)\, dt \\ &= \int_0^1 |(r^{01}\dot{x}^{01}, \dot{r}^{01})|(t)\, dt + \int_0^1 |(r^{12}\dot{x}^{12}, \dot{r}^{12})|(t)\, dt \\ &< \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) + \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2) + 2\varepsilon. \end{aligned}$$

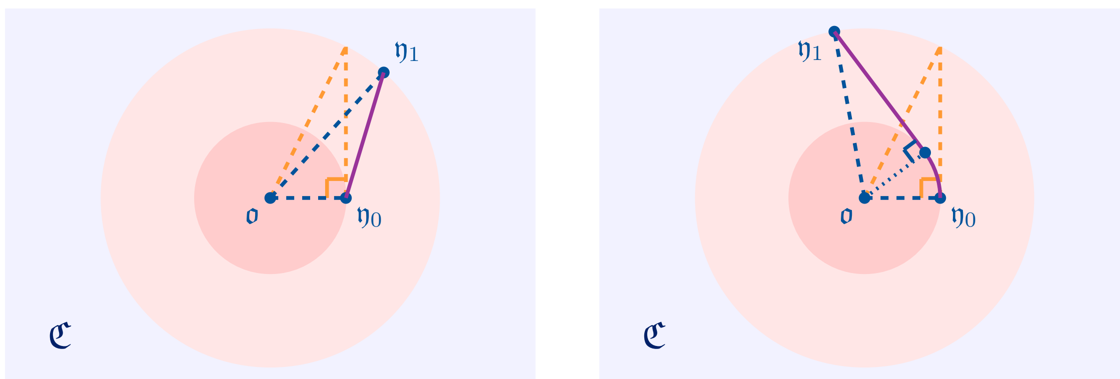
□

Corollary 43. For any $\mathfrak{y}_0, \mathfrak{y}_1, \mathfrak{y}_2 \in \mathfrak{C}$,

$$\min\big\{\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_2), \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_2, \mathfrak{y}_0)\big\} \leqslant \min\big\{\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1), \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_0)\big\} + \min\big\{\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2), \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_2, \mathfrak{y}_1)\big\}. \tag{5.5}$$

5.1.1 Static Formulation of the Modified Cone Metric

Inspired by the observation that the shortest path between $\mathfrak{y}_0 = [x_0, r_0]$ and $\mathfrak{y}_1 = [x_1, r_1]$ corresponding to the cone metric $d_{\mathfrak{C}}$ is a straight line, we seek to investigate the shortest path associated with our cone quasi-metric $\tilde{d}_{\mathfrak{C}}$. Visualized in Figure 5.1, our assumptions are that when $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) \leqslant \cos^{-1}(\frac{r_0}{r_1})$, the shortest path from $\mathfrak{y}_0$ to $\mathfrak{y}_1$ is indeed a straight line. Conversely, if $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$, the optimal path first traverses along the arc with initial radius r_0, and then continues directly along the tangent line toward $[x_1, r_1]$. To verify our assumptions, we will find a feasible curve $(x, r) \in \widetilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)$ and then show it is the only optimal curve between $\mathfrak{y}_0$ and $\mathfrak{y}_1$.



(a) when $d(x_0, x_1) \leqslant \cos^{-1}(\frac{r_0}{r_1})$ (b) when $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$

FIGURE 5.1: The new cone distance $\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$

At first, we hope to find an explicit formulation of a curve $(x, r) \in \widetilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)$, satisfying our assumption, visualized in Figure 5.2, that there is a value $\tilde{t} \in [0, 1]$ such that

- for $t \in [0, \tilde{t}]$, $\dot{r}(t) = 0$, or equivalently, $r(t) = r_0$, which means that it travels along the arc with initial radius r_0 within the time interval $[0, \tilde{t}]$.
- for $t \in [\tilde{t}, 1]$, $|y'|(t) = \sqrt{r^2(t)|\dot{x}(t)|^2 + |\dot{r}(t)|^2}$ is a constant, which means that it travels along a straight line within the time interval $[\tilde{t}, 1]$.

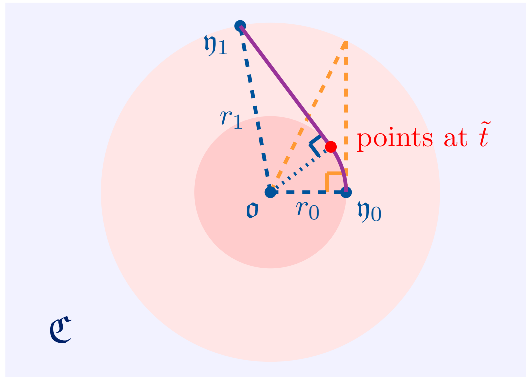
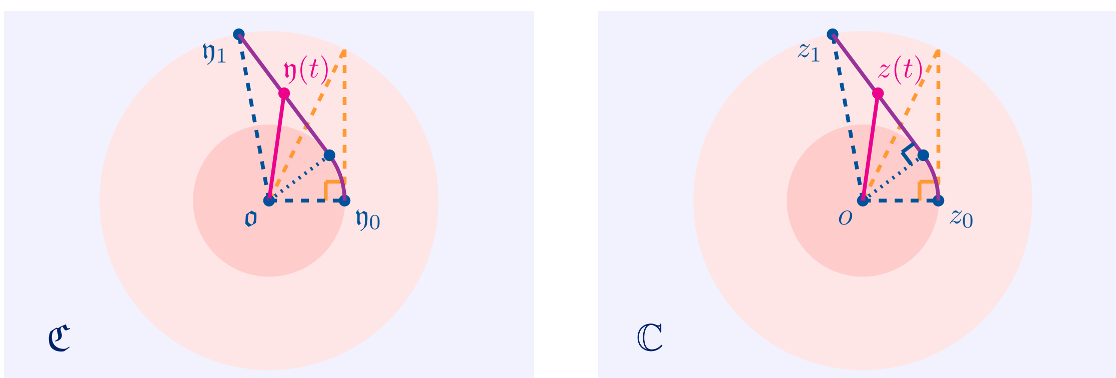


FIGURE 5.2: Curve Between $\mathfrak{y}_0$ and $\mathfrak{y}_1$ with Non-Decreasing Radius

To achieve that, we will use similar ideas discussed in Subsection 3.3.2. For simplicity in notation, let $\theta(t) := d(x_0, x(t))$ with $\theta_1 = d(x_0, x_1)$. To find $x(t)$ and $r(t)$, we use the complex plane $\mathbb{C}$ and write the curve $z(t) = r(t)\exp(i\theta(t))$ connecting $z_0 = r_0 \in \mathbb{C}$ and $z_1 = r_1\exp(i\theta_1) \in \mathbb{C}$ in polar coordinates and $x(t) = x_0 + \frac{x_1 - x_0}{\theta_1} \cdot \theta(t)$. Figure 5.3 shows the correspondence between the cone space and the complex space in polar coordinates.



Cone Space | Complex Space in Polar Coordinates

FIGURE 5.3: Correspondence Between the Two Spaces

Let $\tilde{\theta} = \theta_1 - \cos^{-1}(\frac{r_0}{r_1})$. Using linear interpolation, we obtain that

(i) When $t \in [0, \tilde{t}]$,

$$r(t) = r_0 \quad \text{and} \quad \theta(t) = t\tilde{t}^{-1}\tilde{\theta}. \tag{5.6}$$

(ii) When $t \in [\tilde{t}, 1]$, let $s = \frac{t-\tilde{t}}{1-\tilde{t}} \in [0,1]$, then

$$\begin{aligned} r(t)^2 &= \big[(1-s)r_0\cos(\tilde{\theta}) + sr_1\cos(\theta_1)\big]^2 + \big[(1-s)r_0\sin(\tilde{\theta}) + sr_1\sin(\theta_1)\big]^2 \\ &= (1-s)^2r_0^2 + s^2r_1^2 + 2s(1-s)r_0r_1\cos(\theta_1 - \tilde{\theta}) \\ &= (1-s)^2r_0^2 + s^2r_1^2 + 2s(1-s)r_0^2 \\ &= (1-s^2)r_0^2 + s^2r_1^2 \end{aligned} \tag{5.7}$$

$$\theta(t) = \tilde{\theta} + \cos^{-1}\left(\frac{r_0}{r(t)}\right). \tag{5.8}$$

If such $\tilde{t}$ exists, we can conclude that the construction of (x,r) satisfies that $(x,r) \in \widetilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)$ and

$$\tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \leqslant \int_0^1 |y'|^2(t)\,dt.$$

The next step is to find $\tilde{t} \in [0,1]$ such that the curve (x,r) obtained from above steps is the only curve minimizing $\tilde{d}(\mathfrak{y}_0, \mathfrak{y}_1)$, i.e.

$$\tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) = \int_0^1 |y'|^2(t)\,dt.$$

Lemma 44. Given $\mathfrak{y}_0 = [x_0, r_0] \in \mathfrak{C}$ and $\mathfrak{y}_1 = [x_1, r_1] \in \mathfrak{C}$. If $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$, then for any $(x,r) \in \widetilde{\Gamma}(\mathfrak{y}_0, \mathfrak{y}_1)$,

$$\int_0^1 |y'(t)|^2\,dt \geqslant \left[r_0\left(d(x_0,x_1) - \cos^{-1}\left(\frac{r_0}{r_1}\right)\right) + \sqrt{r_1^2 - r_0^2}\right]^2. \tag{5.9}$$

and the equality holds only when (x,r) is given by that $x(t) = x_0 + \frac{x_1 - x_0}{\theta_1} \cdot \theta(t)$ and

(i) for $t \in [0, \tilde{t}]$, $r(t)$ and $\theta(t)$ are given in (5.6).

(ii) for $t \in [\tilde{t}, 1]$, $r(t)$ and $\theta(t)$ are given in (5.7).

where

$$\tilde{t} = \frac{r_0\big(d(x_0,x_1) - \cos^{-1}(\frac{r_0}{r_1})\big)}{r_0\big(d(x_0,x_1) - \cos^{-1}(\frac{r_0}{r_1})\big) + \sqrt{r_1^2 - r_0^2}}. \tag{5.10}$$

Proof. Pick any $\tilde{t} \in [0,1]$ such that for $t \in [0,\tilde{t}]$, $\dot{r}(t) = 0$ and for $t \in [\tilde{t},1]$, $\dot{r}(t) \geqslant 0$. Denote

$$A = r_0 d(x_0, x(\tilde{t})) = r_0\left(d(x_0,x_1) - \cos^{-1}\left(\frac{r_0}{r_1}\right)\right) \quad \text{and} \quad B = \sqrt{r_1^2 - r_0^2}. \tag{5.11}$$

We can separate the integral into two parts,

$$\int_0^1 |y'|^2(t)\,dt = \left(\int_0^{\tilde{t}} + \int_{\tilde{t}}^1\right)|y'|^2(t)\,dt.$$

For the first integral, by Hölder's inequality and absolute continuity of x w.r.t. Euclidean norm, we have that

$$\int_0^{\tilde{t}} |y'|^2(t)\,dt = \int_0^{\tilde{t}} r_0^2|\dot{x}|_d^2(t)\,dt \geqslant r_0^2\left(\int_0^{\tilde{t}} 1^2 dt\right)^{-1} \cdot \left(\int_0^{\tilde{t}} |\dot{x}|_d(t)\,dt\right)^2$$
$$\geqslant (\tilde{t})^{-1} r_0^2 d(x_0, x(\tilde{t}))^2 = (\tilde{t})^{-1}A^2.$$

For the second integral, by Hölder's inequality and the equivalent formulation of cone metric $d_{\mathfrak{C}}$ in (3.32), we have

$$\int_{\tilde{t}}^1 |y'|^2(t)\,dt \geqslant \left(\int_{\tilde{t}}^1 1^2 dt\right)^{-1} \cdot \left(\int_{\tilde{t}}^1 |y'|(t)\,dt\right)^2$$
$$\geqslant (1-\tilde{t})^{-1} d_{\mathfrak{C}}^2(\mathfrak{y}(\tilde{t}), \mathfrak{y}_1) = (1-\tilde{t})^{-1}B^2.$$

Combining them, we obtain

$$\int_0^{\tilde{t}} |y'|^2(t)dt \geqslant (\tilde{t})^{-1}A^2 + (1-\tilde{t})^{-1}B^2 := f(\tilde{t}).$$

f attains the minimum at $\tilde{t}_*$ where

$$\tilde{t}_* = \frac{A}{A+B} \tag{5.12}$$

and

$$f(\tilde{t}_*) = (A+B)^2 = \left[r_0\big(\theta - \cos^{-1}(\frac{r_0}{r_1})\big) + \sqrt{r_1^2 - r_0^2}\right]^2. \tag{5.13}$$

The equality holds when

$$|y'|(t) = \begin{cases} (\tilde{t}_*)^{-1}A & \text{if } t \in [0,\tilde{t}] \\ (1-\tilde{t}_*)^{-1}B & \text{if } t \in [\tilde{t},1] \end{cases} = A + B. \tag{5.14}$$

□

Proposition 45 (Static Formulation of Modified Cone Metric)**.**

- If $r_0 > r_1$, $\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) := +\infty$.
- If $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) \leqslant \cos^{-1}(\frac{r_0}{r_1})$, then

$$\tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) = d_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) = r_0^2 + r_1^2 - 2r_0 r_1 \cos(d(x_0, x_1)). \tag{5.15}$$

- If $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$, then

$$\tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) = \left[r_0 \left(d(x_0, x_1) - \cos^{-1}\left(\frac{r_0}{r_1}\right)\right) + \sqrt{r_1^2 - r_0^2}\right]^2. \tag{5.16}$$

Proof. For some $\tilde{t} \in [0, 1]$, let $(\tilde{x}, \tilde{r})$ be given by that $\tilde{x}(t) = x_0 + \frac{x_1 - x_0}{\theta_1} \cdot \tilde{\theta}(t)$ and

(i) for $t \in [0, \tilde{t}]$, $\tilde{r}(t)$ and $\tilde{\theta}(t)$ are given in (5.6).

(ii) for $t \in [\tilde{t}, 1]$, $\tilde{r}(t)$ and $\tilde{\theta}(t)$ are given in (5.7).

If $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) \leqslant \cos^{-1}(\frac{r_0}{r_1})$, then

$$\int_0^1 |y'|^2(t)\, dt \geqslant d_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1),$$

and the lower bound is achieved by $(\tilde{x}, \tilde{r})$ when $\tilde{t} = 0$. If $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$, then by Lemma 44,

$$\int_0^1 |y'|^2(t)\, dt \geqslant \left[r_0 \left(d(x_0, x_1) - \cos^{-1}\left(\frac{r_0}{r_1}\right)\right) + \sqrt{r_1^2 - r_0^2}\right]^2, \tag{5.17}$$

and the lower bound is achieved by $(\tilde{x}, \tilde{r})$ when

$$\tilde{t} = \frac{r_0\big(d(x_0, x_1) - \cos^{-1}(\frac{r_0}{r_1})\big)}{r_0\big(d(x_0, x_1) - \cos^{-1}(\frac{r_0}{r_1})\big) + \sqrt{r_1^2 - r_0^2}}. \tag{5.18}$$

□

To connect the optimal curve $(x, r) \in \widetilde{AC}^2(I; \mathcal{Y})$ and $[x, r] \in AC^2(I; (\mathfrak{C}, \tilde{d}_{\mathfrak{C}}))$, Lemma 23 can be generalized as follows,

Corollary 46. Let $(\mathfrak{C}, \tilde{d}_{\mathfrak{C}})$ be a quasi-metric space and $\gamma \in C(I; \mathfrak{C})$ be lifted to $y = \mathsf{y} \circ \gamma \in \widetilde{C}(I; \mathcal{Y})$. Then $\gamma = [x, r] \in AC^2(I; (\mathfrak{C}, \tilde{d}_{\mathfrak{C}}))$ if and only if $y = (x, r) \in \widetilde{AC}^2(I; \mathcal{Y})$ where $\dot{r} \geqslant 0$ if $\dot{r}$ exists. Moreover,

$$|\gamma'|_{\tilde{d}_{\mathfrak{C}}}(t) = |y'|(t), \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in I. \tag{5.19}$$

Definition 5.20. Given $\mathfrak{y}_0, \mathfrak{y}_1 \in \mathfrak{C}$, define a function $\widetilde{\Sigma} : \mathfrak{C} \times \mathfrak{C} \to C(I; \mathfrak{C})$ mapping the cone points $\mathfrak{y}_0$ and $\mathfrak{y}_1$ to an optimal and unique time-dependent curve between them, as

$$\widetilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1) := (\bar{\gamma}(t))_{t\in I} = ([\bar{x}(t), \bar{r}(t)])_{t\in I}, \tag{5.21}$$

where $\bar{x}(t) = x_0 + \frac{x_1 - x_0}{d(x_0,x_1)} \cdot \bar{\theta}(t)$ such that $\bar{r}(t), \bar{\theta}(t)$ are provided in(5.6) and (5.7). Note, $\widetilde{\Sigma}(\mathfrak{o}, [y, r_1]) = ([y, tr_1])_{t\in I}$.

Lemma 47. Given $\mathfrak{y}_0, \mathfrak{y}_1 \in \mathfrak{C}$, there exists an unique optimal curve $\bar{y} = (\bar{x}, \bar{r}) \in \widetilde{AC}^2(I; \mathcal{Y})$ between $\mathfrak{y}_0$ and $\mathfrak{y}_1$, given in (5.6), (5.7), minimizing $\tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$. According to Corollary 46, $\bar{\gamma} = \mathsf{p} \circ \bar{y} = [\bar{x}, \bar{r}] \in AC^2(I; (\mathfrak{C}, d_{\mathfrak{C}}))$ is also the unique optimal curve of $\tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$ such that

$$\tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) = \int_0^1 |\bar{\gamma}'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, dt.$$

Proposition 48. $\tilde{d}^2_{\mathfrak{C}}([x_0, \cdot], [x_1, \cdot])$ is 2-homogeneous, i.e.

$$\tilde{d}^2_{\mathfrak{C}}([x_0, r_0], [x_1, r_1]) = \tilde{d}^2_{\mathfrak{C}}([x_0, \frac{r_0}{v}], [x_1, \frac{r_1}{v}]) \cdot v^2 \quad \forall v > 0. \tag{5.22}$$

Proof. It suffices to check when $r_0 \leqslant r_1 \neq 0$ and $d(x_0, x_1) > \cos^{-1}(\frac{r_0}{r_1})$.

$$\begin{aligned} \tilde{d}^2_{\mathfrak{C}}([x_0, \frac{r_0}{v}], [x_1, \frac{r_1}{v}]) \cdot v^2 &= \left[\frac{r_0}{v} \left(d(x_0, x_1) - \cos^{-1}\left(\frac{r_0}{r_1}\right) \right) + \sqrt{(\frac{r_1}{v})^2 - (\frac{r_0}{v})^2} \right]^2 \cdot v^2 \\ &= \tilde{d}^2_{\mathfrak{C}}([x_0, r_0], [x_1, r_1]). \end{aligned}$$

□

Here, if we treat $\tilde{d}^2$ as a cost function on $\mathfrak{C} \times \mathfrak{C}$, where the cone space $\mathfrak{C}$ is endowed with the cone metric $d_{\mathfrak{C}}$ defined in (3.1), then we can show that this cost function is lower semicontinuous w.r.t. the cone metric $d_{\mathfrak{C}}$.

Proposition 49. Let the cost function be $c(\mathfrak{y}_0, \mathfrak{y}_1) := \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$. Then c is lower semi-continuous w.r.t the cone metric $d_{\mathfrak{C}}$, i.e. show that

$$c(\mathfrak{y}'_0, \mathfrak{y}'_1) \leqslant \liminf_{\substack{\mathfrak{y}_0 \to \mathfrak{y}'_0 \\ \mathfrak{y}_1 \to \mathfrak{y}'_1}} c(\mathfrak{y}_0, \mathfrak{y}_1), \tag{5.23}$$

that is, for every $y < c(\mathfrak{y}'_0, \mathfrak{y}'_1)$, there exists a neighborhood U of $(\mathfrak{y}'_0, \mathfrak{y}'_1)$ such that $y < c(\mathfrak{y}_0, \mathfrak{y}_1)$ for all $(\mathfrak{y}_0, \mathfrak{y}_1) \in U$.

Proof. Given $\varepsilon > 0$, let

$$U_\varepsilon := \left\{ (\mathfrak{y}_0, \mathfrak{y}_1) \in \mathfrak{C} \times \mathfrak{C} : d_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i') < \varepsilon, i \in \{0,1\} \right\}$$

be a neighborhood of $(\mathfrak{y}_0', \mathfrak{y}_1')$. Fixing $y < c(\mathfrak{y}_0', \mathfrak{y}_1')$, the goal is to find ε such that $y < c(\mathfrak{y}_0, \mathfrak{y}_1)$ for all $(\mathfrak{y}_0, \mathfrak{y}_1) \in U_\varepsilon$. Observe from the definition of $d_{\mathfrak{C}}$ in (3.1),

$$d_{\mathfrak{C}}^2(\mathfrak{y}_i, \mathfrak{y}_i') = |r_i - r_i'|^2 + 4 r_i r_i' \sin^2(d_{\frac{\pi}{2}}(x_i, x_i')/2) < \varepsilon^2, \quad \forall i \in \{0,1\}$$

that we obtain $|r_i - r_i'| < \varepsilon$ and

$$4 r_i r_i' \sin^2(d_{\frac{\pi}{2}}(x_i, x_i')/2) < \varepsilon^2 \implies \sin(d_{\frac{\pi}{2}}(x_i, x_i')/2) < \frac{\varepsilon}{2|r_i' - \varepsilon|}.$$

When $\varepsilon < \frac{2r_i'}{3}$, we obtain

$$d(x_i, x_i') < \frac{\varepsilon}{r_i' - \varepsilon} < 3(r_i')^{-1}\varepsilon.$$

Then for any $(\mathfrak{y}_0, \mathfrak{y}_1) \in U_\varepsilon$, when $r_i \leqslant r_i'$,

(i) If $d(x_i, x_i') \leqslant \cos^{-1}(\frac{r_i}{r_i'})$, then

$$\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i') = d_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i') < \varepsilon.$$

(ii) If $d(x_i, x_i') > \cos^{-1}(\frac{r_i}{r_i'})$, then

$$\begin{aligned} \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i') &= r_i \left(d(x_i, x_i') - \cos^{-1}(\frac{r_i}{r_i'}) \right) + \sqrt{(r_i')^2 - r_i^2} \\ &< r_i' \cdot 3(r_i')^{-1}\varepsilon + \sqrt{(2r_i' - \varepsilon)\varepsilon} < 5\sqrt{r_i'\varepsilon}. \end{aligned}$$

In summary,

$$\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i') < 5\sqrt{r_i'\varepsilon}.$$

Similarly, when $r_i > r_i'$, we get

$$\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i', \mathfrak{y}_i) < 5\sqrt{r_i\varepsilon} < 5\sqrt{(r_i' + \varepsilon)\varepsilon} < 5\sqrt{2r_i'\varepsilon}.$$

Therefore,

$$\min\left\{ \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i'), \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i', \mathfrak{y}_i) \right\} < 10\sqrt{r_i'\varepsilon}.$$

Then

(i) If $r_0' > r_1'$, further choose $\varepsilon \leqslant \frac{r_0'-r_1'}{4}$, we obtain for any $(\mathfrak{y}_0, \mathfrak{y}_1) \in U_\varepsilon$,

$$r_0 - r_1 > (r_0' - \varepsilon) - (r_1' + \varepsilon) > r_0' - r_1' - 2\varepsilon > 0.$$

Therefore, $c(\mathfrak{y}_0', \mathfrak{y}_1') = \infty = c(\mathfrak{y}_0, \mathfrak{y}_1)$.

(ii) If $r_0' \leqslant r_1'$, by Corollary 43,

$$\begin{aligned}\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0', \mathfrak{y}_1') &\leqslant \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) + \sum_{i=0}^{1} \min\left\{\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i, \mathfrak{y}_i'), \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_i', \mathfrak{y}_i)\right\} \\ &\leqslant \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) + \sum_{i=0}^{1} 10\sqrt{r_i'\varepsilon}.\end{aligned}$$

Therefore, if we further choose

$$\varepsilon < \left[\left(10\sum_{i=0}^{1}\sqrt{r_i'}\right)^{-1} \cdot \left(\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0', \mathfrak{y}_1') - \sqrt{y}\right)\right]^2,$$

we have $\tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1) > \sqrt{y}$ so that $c(\mathfrak{y}_0, \mathfrak{y}_1) > y$.

□

5.2 A Variant of mHK Problem within a Cone Setting

Similar to the original construction of the Hellinger Kantorovich problem defined in (3.18), we also want to construct a modified variant on a cone space.

Definition 5.24 (Modified Hellinger Kantorovich Problem within a Cone Setting)**.** Given $\mu_1, \mu_2 \in \mathcal{M}_2(X)$,

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_1, \mu_2) := \inf\left\{\int_{\mathfrak{C}\times\mathfrak{C}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) : \alpha \in \mathcal{M}_2(\mathfrak{C}\times\mathfrak{C}),\ \mathfrak{h}_i^2\alpha = \mu_i,\ i = 1, 2\right\}. \tag{5.25}$$

We denote $Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2)$ by the set of all optimal plans α for $\mathbf{mHK}^2_{\mathfrak{C}}(\mu_1, \mu_2)$.

Proposition 50. Let $\boldsymbol{\mathfrak{C}} = \mathfrak{C}^{\otimes 2}$ and $\boldsymbol{\mathfrak{C}}[R] := \{\boldsymbol{\mathfrak{y}} = (\mathfrak{y}_1, \mathfrak{y}_2) \in \boldsymbol{\mathfrak{C}} : |\boldsymbol{\mathfrak{y}}|_2 \leqslant R\}$. Then

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_1, \mu_2) = \inf_{\alpha\in S(\mu_1,\mu_2)} \int_{\boldsymbol{\mathfrak{C}}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) := \mathcal{C}(\alpha), \tag{5.26}$$

where $S(\mu_1, \mu_2) := \left\{\alpha \in \mathcal{P}(\boldsymbol{\mathfrak{C}}) : \alpha(\boldsymbol{\mathfrak{C}}\backslash\boldsymbol{\mathfrak{C}}[R]) = 0,\ \mathfrak{h}_i^2\alpha = \mu_i,\ i = 1, 2\right\}$.

Proof of Proposition 50. By Remark 3.21, we can obtain a rescaled measure $\tilde{\alpha} := \mathrm{dil}_{\vartheta,2}\alpha$ such that $\tilde{\alpha} \in \mathcal{P}_2(\mathfrak{C})$, $\mathfrak{h}_i^2\tilde{\alpha} = \mathfrak{h}_i^2\alpha$, and $\tilde{\alpha}(\mathfrak{C}\backslash\mathfrak{C}[R]) = 0$. Since $\tilde{d}_{\mathfrak{C}}^2([x_1,\cdot],[x_2,\cdot])$ is 2-homogeneous, then

$$\begin{aligned}\int \tilde{d}_{\mathfrak{C}}^2([x_1,r_1],[x_2,r_2])\,d\tilde{\alpha} &= \int \tilde{d}_{\mathfrak{C}}^2(\left[x_1,\frac{r_1}{\vartheta(\mathfrak{y})}\right],\left[x_2,\frac{r_2}{\vartheta(\mathfrak{y})}\right])\cdot\vartheta(\mathfrak{y})^2\,d\alpha \\ &= \int \tilde{d}_{\mathfrak{C}}^2([x_1,r_1],[x_2,r_2])\,d\alpha.\end{aligned}$$

□

Theorem 51 (Existence of $\mathbf{mHK}_{\mathfrak{C}}^2$)**.** The $\mathbf{mHK}_{\mathfrak{C}}^2$ problem always admits an optimal solution $\alpha \in \mathcal{P}(\mathfrak{C})$ concentrated on $\mathfrak{C}[R]\backslash\{(\mathfrak{o},\mathfrak{o})\}$.

Remark 5.27. $\alpha \in \mathcal{P}(\mathfrak{C})$ concentrated on $\mathfrak{C}[R]\backslash\{(\mathfrak{o},\mathfrak{o})\}$ refers to the transport plan α assigns no mass transported from the vertex to the vertex.

Lemma 52 (Tightness, [LMS18, Lemma 7.3])**.** Let $\{\mathcal{K}_i\}_{i=1}^N$ be a finite collection of bounded and equally tight sets in $\mathcal{M}(X)$. Then the set

$$\left\{\alpha \in \mathcal{M}_2(\mathfrak{C}^{\otimes N}) : \mathfrak{h}_i^2\alpha \in \mathcal{K}_i,\ i = \{1,\dots,N\}\right\}$$

is equally tight in $\mathcal{M}(\mathfrak{C}^{\otimes N})$.

Proof of Theorem 51. For simplicity, say $S := S(\mu_1,\mu_2)$.

(i) S is compact w.r.t the weak topology since

  (a) S is relatively compact: since $\sup_{\alpha\in S}\alpha(\mathfrak{C}\times\mathfrak{C}) = 1 < \infty$, then S is bounded and by Lemma 52, S is equally tight in $\mathcal{P}(\mathfrak{C}^{\otimes 2})$.

  (b) S is closed w.r.t weak topology, i.e. the limit of any convergent sequence is also in S. Suppose $\alpha_n \in S \to \alpha_*$ weakly. Since $\mathfrak{C}[R]$ is closed, then α_* is also concentrated on $\mathfrak{C}[R]$. Moreover, α_* satisfies the homogenous marginals $\mathfrak{h}_i^2\alpha_* = \mu_i$. Since $\mathsf{x}_i, \mathsf{r}_i$ are continuous, then for any $\varphi \in C_b(X_i)$,

$$\int_{X_i}\varphi(x_i)\,d(\mathfrak{h}_i^2\alpha_*) = \int_{\mathfrak{C}[R]}\varphi(\mathsf{x}_i)\,\mathsf{r}_i^2\,d\alpha_* = \lim_{n\to\infty}\int_{\mathfrak{C}[R]}\varphi(\mathsf{x}_i)\mathsf{r}_i^2\,d\alpha_n = \int_{X_i}\varphi(x_i)\,d\mu_i.$$

Therefore, $\alpha_* \in S$.

(ii) the map $\alpha \mapsto \mathcal{C}(\alpha)$ is lower semi-continuous, i.e. if $\alpha_n \to \alpha_*$ weakly, then

$$\mathcal{C}(\alpha_*) \leqslant \liminf_{n\to\infty} \mathcal{C}(\alpha_n).$$

For $k \in \mathbb{N}$, let $c(\mathfrak{y}_0, \mathfrak{y}_1) := \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)$ and

$$c_k(\mathfrak{y}_0, \mathfrak{y}_1) := \inf_{\tilde{\mathfrak{y}}_0, \tilde{\mathfrak{y}}_1 \in \mathfrak{C}} \left\{ c(\tilde{\mathfrak{y}}_0, \tilde{\mathfrak{y}}_1) \wedge k + k \cdot d_{\mathfrak{C}}(\mathfrak{y}_0, \tilde{\mathfrak{y}}_0) + k \cdot d_{\mathfrak{C}}(\mathfrak{y}_1, \tilde{\mathfrak{y}}_1) \right\},$$

then $0 \leqslant c_k \leqslant c_{k+1} \leqslant c \wedge k$ and we can prove that $c_k \in Lip_b(\mathfrak{C}^{\otimes 2})$. Given $(\mathfrak{y}'_0, \mathfrak{y}'_1), (\mathfrak{y}''_0, \mathfrak{y}''_1) \in \mathfrak{C}^{\otimes 2}$. For any $\varepsilon > 0$, there exists $(\hat{\mathfrak{y}}_0, \hat{\mathfrak{y}}_1) \in \mathfrak{C}^{\otimes 2}$ such that

$$c_k(\mathfrak{y}'_0, \mathfrak{y}'_1) > c_k(\hat{\mathfrak{y}}_0, \hat{\mathfrak{y}}_1) + k \cdot d_{\mathfrak{C}}(\mathfrak{y}'_0, \hat{\mathfrak{y}}_0) + k \cdot d_{\mathfrak{C}}(\mathfrak{y}'_1, \hat{\mathfrak{y}}_1) - \varepsilon,$$

$$c_k(\mathfrak{y}''_0, \mathfrak{y}''_1) \leqslant c_k(\hat{\mathfrak{y}}_0, \hat{\mathfrak{y}}_1) + k \cdot d_{\mathfrak{C}}(\mathfrak{y}''_0, \hat{\mathfrak{y}}_0) + k \cdot d_{\mathfrak{C}}(\mathfrak{y}''_1, \hat{\mathfrak{y}}_1).$$

Thus, by the metric property of $d_{\mathfrak{C}}$,

$$\begin{aligned} c_k(\mathfrak{y}''_0, \mathfrak{y}''_1) - c_k(\mathfrak{y}'_0, \mathfrak{y}'_1) &< k \cdot \left[d_{\mathfrak{C}}(\mathfrak{y}''_0, \hat{\mathfrak{y}}_0) - d_{\mathfrak{C}}(\mathfrak{y}'_0, \hat{\mathfrak{y}}_0) \right] + k \cdot \left[d_{\mathfrak{C}}(\mathfrak{y}''_1, \hat{\mathfrak{y}}_1) - d_{\mathfrak{C}}(\mathfrak{y}'_1, \hat{\mathfrak{y}}_1) \right] \\ &< k \cdot \left[d_{\mathfrak{C}}(\mathfrak{y}''_0, \mathfrak{y}'_0) + d_{\mathfrak{C}}(\mathfrak{y}''_1, \mathfrak{y}'_1) \right]. \end{aligned}$$

Similar for $c_k(\mathfrak{y}'_0, \mathfrak{y}'_1) - c_k(\mathfrak{y}''_0, \mathfrak{y}''_1)$, we obtain

$$|c_k(\mathfrak{y}''_0, \mathfrak{y}''_1) - c_k(\mathfrak{y}'_0, \mathfrak{y}'_1)| < k \cdot \left[d_{\mathfrak{C}}(\mathfrak{y}''_0, \mathfrak{y}'_0) + d_{\mathfrak{C}}(\mathfrak{y}''_1, \mathfrak{y}'_1) \right].$$

Since c is lower semi-continuous, then $c_k \uparrow c$ and $c = \sup_k c_k$ for $c_k \in \mathrm{Lip}_b(\mathfrak{C}^{\otimes 2})$. Thus,

$$\liminf_{n\to\infty} \mathcal{C}(\alpha_n) \geqslant \liminf_{n\to\infty} \int c_k \, d\alpha_n = \int c_k \, d\alpha_* \to \mathcal{C}(\alpha_*).$$

The second equality is from weak convergence and the last one is by monotone convergence theorem.

(iii) Since $\tilde{d}^2_{\mathfrak{C}}([x_0, 0], [x_1, 0]) = 0$, then

$$\begin{aligned} &\int_{\mathfrak{C}^{\otimes 2}} \tilde{d}^2_{\mathfrak{C}}([x_0, r_0], [x_1, r_1]) \, d\alpha_* \\ &= \int_{\mathfrak{C}^{\otimes 2} \setminus (\mathfrak{o}, \mathfrak{o})} \tilde{d}^2_{\mathfrak{C}}([x_0, r_0], [x_1, r_1]) \, d\alpha_* + \int_{(\mathfrak{o}, \mathfrak{o})} \tilde{d}^2_{\mathfrak{C}}([x_0, 0], [x_1, 0]) \, d\alpha_* \\ &= \int_{\mathfrak{C}^{\otimes 2} \setminus (\mathfrak{o}, \mathfrak{o})} \tilde{d}^2_{\mathfrak{C}}([x_0, r_0], [x_1, r_1]) \, d\alpha_*. \end{aligned}$$

□

5.2.1 A Modified Variant of Wasserstein Distance

The $L2-$Kantorovich-Wasserstein distance $W_{d_{\mathfrak{C}}}$ in (3.9) can be generalized as follows: Let $p^i : \mathfrak{C}^{\otimes N} \to \mathfrak{C}$ be the projections on the i-th coordinate.

Definition 5.28. The modified **Wasserstein (extended) distance** $W_{\tilde{d}_{\mathfrak{C}}}$ in $\mathcal{P}_2(\mathfrak{C})$, induced by the quasi-cone metric $\tilde{d}_{\mathfrak{C}}$, between two measure $\nu_1, \nu_2 \in \mathcal{P}_2(\mathfrak{C})$, is defined as

$$W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) := \min\left\{ \int \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) : \alpha \in \mathcal{M}(\mathfrak{C} \times \mathfrak{C}),\, p^i_{\#}\alpha = \nu_i \right\}, \tag{5.29}$$

and $W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) := +\infty$ if $\nu_1(\mathfrak{C}) \neq \nu_2(\mathfrak{C})$. We denote $Opt_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1, \nu_2)$ as the set of all optimal plans α for $W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2)$.

Proposition 53 (mHK and Wasserstein distance $W_{\tilde{d}_{\mathfrak{C}}}$ on $\mathcal{P}_2(\mathfrak{C})$)**.** For every $\mu_1, \mu_2 \in \mathcal{M}(X)$, we have

$$\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2) = \min\left\{ W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) : \nu_i \in \mathcal{P}_2(\mathfrak{C}),\, \mathfrak{h}^2\nu_i = \mu_i \right\}. \tag{5.30}$$

Proof. If there exist $\nu_i \in \mathcal{P}_2(\mathfrak{C})$ satisfying $\nu_1(\mathfrak{C}) = \nu_2(\mathfrak{C})$ and $\mathfrak{h}^2\nu_i = \mu_i$, then there exists an optimal transport plan α of $W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2)$ and $\mathfrak{h}^2_i\alpha = \mathfrak{h}^2\nu_i = \mu_i$. Since α is concentrated only when $r_1 \leqslant r_2$, then $\mu_1(X) = \mathfrak{h}^2\nu_1(X) \leqslant \mathfrak{h}^2\nu_2(X) = \mu_2(X)$. Thus, $W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) \geqslant \mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2)$. On the other hand, let α be the optimal solution of $\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2)$. Then $\nu_i = (p^i)_{\#}\alpha$ satisfying $\mathfrak{h}^2\nu_i = \mu_i$, so $\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2) \geqslant W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2)$. □

5.2.2 Triangle Inequality

Lemma 54 (Dudley, [AGS05, Lemma 5.3.2], [ABS21, Lemma 8.4])**.** Let Z_1, Z_2, Z_3 be Radon separable metric spaces and let $\alpha^{12} \in \mathcal{P}(Z_1 \times Z_2)$ and $\alpha^{23} \in \mathcal{P}(Z_2 \times Z_3)$ such that $p^2_{\#}\alpha^{12} = p^1_{\#}\alpha^{23}$. Then there exists $\alpha \in \mathcal{P}(Z_1 \times Z_2 \times Z_3)$ such that

$$p^{1,2}_{\#}(\alpha) = \alpha^{12} \quad \text{and} \quad p^{2,3}_{\#}(\alpha) = \alpha^{23}, \tag{5.31}$$

where $p^{1,2}(z_1, z_2, z_3) = (z_1, z_2)$ and $p^{2,3}(z_1, z_2, z_3) = (z_2, z_3)$. More specifically, using the disintegration theorem,

$$\alpha^{12}(dz_1, dz_2) = \alpha^{12}_{z_2}(dz_1)\mu_2(dz_2) \quad \text{and} \quad \alpha^{23}(dz_2, dz_3) = \alpha^{23}_{z_2}(dz_3)\mu_2(dz_2),$$

Then

$$\alpha := \alpha^{12}_{z_2} \times \alpha^{23}_{z_2}(dz_1, dz_3)\mu_2(dz_2).$$

Proposition 55. The $W_{\tilde{d}_{\mathfrak{C}}}$ is a quasi-metric in $\mathcal{M}_2(\mathfrak{C})$, i.e. the symmetry does not hold.

Proof. By Jensen's inequality,

$$W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) := \min\left\{ \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha(\mathfrak{y}_1, \mathfrak{y}_2) : \alpha \in \mathcal{M}(\mathfrak{C} \times \mathfrak{C}),\, p^i_{\#}\alpha = \nu_i \right\}. \tag{5.32}$$

It suffices to show that for any $\nu_1, \nu_2, \nu_3 \in \mathcal{P}_2(\mathfrak{C})$,

$$W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_3) \leqslant W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) + W_{\tilde{d}_{\mathfrak{C}}}(\nu_2, \nu_3).$$

Let α^{12}, α^{23} be the optimal solutions of $W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2)$, $W_{\tilde{d}_{\mathfrak{C}}}(\nu_2, \nu_3)$ respectively such that $p^2_{\#}\alpha^{12} = \nu_2 = p^1_{\#}\alpha^{23}$. By Dudley's lemma, there exists $\alpha^{13} \in \mathcal{P}(\mathfrak{C}^{\otimes 3})$ such that

$$p^{1,2}(\alpha^{13}) = \alpha^{12} \quad \text{and} \quad p^{2,3}_{\#}(\alpha^{13}) = \alpha^{23}.$$

Therefore, applying the triangle inequality for $\tilde{d}_{\mathfrak{C}}$

$$\begin{aligned} W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_3) \leqslant \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_3)\, d\alpha^{13} &\leqslant \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha^{13} + \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_2, \mathfrak{y}_3)\, d\alpha^{13} \\ &= \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_2)\, d\alpha^{12} + \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_2, \mathfrak{y}_3)\, d\alpha^{23} \\ &= W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) + W_{\tilde{d}_{\mathfrak{C}}}(\nu_2, \nu_3). \end{aligned}$$

□

Lemma 56 ([LMS18, Lemma 7.10])**.** Given $N \geqslant 2$ and $\alpha \in \mathcal{M}_2(\mathfrak{C}^{\otimes N})$ satisfying $\mathfrak{h}^2_i\alpha = \mu_i$ for $i = 1, \ldots, N$. Suppose that $\omega_j := \alpha(\{\mathfrak{y} \in \mathfrak{C}^{\otimes N} : p^j(\mathfrak{y}) = \mathfrak{o}\}) \geqslant 1$. We can choose a scalar

$$\vartheta(\mathfrak{y}) := \begin{cases} \mathsf{r}_j(\mathfrak{y}) & \text{if } \mathfrak{y}_j \neq \mathfrak{o}, \\ \omega_j^{-\frac{1}{2}} & \text{otherwise,} \end{cases} \tag{5.33}$$

then $\bar{\alpha} := \mathrm{dil}_{\vartheta,2}(\alpha)$ satisfies

1. $\mathfrak{h}_i^2 \bar{\alpha} = \mu_i$ for $i = 1, \ldots, N$.
2. $\int \tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i)\, d\alpha = \int \tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_{i-1}, \mathfrak{y}_i)\, d\bar{\alpha}$ for $i = 2, \ldots, N$.
3. the normalization of jth lift,

$$p^j_{\#}(\bar{\alpha}) = \delta_{\mathfrak{o}} + \mathsf{p}_{\#}(\mu_j \otimes \delta_1), \tag{5.34}$$

i.e. for any $\varphi \in \mathcal{B}_b(\mathfrak{C})$,

$$\int_{\mathfrak{C}} \varphi(\mathfrak{y})\, d[p^j_{\#}\bar{\alpha}] = \varphi(\mathfrak{o}) + \int_{\mathfrak{C}} \varphi([x, r])\, d[\mathsf{p}_{\#}(\mu_j \otimes \delta_1)]. \tag{5.35}$$

Lemma 57 (Gluing Lemma)**.** Given $N \geqslant 2$ and a sequence of non-negative Radon measures $\mu_i \in \mathcal{M}(X)$, $i = 1, \ldots, N$ such that $\mu_1(X) \leqslant \mu_2(X) \ldots \leqslant \mu_N(X)$. Then there exists $\alpha \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ such that

$$\begin{aligned} \mathfrak{h}_i^2 \alpha &= \mu_i, \quad \text{ for } i = 1, \ldots, N, \\ \mathbf{nHK}^2_{\mathfrak{C}}(\mu_{i-1}, \mu_i) &= \int \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_{i-1}, \mathfrak{y}_i)\, d\alpha, \quad \text{ for } i = 2, \ldots, N. \end{aligned} \tag{5.36}$$

Moreover, there exists $\beta \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ such that

$$\begin{aligned} \mathfrak{h}_i^2 \beta &= \mu_i, \quad \text{ for } i = 1, \ldots, N, \\ \mathbf{nHK}^2_{\mathfrak{C}}(\mu_1, \mu_i) &= \int \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_1, \mathfrak{y}_i)\, d\beta, \quad \text{ for } i = 2, \ldots, N. \end{aligned} \tag{5.37}$$

This lemma is generalized from [AGS05, Lemma 5.3.4] and [LMS18, Lemma 7.11].

Proof. We will prove the first statement in (5.36) by induction. The case when $N = 2$ holds by Theorem 51. Assume it holds for $N = k$, then there exist $\alpha_k \in \mathcal{P}_2(\mathfrak{C}^{\otimes k})$ satisfying (5.36) and $\alpha' \in \mathrm{Opt}_{\mathbf{nHK}_{\mathfrak{C}}}(\mu_k, \mu_{k+1})$. By Lemma 56, applying the scaling to α_k and α', we obtain $p^k_{\#}\alpha_k = p^1_{\#}\alpha'$. By Lemma 54, there exists $\alpha_{k+1} \in \mathcal{P}_2(\mathfrak{C}^{\otimes k+1})$ satisfying $p^{1,\ldots,k}_{\#}\alpha_{k+1} = \alpha_k$ and $p^{k,k+1}_{\#}\alpha_{k+1} = \alpha'$. In particular, α_{k+1} satisfies (5.36).

Similarly, we will prove the second statement in (5.37) by induction. The case when $N = 2$ holds by Theorem 51. Assume it holds for $N = k$, then there exist $\beta_k \in \mathcal{P}_2(\mathfrak{C}^{\otimes k})$ satisfying (5.37) and $\beta' \in \mathrm{Opt}_{\mathbf{nHK}_{\mathfrak{C}}}(\mu_1, \mu_{k+1})$. By Lemma 54, there exists $\beta_{k+1} \in \mathcal{P}_2(\mathfrak{C}^{\otimes k+1})$ satisfying $p^{1,\ldots,k}_{\#}\beta_{k+1} = \beta_k$ and $p^{1,k+1}_{\#}\beta_{k+1} = \beta'$. In particular, α_{k+1} satisfies (5.37). $\square$

Lemma 58 ([LMS18, Section 7.4])**.** Continuing with the Gluing Lemma 57. Assume that $\alpha((\mathfrak{o},\mathfrak{o})) = 0$. Then

i) Let $M^2 := \sum_{i=1}^{N} \mu_i(X)$ and suppose that α satisfies (5.36). If the scalar function is chosen as

$$\vartheta(\mathfrak{y}) = (r_*)^{-1}|\mathfrak{y}|_2, \quad \text{where} \quad r_*^2 = \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_2^2 \, d\alpha = M^2, \tag{5.38}$$

then $\alpha' := \mathrm{dil}_{\vartheta,2}(\alpha) \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ satisfies (5.36) and it is concentrated on

$$\mathfrak{C}[M]^{\otimes N} = \left\{ \mathfrak{y} \in \mathfrak{C}^{\otimes N} : |\mathfrak{y}|_2 \leqslant M \right\}. \tag{5.39}$$

ii) Let $\Theta := \sqrt{\mu_1(X)} + \sum_{i=2}^{N} \mathbf{mHK}_{\mathfrak{C}}(\mu_{i-1}, \mu_i)$ and suppose that α satisfies (5.36). If the scalar function is chosen as

$$\vartheta(\mathfrak{y}) = (r_*)^{-1}|\mathfrak{y}|_\infty = (r_*)^{-1} \sup_i \mathsf{r}_i(\mathfrak{y}), \quad \text{where} \quad r_*^2 = \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_\infty^2 \, d\alpha, \tag{5.40}$$

then $\alpha'' := \mathrm{dil}_{\vartheta,2}(\alpha) \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ satisfies (5.36) and it is concentrated on

$$\mathfrak{C}[\Theta]^{\otimes N} = \left\{ \mathfrak{y} \in \mathfrak{C}^{\otimes N} : |\mathfrak{y}|_\infty \leqslant \Theta \right\}. \tag{5.41}$$

iii) Let $\Xi := \sqrt{\mu_1(X)} + \sum_{i=2}^{N} \mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_i)$ and suppose that α satisfies (5.37). If the scalar function is chosen as

$$\vartheta(\mathfrak{y}) = (r_*)^{-1}|\mathfrak{y}|_\infty = (r_*)^{-1} \sup_i \mathsf{r}_i(\mathfrak{y}), \quad \text{where} \quad r_*^2 = \int_{\mathfrak{C}^{\otimes N}} |\mathfrak{y}|_\infty^2 \, d\alpha, \tag{5.42}$$

then $\alpha''' := \mathrm{dil}_{\vartheta,2}(\alpha) \in \mathcal{P}_2(\mathfrak{C}^{\otimes N})$ is concentrated on

$$\mathfrak{C}[\Xi]^{\otimes N} = \left\{ \mathfrak{y} \in \mathfrak{C}^{\otimes N} : |\mathfrak{y}|_\infty \leqslant \Xi \right\}. \tag{5.43}$$

Corollary 59. Given a sequence of non-negative Radon measures $\mu_i \in \mathcal{M}(X)$, $i = 1, \dots, N$ such that $\mu_1(X) \leqslant \mu_2(X) \dots \leqslant \mu_N(X)$, then there exist $\nu_i \in \mathcal{P}_2(\mathfrak{C})$ such that

(1) ν_i is concentrated in $\mathfrak{C}[r]$ where $r = \min\{M, \Theta\}$.

(2) $\mathfrak{h}^2 \nu_i = \mu_i$ for $i = 1, \dots, N$.

(3) $\mathbf{mHK}_{\mathfrak{C}}(\mu_i, \mu_{i+1}) = W_{\tilde{d}_{\mathfrak{C}}}(\nu_i, \nu_{i+1})$ for $i = 2, \dots, N$.

Moreover, there exists $\beta_i \in \mathcal{P}_2(\mathfrak{C})$ such that

(1) β_i is concentrated in $\mathfrak{C}[\Xi]$.

(2) $\mathfrak{h}^2\beta_i = \mu_i$ for $i = 1, \ldots, N$.

(3) $\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_i) = W_{\tilde{d}_{\mathfrak{C}}}(\beta_1, \beta_i)$ for $i = 2, \ldots, N$.

Corollary 60. $\mathbf{mHK}_{\mathfrak{C}}$ is a quasi-metric on $\mathcal{M}(X)$, i.e. the symmetry does not hold.

Proof. It is obvious that the identity holds and symmetry does not hold. Now, let us prove that triangle inequality holds, i.e. for any $\mu_1, \mu_2, \mu_3 \in \mathcal{M}(X)$,

$$\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_3) \leqslant \mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2) + \mathbf{mHK}_{\mathfrak{C}}(\mu_2, \mu_3). \tag{5.44}$$

It suffices to assume that $\mu_1(X) \leqslant \mu_2(X) \leqslant \mu_3(X)$, otherwise the right-hand side must be ∞. Apply the corollary 59, there exists $\nu_i \in \mathcal{P}_2(\mathfrak{C})$ for $i = 1, 2, 3$ such that $\mathfrak{h}^2\nu_i = \mu_i$ and

$$\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2) = W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) \quad \text{and} \quad \mathbf{mHK}_{\mathfrak{C}}(\mu_2, \mu_3) = W_{\tilde{d}_{\mathfrak{C}}}(\nu_2, \nu_3).$$

Applying the triangle inequality for $W_{\tilde{d}_{\mathfrak{C}}}$ in Proposition 55, we obtain

$$\begin{aligned}\mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_3) \leqslant W_{\tilde{d}_{\mathfrak{C}}}(\mu_1, \mu_3) &\leqslant W_{\tilde{d}_{\mathfrak{C}}}(\nu_1, \nu_2) + W_{\tilde{d}_{\mathfrak{C}}}(\nu_2, \nu_3) \\ &= \mathbf{mHK}_{\mathfrak{C}}(\mu_1, \mu_2) + \mathbf{mHK}_{\mathfrak{C}}(\mu_2, \mu_3).\end{aligned}$$

□

5.3 The Support of Optimal Transport Plan

In this subsection, we will explore several properties of the support of optimal transport plans for the $\mathbf{mHK}^2_{\mathfrak{C}}$ problem. These properties will help compute the explicit solution of $\mathbf{mHK}^2_{\mathfrak{C}}$ between two Dirac measures as discussed in Subsection 5.3.1. Furthermore, the findings of this subsection will contribute to establishing dual pairs for the dual problem, as discussed in Section 5.4.

Lemma 61. Given any optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ with $\int c \, d\hat{\alpha} < \infty$ and $\hat{\alpha}((\mathfrak{o}, \mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{\mathsf{r}_2, 2}\hat{\alpha}$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$, then $\tilde{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$.

Proof. Since $\tilde{d}^2_{\mathfrak{C}}([x,\cdot],[y,\cdot])$ is 2-homogeneous, we obtain

$$\int \tilde{d}^2_{\mathfrak{C}}([x,\rho],[y,1])\, d\tilde{\alpha}([x,\rho],[y,1]) = \int \tilde{d}^2_{\mathfrak{C}}([x,\frac{r_0}{r_1}],[y,\frac{r_1}{r_1}]) \cdot r_1^2 \, d\hat{\alpha}([x,r_0],[y,r_1])$$
$$= \int \tilde{d}^2_{\mathfrak{C}}([x,r_0],[y,r_1])\, d\hat{\alpha}([x,r_0],[y,r_1]).$$

□

Moreover, we can write the support of $\tilde{\alpha}$ as

$$\Gamma := \operatorname{spt}(\tilde{\alpha}) = \left\{([x,\rho],[y,1]) : ([x,r_0],[y,r_1]) \in spt(\hat{\alpha}),\ \rho = \frac{r_0}{r_1} \in [0,1]\right\}. \tag{5.45}$$

We can decompose Γ into $\Gamma_0 \cup \Gamma^+$, where $\Gamma^+ := \Gamma_= \cup \Gamma_>$ and

$$\Gamma_0 := \Gamma \cap \{\rho = 0\} \subseteq \left\{(\mathfrak{o},[y,1]) : y \in spt(\mu_1)\right\}, \tag{5.46}$$

$$\Gamma_= := \Gamma \cap \{\rho = 1\} \subseteq \left\{([x,1],[y,1]) : x \in spt(\mu_0), y \in spt(\mu_1)\right\}, \tag{5.47}$$

$$\Gamma_> := \Gamma \cap \{0 < \rho < 1\} \subseteq \left\{([x,\rho],[y,1]) : x \in spt(\mu_0), y \in spt(\mu_1), \rho \in (0,1)\right\}. \tag{5.48}$$

Here, Γ_0 is a set where the mass grows from the vertex. We denote $\Gamma_=$ as a balanced transfer set, where the initial-to-target mass ratio is 1, and $\Gamma_>$ as an unbalanced transfer set, where the initial-to-target mass ratio falls between 0 and 1. For simplicity in notation, define $\theta : X \times X \to \mathbb{R}_+$, $\beta : [0,1] \to [0,\frac{\pi}{2}]$, and $\gamma : [0,1] \to [0,1]$ such that

$$\theta = \theta(x,y) := d(x,y), \quad \beta = \beta(\rho) := \cos^{-1}(\rho), \quad \gamma = \gamma(\rho) := \sqrt{1-\rho^2}. \tag{5.49}$$

Here, shown in Figure 5.4, the variable θ represents the angle between $[x,\rho]$ and $[y,1]$, the variable β denotes the angle between $[z,\rho]$ and $[y,1]$, and the variable γ is the length of straight line from $[z,\rho]$ to $[y,1]$.

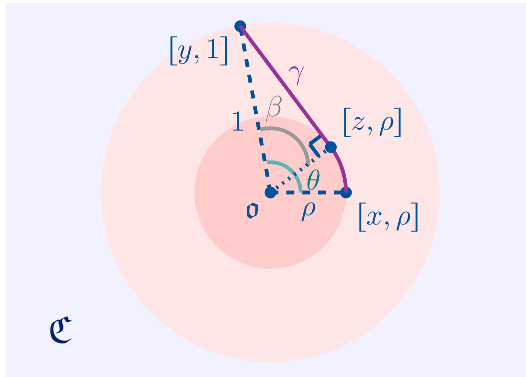


FIGURE 5.4: Representations of θ, β, γ

Thus, we can simplify the cost $\tilde{d}^2_{\mathfrak{C}}([x,\rho],[y,1])$ for all $\rho \in [0,1]$ as

$$c(x,\rho,y,1) = \begin{cases} \rho^2 + 1 - 2\rho\cos(\theta) & \text{if } \theta \leqslant \beta, \\ (\rho(\theta-\beta)+\gamma)^2 & \text{if } \theta > \beta. \end{cases} \tag{5.50}$$

When $\theta > \frac{\pi}{2}$, let ρ_θ denote the unique solution of $(\theta-\beta)\gamma = \rho$ in $(0,1)$, where the uniqueness is proved in Claim B.1. We will now elaborate on the significance of ρ_θ. This value serves as a threshold: as demonstrated in the following Lemma, it illustrates that the cost of transporting ρ units of mass from location x to location y with a unit of mass can be minimized if the mass is partially transferred from location x to y while partially growing in place.

Lemma 62. Shown in Claim B.3, given $\tilde{r} \in (\rho\rho_\theta^{-1}, 1)$, then

$$c(x,\rho,y,\tilde{r}) + c(x,0,y,\sqrt{1-\tilde{r}^2}) < c(x,\rho,y,1). \tag{5.51}$$

The investigation into ρ_θ aids in refining our understanding of the support of the optimal transport plan. As evidenced by the subsequent proposition, we conclude that the optimal transport plan $\tilde{\alpha}$ is not concentrated on the set $\{\theta > \frac{\pi}{2}, 0 < \rho < \rho_\theta\}$.

Proposition 63. *$D := \Gamma^+ \cap \{\theta > \frac{\pi}{2}, \rho < \rho_\theta\} = \varnothing$.*

Proof. Assume that $D \neq \varnothing$. Fix $\tilde{r} \in (\rho\rho_\theta^{-1}, 1)$ and define two measurable maps T_1, T_2 on D by

$$T_1\big(([x,\rho],[y,1])\big) := ([x,\rho],[y,\tilde{r}]),$$

$$T_2\big(([x,\rho],[y,1])\big) := ([x,0],[y,\sqrt{1-\tilde{r}^2}]),$$

we can construct a better plan α' than $\tilde{\alpha}$ where

$$\alpha' := \tilde{\alpha}\mathbb{1}_{D^c} + (T_1)_\#(\tilde{\alpha}\mathbb{1}_D) + (T_2)_\#(\tilde{\alpha}\mathbb{1}_D).$$

By Lemma 62,

$$\int c\, d(\alpha' - \tilde{\alpha}) = \int_D \Big[c(x,\rho,y,\tilde{r}) + c(x,0,y,\sqrt{1-\tilde{r}^2}) - c(x,\rho,y,1)\Big]\, d\tilde{\alpha} < 0,$$

which contradicts the optimality of $\tilde{\alpha}$. Hence, $D = \varnothing$. □

Corollary 64. By Proposition 63, we can rewrite the cost function in (5.50)

$$c(x,\rho,y,1) = \begin{cases} \rho^2 + 1 - 2\rho\cos(\theta) & \text{if } \theta \leqslant \beta, \\ \rho^2[(\theta - \beta(\rho_\theta))^2 + 1] + 1 & \text{if } \theta > \frac{\pi}{2}, \rho < \rho_\theta, \\ (\rho(\theta-\beta)+\gamma)^2 & \text{else.} \end{cases} \tag{5.52}$$

Due to that $c(x,\cdot,y,\cdot)$ is 2-homogeneous, we have that for $r_0 \leqslant r_1$,

$$c(x,r_0,y,r_1) = \begin{cases} r_0^2 + r_1^2 - 2r_0r_1\cos(\theta) & \text{if } \theta \leqslant \beta, \\ r_0^2[(\theta - \beta(\rho_\theta))^2 + 1] + r_1^2 & \text{if } \theta > \frac{\pi}{2}, \rho < \rho_\theta, \\ (r_0(\theta-\beta)+\sqrt{r_1^2 - r_0^2})^2 & \text{else.} \end{cases} \tag{5.53}$$

Building upon our earlier discoveries regarding the support of the optimal transport plan, we proceed to develop two functions, F and G, and analyze their properties. This exploration serves as a foundation for deriving the optimality conditions for the **One-To-Two** and **Two-To-One** cases in Subsection 5.3.1, as well as for establishing dual pairs in Subsubsection 5.4.1.2.

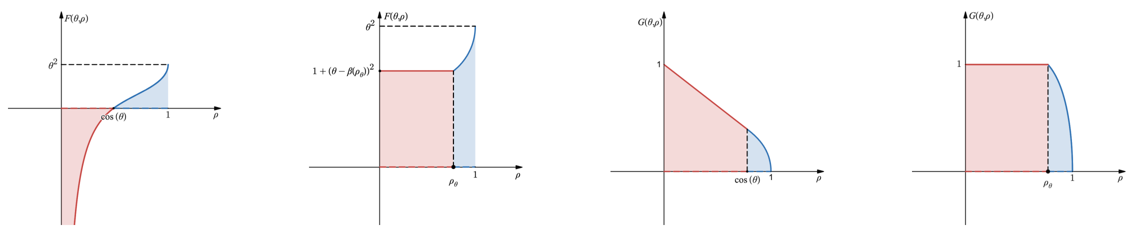
Definition 5.54. Let $F(\theta,\rho) := \mathcal{F}(x,y,\rho) = \frac{1}{2\rho}\partial_\rho c(x,\rho,y,1)$ be defined on $\{\theta \geqslant 0, \rho \in (0,1]\}$. Then

$$F(\theta,\rho) = \begin{cases} 1 - \rho^{-1}\cos(\theta) & \text{if } \theta \leqslant \beta \\ (\theta - \beta(\rho_\theta))^2 + 1 & \text{if } \theta > \frac{\pi}{2}, \rho < \rho_\theta\ . \\ \big(\theta - \beta + \frac{\gamma}{\rho}\big)\cdot(\theta - \beta) & \text{else} \end{cases} \tag{5.55}$$

Let $G(\theta,\rho) := \mathcal{G}(y,x,\rho) = c(x,\rho,y,1) - \mathcal{F}(x,y,\rho)\cdot\rho^2$ on $\{\theta \geqslant 0, \rho \in [0,1]\}$. Then

$$G(\theta,\rho) = \begin{cases} 1-\rho\cos(\theta) & \text{if } \theta \leqslant \beta \\ 1 & \text{if } \theta > \frac{\pi}{2}, \rho < \rho_\theta \\ \big(\rho(\theta-\beta)+\gamma\big)\cdot\gamma & \text{else} \end{cases}. \tag{5.56}$$

Remark 5.57. *F is increasing and G is decreasing. These graphs are drawn from the link:* `https://www.desmos.com/calculator`.



(a) $F(\theta,\cdot)$ when $\theta < \frac{\pi}{2}$ (b) $F(\theta,\cdot)$ when $\theta \geqslant \frac{\pi}{2}$ (c) $G(\theta,\cdot)$ when $\theta < \frac{\pi}{2}$ (d) $G(\theta,\cdot)$ when $\theta \geqslant \frac{\pi}{2}$

5.3.1 Examples Between Dirac Measures

In this section, we will discuss explicit solutions of $\mathbf{nHK}_{\mathfrak{C}}$ between two Dirac measures.

5.3.1.1 One-to-One

Fix $a, b \in \mathbb{R}^d$ with $\theta = d(a,b)$ and $0 < n \leqslant m$ with $\rho_* := \sqrt{\frac{n}{m}}$. Let $\mu = n\delta_a$ and $\nu = m\delta_b$.

Theorem 65.

$$\mathbf{HK}^2(\mu,\nu) = n + m - 2\sqrt{n\cdot m}\cos(\theta_\pi). \tag{5.58}$$

Proof. Pick $\alpha \in Opt_{\mathbf{HK}}(\mu,\nu)$. By Hölder's inequality,

$$\begin{aligned}\int \big(r_0^2 + r_1^2 - 2r_0r_1\cos(\theta_\pi)\big)\,d\alpha &= n + m - 2\cos(\theta_\pi)\int r_0r_1\,d\alpha \\ &\geqslant n + m - 2\cos(\theta_\pi)\left(\int r_0^2\,d\alpha\right)^{\frac{1}{2}}\left(\int r_1^2\,d\alpha\right)^{\frac{1}{2}} \\ &\geqslant n + m - 2\sqrt{n\cdot m}\cos(\theta_\pi).\end{aligned}$$

and the equality holds when $cr_1 = r_0$ where

$$c^2 = \frac{\int r_0^2 \, d\alpha}{\int r_1^2 \, d\alpha} = \frac{n}{m} = \rho_*^2.$$

The lower bound is achieved when $\alpha = \delta_{([a,\sqrt{n}],[b,\sqrt{m}])}$. □

Theorem 66. When $\theta > \frac{\pi}{2}$, let ρ_θ be the unique solution of $(\theta - \beta)\gamma = \rho$ in $(0, 1)$ where the uniqueness is proved in Claim B.1.

- **Case I:** Suppose $\theta \leqslant \beta(\rho_*)$, then

$$\mathbf{mHK}_{\mathfrak{C}}^2(\mu, \nu) = \mathbf{HK}^2(\mu, \nu) = n + m - 2\sqrt{n \cdot m}\cos(\theta), \tag{5.59}$$

  and one of the optimal transport plans is $\alpha = \delta_{([a,\sqrt{n}],[b,\sqrt{m}])}$.

- **Case II:** Suppose $\theta > \beta(\rho_*)$. If $\theta > \frac{\pi}{2}$ and $\rho_\theta > \rho_*$, then

$$\mathbf{mHK}_{\mathfrak{C}}^2(\mu, \nu) = n \cdot \left[\big(\theta - \beta(\rho_\theta)\big)^2 + 1\right] + m, \tag{5.60}$$

  and there must be mass transported from nothing and one of the optimal transport plans is

$$\alpha = (1 - \varepsilon)\delta_{([a,0],[b,\sqrt{\frac{m-m_\theta}{1-\varepsilon}}])} + \varepsilon\delta_{([a,\sqrt{\frac{n}{\varepsilon}}],[b,\sqrt{\frac{m_\theta}{\varepsilon}}])} \quad \text{with } m_\theta = n\rho_\theta^{-2},$$

  for any $\varepsilon \in (0, 1)$.

- **Case III:** Suppose $\theta > \beta(\rho_*)$. If $\theta > \frac{\pi}{2}$ and $\rho_\theta \leqslant \rho_*$ or $\theta \leqslant \frac{\pi}{2}$, then

$$\mathbf{mHK}_{\mathfrak{C}}^2(\mu, \nu) = m \cdot \left[k\big(\theta - \beta(\rho_*)\big) + \gamma(\rho_*)\right]^2, \tag{5.61}$$

  and one of the optimal transport plans is $\alpha = \delta_{([a,\sqrt{n}],[b,\sqrt{m}])}$.

In summary,

$$\mathbf{mHK}_{\mathfrak{C}}^2(n\delta_a, m\delta_b) = c(a, \sqrt{n}, b, \sqrt{m}), \tag{5.62}$$

where c is defined in (5.53).

Remark 5.63. The **Case I** is obvious since $\mathbf{mHK}_{\mathfrak{C}}^2 \geqslant \mathbf{HK}^2$ and the equality holds when we choose a transport plan $\alpha = \delta_{[a,\sqrt{n}],[b,\sqrt{m}]}$ for $\mathbf{mHK}_{\mathfrak{C}}^2(\mu, \nu)$.

*Proof of **Case II** in the Theorem 66.* For any $\alpha \in Opt_{\mathbf{nHK}_{\mathfrak{C}}}(\mu, \nu)$ such that $\alpha((\mathfrak{o}, \mathfrak{o})) = 0$, let $\tilde{\alpha} = \mathrm{dil}_{\mathsf{r}_2,2}\alpha$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$. From the cost redefined in (5.52), if $\theta > \frac{\pi}{2}$, then

$$\int c(x, \rho, y, 1)\, d\tilde{\alpha} = \int c(a, \rho, b, 1)\, d\tilde{\alpha}$$

$$= \int_{\{\rho < \rho_\theta\}} \left(\rho^2\left[(\theta - \beta(\rho_\theta))^2 + 1\right] + 1\right) d\tilde{\alpha} + \int_{\{\rho \geqslant \rho_\theta\}} \left(\rho(\theta - \beta) + \gamma\right)^2 d\tilde{\alpha}. \tag{5.64}$$

Let $g(\rho) := (\theta - \beta)^2 + 2(\theta - \beta)\frac{\gamma}{\rho} - 1$. Since

$$g'(\rho) = -2\rho^{-2} \cdot \left[(\theta - \beta)\gamma - \rho\right],$$

then from the proof of Claim B.1, $g'(\rho) < 0$ if $\rho \in (0, \rho_\theta)$ and $g'(\rho) > 0$ if $\rho \in (\rho_\theta, 1]$. Therefore, $g(\rho) \geqslant g(\rho_\theta) = \left[(\theta - \beta(\rho_\theta))^2 + 1\right]$ and

$$\int_{\{\rho \geqslant \rho_\theta\}} \left(\rho(\theta - \beta) + \gamma\right)^2 d\tilde{\alpha} = \int_{\{\rho \geqslant \rho_\theta\}} \left(\rho^2 \cdot g(\rho) + 1\right) d\tilde{\alpha} \geqslant \int_{\{\rho \geqslant \rho_\theta\}} \left(\rho^2 \cdot g(\rho_\theta) + 1\right) d\tilde{\alpha}.$$

Then we get the lower bound of (5.64),

$$\int c(x, \rho, y, 1)\, d\tilde{\alpha} \geqslant \int \left(\rho^2\left[(\theta - \beta(\rho_\theta))^2 + 1\right] + 1\right) d\tilde{\alpha} = n \cdot \left[(\theta - \beta(\rho_\theta))^2 + 1\right] + m.$$

The lower bound is achieved when

$$\alpha = (1 - \varepsilon)\delta_{([a,0],[b,\sqrt{\frac{m - m_\theta}{1-\varepsilon}}])} + \varepsilon\delta_{([a,\sqrt{\frac{n}{\varepsilon}}],[b,\sqrt{\frac{m_\theta}{\varepsilon}}])} \text{ with } m_\theta = n\rho_\theta^{-2} \tag{5.65}$$

for any $\varepsilon \in (0, 1)$. □

*Proof of **Case III** in the Theorem 66.* For any $\alpha \in Opt_{\mathbf{nHK}_{\mathfrak{C}}}(\mu, \nu)$ such that $\alpha((\mathfrak{o}, \mathfrak{o})) = 0$, let $\tilde{\alpha} = \mathrm{dil}_{\mathsf{r}_2,2}\alpha$. Let the Lagrangian be

$$g(\rho) := c(a, \rho, y, 1) - \lambda \cdot (\rho^2 - \rho_*^2) \quad \text{with } \lambda = F(\theta, \rho_*)$$

for $\rho \in (0, 1]$ where F is defined in (5.55). By Claim B.6, $g(\rho) \geqslant g(\rho_*)$. Then

$$\int c(x, \rho, y, 1)\, d\tilde{\alpha} = \int c(a, \rho, b, 1)\, d\tilde{\alpha} = \int \left[g(\rho) + \lambda \cdot (\rho^2 - \rho_*^2)\right] d\tilde{\alpha}$$

$$\geqslant \int g(\rho_*)\, d\tilde{\alpha} + \lambda \cdot \int (\rho^2 - \rho_*^2)\, d\tilde{\alpha}$$

$$= m \cdot c(a, \rho_*, y, 1) + \lambda \cdot (n - \rho_*^2 m) = c(a, \sqrt{n}, y, \sqrt{m}).$$

The lower bound is achieved when $\alpha = \delta_{[a,\sqrt{n}],[b,\sqrt{m}]}$. □

5.3.1.2 One-To-Two

Fix $a, b_1, b_2 \in \mathbb{R}^d$ with $\theta_i := d(a, b_i)$ and $n \leqslant m := m_1 + m_2$. Let $\mu = n\delta_a$ and $\nu = \sum_{i=1}^{2} m_i \delta_{b_i}$. We want to find the optimality conditions for computing $\mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu)$.

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu) = \min_{n_1, n_2} \left\{ \sum_{i=1}^{2} \mathbf{mHK}^2_{\mathfrak{C}}(n_i\delta_a, m_i\delta_{b_i}) : \sum_{i=1}^{2} n_i = n \text{ and } n_i \in [0, m_i] \right\}. \tag{5.66}$$

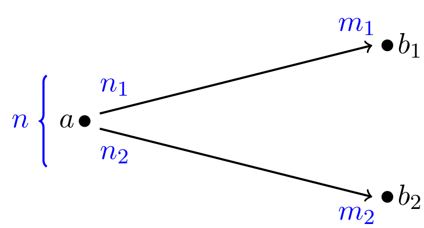


FIGURE 5.5: One-To-Two Transportation

By (5.62) in One-to-One example,

$$\mathbf{mHK}^2_{\mathfrak{C}}(n_i\delta_a, m_i\delta_{b_i}) = c(a, \sqrt{n_i}, b_i, \sqrt{m_i}). \tag{5.67}$$

Then the problem can be reduced to

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu) = \min_{n_1, n_2} \left\{ \sum_{i=1}^{2} c(a, \sqrt{n_i}, b_i, \sqrt{m_i}) : \sum_{i=1}^{2} n_i = n \text{ and } n_i \in [0, m_i] \right\} \tag{5.68}$$

$$= \min_{n_1} \left\{ \mathrm{Obj}(n_1) : n_1 \in [0, m_1] \cap [n - m_2, n] \right\}, \tag{5.69}$$

where $\mathrm{Obj}(n_1) := c(a, \sqrt{n_1}, b_1, \sqrt{m_1}) + c(a, \sqrt{n - n_1}, b_2, \sqrt{m_2})$.

Theorem 67. If $n - m_2 = m_1$, then $\mathrm{Obj} \equiv \mathrm{Obj}(m_1)$. Suppose that $n - m_2 < m_1$. Let

$$f(n_1) := \partial_{n_1} \mathrm{Obj}(n_1) = F\Big(\theta_1, \sqrt{\frac{n_1}{m_1}}\Big) - F\Big(\theta_2, \sqrt{\frac{n - n_1}{m_2}}\Big), \tag{5.70}$$

where F is defined in (5.55). Then $\mathrm{Obj}(n_1)$ attains the minimum within the interior $(0 \vee (n - m_2), n \wedge m_1)$ if and only if there exists $\hat{n}_1 \in (0 \vee (n - m_2), n \wedge m_1)$ such that $f(\hat{n}_1) = 0$. Otherwise, $\mathrm{Obj}(n_1)$ attains the minimum at the boundary, more specifically,

i) If $f(0 \vee (n - m_2)) \geqslant 0$, then f attains the minimum at $0 \vee (n - m_2)$.

ii) If $f(n \wedge m_1) \leqslant 0$, then f attains the minimum at $n \wedge m_1$.

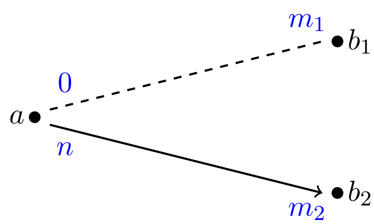


FIGURE 5.6: When $n \leqslant m_2, f(0) \geqslant 0$

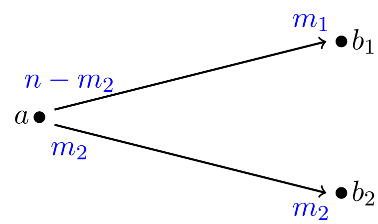


FIGURE 5.7: When $n > m_2, f(n - m_2) \geqslant 0$

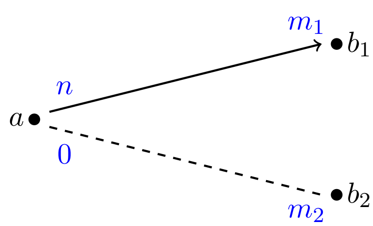


FIGURE 5.8: When $n \leqslant m_1, f(n) \leqslant 0$

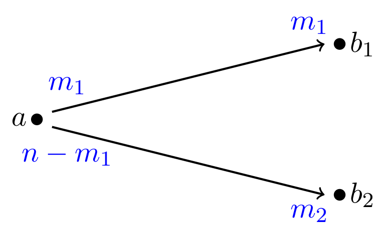


FIGURE 5.9: When $n > m_1, f(n - m_1) \leqslant 0$

Proof. f is increasing in n_1 since it's the sum of increasing functions. Therefore, $\text{Obj}(n_1)$ attains the minimum within $(0 \vee (n - m_2), n \wedge m_1)$ if and only if there exists $\hat{n}_1 \in (0 \vee (n - m_2), n \wedge m_1)$ such that $f(\hat{n}_1) = 0$. Else, if $f(0 \vee (n - m_2) \geqslant 0$, then $f(n_1) \geqslant f(0 \vee (n - m_2) \geqslant 0$. Thus, $\text{Obj}(n_1)$ is increasing. If $f(n \wedge m_1) \leqslant 0$, then $f(n_1) \leqslant f(n \wedge m_1) \leqslant 0$ so that $\text{Obj}(n_1)$ is decreasing. □

5.3.1.3 Two-To-One

Fix $a_1, a_2, b \in \mathbb{R}^d$ with $\theta_i := d(a_i, b)$ and $n := n_1 + n_2 \leqslant m$. Let $\mu = \sum_{i=1}^{2} n_i \delta_{a_i}$ and $\nu = m\delta_b$. We want to find the optimality conditions for computing $\mathbf{nHK}^2_{\mathfrak{C}}(\mu, \nu)$.

$$\mathbf{nHK}^2_{\mathfrak{C}}(\mu, \nu) = \min_{m_1, m_2} \left\{ \sum_{i=1}^{2} \mathbf{nHK}^2_{\mathfrak{C}}(n_i \delta_{a_i}, m_i \delta_b) : \sum_{i=1}^{2} m_i = m \text{ and } m_i \in [n_i, m] \right\}.$$

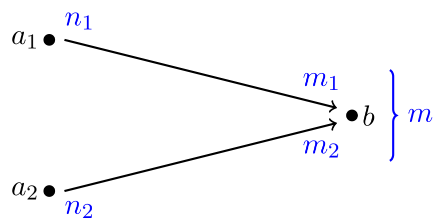


FIGURE 5.10: Two-To-One Transportation

By (5.62) in One-to-One example,

$$\mathbf{mHK}^2_{\mathfrak{C}}(n_i\delta_{a_i}, m_i\delta_b) = c(a_i, \sqrt{n_i}, b, \sqrt{m_i}). \tag{5.71}$$

Then the problem can be reduced to

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu) = \min_{m_1,m_2}\left\{\sum_{i=1}^{2} c(a_i,\sqrt{n_i},b,\sqrt{m_i}) : \sum_{i=1}^{2} m_i = m \text{ and } m_i \in [n_i, m]\right\} \tag{5.72}$$

$$= \min_{m_1}\left\{\mathrm{Obj}(m_1) : m_1 \in [n_1, m-n_2]\right\}, \tag{5.73}$$

where $\mathrm{Obj}(m_1) = c(a_1,\sqrt{n_1},b,\sqrt{m_1}) + c(a_2,\sqrt{n_2},b,\sqrt{m-m_1})$

Theorem 68. Let

$$f(m_1) := \partial_{m_1}\mathrm{Obj}(m_1) = G\big(\theta_1, \sqrt{\frac{n_1}{m_1}}\big) - G\big(\theta_2, \sqrt{\frac{n_2}{m-m_1}}\big). \tag{5.74}$$

where G is defined in (5.56). If $n_1 = m - n_2$, then $\mathrm{Obj} \equiv \mathrm{Obj}(n_1)$ and $f(n_1) = 0$. Suppose that $n_1 < m - n_2$. Then there exists $\hat{m}_1 \in (n_1, m-n_2)$ such that $f(\hat{m}_1) = 0$ and $\mathrm{Obj}(m_1)$ attains the minimum at $\hat{m}_1$.

Proof. Suppose that $n_1 < m - n_2$. Observe that f is increasing in m_1 since it's the sum of increasing functions. Moreover,

$$f(n_1) = G(\theta_1, 1) - G(\theta_2, \sqrt{\frac{n_2}{m-n_1}}) = -G(\theta_2, \sqrt{\frac{n_2}{m-n_1}}) < -G(\theta_2, 1) = 0$$

$$f(m-n_2) = G(\theta_1, \sqrt{\frac{n_1}{m-n_2}}) - G(\theta_2, 1) = G(\theta_1, \sqrt{\frac{n_1}{m-n_2}}) > G(\theta_1, 1) = 0,$$

then by the intermediate value theorem, there exists $\hat{m}_1 \in (n_1, m-n_2)$ such that $f(\hat{m}_1) = 0$. Therefore, $\mathrm{Obj}(m_1)$ attains the minimum at $\hat{m}_1$. □

5.3.2 The Costs of Sending and Receiving the mass

After establishing the optimality conditions for the **One-To-Two** and **Two-To-One** cases in Section 5.3.1, we will further explore the properties of functions $\mathcal{F}$ and $\mathcal{G}$ defined in (5.55) and (5.56). This study will help establish a well-defined dual pair in Subsubsection 5.4.1.2. In this subsection, we will continue using the setting at the beginning of Section 5.3: Given any optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{nHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ with $\int c\, d\hat{\alpha} < \infty$ and $\hat{\alpha}((\mathfrak{o}, \mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{\mathsf{r}_2,2}\hat{\alpha}$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$.

Theorem 69. For any $([x, \rho_1], [y_1, 1]), ([x, \rho_2], [y_2, 1]) \in \Gamma_>$,

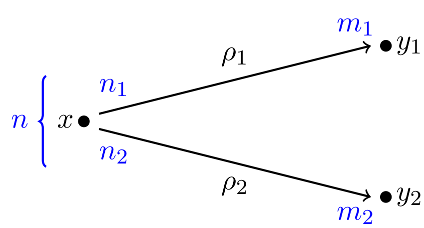
$$\mathcal{F}(x, y_1, \rho_1) = \mathcal{F}(x, y_2, \rho_2). \tag{5.75}$$

Remark 5.76. Since $\mathcal{F}(x, \cdot, \cdot)$ is independent of the target location and the initial-to-target mass ratio, we can understand $\mathcal{F}(x, \cdot, \cdot)$ as the cost of sending ρ unit of mass from x.

Proof. Assume that $\mathcal{F}(x, y_1, \rho_1) \neq \mathcal{F}(x, y_2, \rho_2)$. Let

$$m_i := \tilde{\alpha}([x, \rho_i], [y_i, 1]) = \int_{\left\{([x,r_0],[y_i,r_1]) \in \mathfrak{C}^{\otimes 2} : \frac{r_0}{r_1} = \rho_i\right\}} d\hat{\alpha},$$

and $n_i = m_i \rho_i^2$. Since $0 < \rho_1, \rho_2 < 1$, then $n := m_1 \cdot \rho_1^2 + m_2 \cdot \rho_2^2 < m_1 + m_2$.



Then we can define an objective function

$$\mathrm{Obj}(u) := c(x, \sqrt{u}, y_1, \sqrt{m_1}) + c(x, \sqrt{n-u}, y_2, \sqrt{m_2}) \quad \forall u \in [0, m_1] \cap [n - m_2, n].$$

By Theorem 67, there exists $\hat{u} \in [0, m_1] \cap [n - m_2, n]$ such that $\mathrm{Obj}(u) > \mathrm{Obj}(\hat{u})$. From that, we can define two measurable maps T_1, T_2 by

$$T_1\big(([x, \rho_1], [y_1, 1])\big) := \Big([x, \hat{\rho}_1], [y_1, 1]\Big) \quad \text{with } \hat{\rho}_1 = \sqrt{\frac{\hat{u}}{m_1}},$$

$$T_2\big(([x, \rho_2], [y_2, 1])\big) := \Big([x, \hat{\rho}_2], [y_2, 1]\Big) \quad \text{with } \hat{\rho}_2 = \sqrt{\frac{n - \hat{u}}{m_2}},$$

and we can construct a better plan α' than $\tilde{\alpha}$ by

$$\alpha' := \tilde{\alpha}\mathbb{1}_{(A_1\cup A_2)^c} + (T_1)_\#(\tilde{\alpha}\mathbb{1}_{A_1}) + (T_2)_\#(\tilde{\alpha}\mathbb{1}_{A_2}),$$

where $A_i = \big\{([x,\rho_i],[y_i,1])\big\}$. Then we obtain

$$\begin{aligned}\int c\,d(\alpha' - \tilde{\alpha}) &= \sum_{i=1}^{2}\int c\,d[(T_i)_\#(\tilde{\alpha}\mathbb{1}_{A_i})] - \int_{A_1\cup A_2} c\,d\tilde{\alpha}\\ &= \sum_{i=1}^{2}\big[c(x,\hat{\rho}_i,y_i,1) - c(x,\rho_i,y_i,1)\big]\cdot\tilde{\alpha}(x,\rho_i,y_i,1)\\ &= \sum_{i=1}^{2} c(x,\sqrt{\hat{n}_i},y,\sqrt{m_i}) - c(x,\sqrt{n_i},y_i,\sqrt{m_i}) \quad \text{with } \hat{n}_1 = \hat{u},\ \hat{n}_2 = n - \hat{u}\\ &= \mathrm{Obj}(\hat{u}) - \mathrm{Obj}(n_1) < 0,\end{aligned}$$

which contradicts the optimality of $\tilde{\alpha}$, so $\mathcal{F}(x,y_1,\rho_1) = \mathcal{F}(x,y_2,\rho_2)$. □

Theorem 70. For any $([x_1,\rho_1],[y,1]),([x_2,\rho_2],[y,1]) \in \Gamma$,

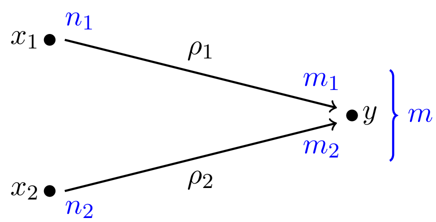
$$\mathcal{G}(y,x_1,\rho_1) = \mathcal{G}(y,x_2,\rho_2). \tag{5.77}$$

Remark 5.78. Since $\mathcal{G}(y,\cdot,\cdot)$ is independent of the initial location and the initial-to-target mass ratio, we can understand $\mathcal{G}(y,\cdot,\cdot)$ as the cost of receiving 1 unit of mass at y.

Proof. Let

$$m_i := \tilde{\alpha}([x,\rho_i],[y_i,1]) = \int_{\big\{([x,r_0],[y_i,r_1])\in\mathfrak{C}^{\otimes 2}:\frac{r_0}{r_1}=\rho_i\big\}} d\hat{\alpha}$$

and $n_i = m_i\rho_i^2$. Then $n_1 + n_2 \leqslant m := m_1 + m_2$.



If $n_1 + n_2 = m$, then $m_1 = n_1$ and $m_2 = n_2$ so that $\rho_1 = \rho_2 = 1$ and

$$\mathcal{G}(y, x_1, \rho_1) = 0 = \mathcal{G}(y, x_2, \rho_2).$$

Now, suppose that $n_1 + n_2 < m$ and assume that $\mathcal{G}(y, x_1, \rho_1) \neq \mathcal{G}(y, x_2, \rho_2)$. We can define an objective function

$$\mathrm{Obj}(u) := c(x_1, \sqrt{n_1}, y, \sqrt{u}) + c(x_2, \sqrt{n_2}, y, \sqrt{m-u}) \quad \forall u \in [n_1, m - n_2].$$

By Theorem 68, there exists $\hat{u} \in (n_1, m - n_2)$ and $\hat{u} \neq m_1$ such that $\mathrm{Obj}(u) > \mathrm{Obj}(\hat{u})$. From that, we can define two measurable maps T_1, T_2 by

$$T_1\big(([x_1, \rho_1], [y, 1])\big) := \Big([x_1, \hat{\rho}_1], [y, 1]\Big) \quad \text{with } \hat{\rho}_1 = \sqrt{\frac{n_1}{\hat{u}}},$$

$$T_2\big(([x_2, \rho_2], [y, 1])\big) := \Big([x_2, \hat{\rho}_2], [y, 1]\Big) \quad \text{with } \hat{\rho}_2 = \sqrt{\frac{n_2}{m - \hat{u}}},$$

and we can construct a better plan α' than $\tilde{\alpha}$ by

$$\alpha' := \tilde{\alpha}\mathbb{1}_{(A_1 \cup A_2)^c} + (T_1)_{\#}(\tilde{\alpha}\mathbb{1}_{A_1}) + (T_2)_{\#}(\tilde{\alpha}\mathbb{1}_{A_2}).$$

where $A_i = \big\{([x_i, \rho_i], [y, 1])\big\}$. Then we obtain

$$\begin{aligned}
\int c \, d(\alpha' - \tilde{\alpha}) &= \sum_{i=1}^{2} \int c \, d[(T_i)_{\#}(\tilde{\alpha}\mathbb{1}_{A_i})] - \int_{A_1 \cup A_2} c \, d\tilde{\alpha} \\
&= \sum_{i=1}^{2} \big[c(x_i, \hat{\rho}_i, y, 1) - c(x_i, \rho_i, y, 1)\big] \cdot \tilde{\alpha}(x_i, \rho_i, y, 1) \\
&= \sum_{i=1}^{2} c(x, \sqrt{n_i}, y, \sqrt{\hat{m}_i}) - c(x, \sqrt{n_i}, y_i, \sqrt{m_i}) \quad \text{with } \hat{m}_1 = \hat{u},\ \hat{m}_2 = m - \hat{u} \\
&= \mathrm{Obj}(\hat{u}) - \mathrm{Obj}(m_1) < 0.
\end{aligned}$$

which contradicts the optimality of $\tilde{\alpha}$, so $\mathcal{G}(y, x_1, \rho_1) = \mathcal{G}(y, x_2, \rho_2)$. □

Corollary 71. Let $\mathcal{P}^2 : \mathfrak{C} \times \mathfrak{C} \to X$ be the projection map, i.e. $\mathcal{P}^2([x, \rho], [y, 1]) = y$. Then $\mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_>) = \varnothing$ and $\mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_0) = \varnothing$. In other words, the mass transported with a ratio of 1 (balanced transfer) cannot be transported to the same target location as the mass transported with a ratio of $\rho \in (0, 1)$ (unbalanced transfer) or mass grew in place.

Proof. Given the definition of $\mathcal{G}$ in (5.56),

1. for any $([x,1],[y,1]) \in \Gamma_=$, $\mathcal{G}(y,x,1) = 0$.
2. for any $([x,\rho],[y,1]) \in \Gamma_>$, $\mathcal{G}(y,x,\rho) \in (0,1]$.
3. for any $([x,0],[y,1]) \in \Gamma_0$, $\mathcal{G}(y,x,0) = 1$.

By Theorem 70, $\mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_>) = \varnothing$ and $\mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_0) = \varnothing$. □

5.3.3 Cyclical Monotonicity-like Property

Similar to the concept of cyclic monotonicity introduced in [ABS21], we observe that the support of the optimal transport plan exhibits a similar cyclical monotonicity-like property.

Lemma 72 ([ABS21, p.31])**.** Let $X_1, \ldots, X_k$ be Polish spaces and $\gamma_i \in \mathcal{P}(X_i)$ for $i = \{1, \ldots, k\}$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable maps $T_i : \Omega \to X_i$ s.t. $(T_i)_\# \mathbb{P} = \gamma_i$ for $i = \{1, \ldots, k\}$. In this case, we can take $\Omega = \prod_{i=1}^k X_i$, $\mathcal{F}$ the product σ-algebra, and $\mathbb{P}$ the product of the γ_i.

Theorem 73. For every $N \geqslant 1$, σ permutations of $\{1, \ldots, N\}$ and $\left\{([x_i,\rho],[y_i,1])\right\}_{i=1}^N \in \Gamma$,

$$\sum_{j=1}^{N} c(x_j, \rho_j, y_{\sigma(j)}, 1) \geqslant \sum_{j=1}^{N} c(x_j, \rho_j, y_j, 1). \tag{5.79}$$

Proof. We will generalize the proof for [ABS21, Theorem 3.17]. Assume it by contradiction, then there exists $N \geqslant 1$, and a permutation σ and points $\left\{([x_i,\rho],[y_i,1])\right\}_{i=1}^N \in \Gamma$ such that

$$\sum_{j=1}^{N} c(x_j, \rho_j, y_{\sigma(j)}, 1) < \sum_{j=1}^{N} c(x_j, \rho_j, y_j, 1). \tag{5.80}$$

It suffices to find η s.t. $\int c\, d(\tilde{\alpha} + \eta) < \int c\, d\tilde{\alpha}$, which will contradict the optimality of $\tilde{\alpha}$. By the continuity of c, there exist neighbourhoods $U_j \times V_j$ of $([x_j,\rho_j],[y_j,1])$ where the inequality in (5.80) still hold. Let $z_j = \tilde{\alpha}(U_j \times V_j)$ and $\alpha_j = \frac{1}{z_j} \tilde{\alpha}\big|_{U_j \times V_j} \in \mathcal{P}(U_j \times V_j)$. Then by the Lemma 72, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable maps

$X_j \times Y_j : \Omega \to U_j \times V_j$ s.t. $\alpha_j = (X_j \times Y_j)_{\#}\mathbb{P}$. Let

$$\eta := \sum_{j=1}^{N} (X_j \times Y_{\sigma(j)})_{\#}\mathbb{P} - (X_j \times Y_j)_{\#}\mathbb{P}.$$

Since $\eta^- \leqslant \sum_{j=1}^{N} \alpha_j \leqslant \frac{N}{\min_j z_j}\tilde{\alpha}$, then let $\tilde{\eta} = c\eta$ where $c = \frac{\min_j z_j}{N}$, we have $\tilde{\eta}^- \leqslant \tilde{\alpha}$ so that $\tilde{\eta} + \tilde{\alpha} \geqslant 0$. Moreover, $\tilde{\eta}$ is null on marginals since

$$\int r_i^2 \, d\tilde{\eta} = c \sum_{j=1}^{N} \int (r_i)^2 - (r_i)^2 \, d\mathbb{P} = 0.$$

Finally, $\int c \, d(\tilde{\eta} + \tilde{\alpha}) < \int c \, d\tilde{\alpha}$ since

$$\int c \, d\tilde{\eta} = \int \sum_{j=1}^{N} \left[c(x_j, \rho_j, y_{\sigma(j)}, 1) - c(x_j, \rho_j, y_j, 1) \right] d\mathbb{P} < 0.$$

$\square$

5.4 Dual Problem

Let us define the dual problem $\mathcal{D}_b$ for the transport only such that for any $\mu^b, \nu^b \in \mathcal{M}(\mathbb{R}^d)$,

$$\mathcal{D}_b(\mu^b, \nu^b) := \sup \left\{ \int \phi_b(x) \, d\mu^b(x) + \int \psi_b(y) \, d\nu^b(y) \right\}, \tag{5.81}$$

where the supremum is taken over a set

$$I_b := \left\{ (\phi_b, \psi_b) \in \mathrm{Lip}_b(\mathbb{R}^d) \times \mathrm{Lip}_b(\mathbb{R}^d) : \phi_b(x) + \psi_b(y) \leqslant c(x, 1, y, 1) \right\}. \tag{5.82}$$

When $\mu^b(\mathbb{R}^d) = \mu^b(\mathbb{R}^d)$, the problem $\mathcal{D}_b$ coincides with the dual problem of balanced optimal transport defined in (2.2). Further, we define the dual problem $\mathcal{D}_g$ for growth such that for any $\mu^g, \nu^g \in \mathcal{M}(\mathbb{R}^d)$,

$$\mathcal{D}_g(\mu^g, \nu^g) := \sup \left\{ \int \phi_g(x) \, d\mu^g(x) + \int \psi_g(y) \, d\nu^g(y) \right\}, \tag{5.83}$$

where the supremum is taken over a set

$$I_g := \Big\{(\phi_g, \psi_g) \in \mathrm{Lip}_b(\mathbb{R}^d) \times \mathrm{Lip}_b(\mathbb{R}^d) : \phi_g(x) \cdot \rho^2 + \psi_g(y) \leqslant c(x, \rho, y, 1)$$

$$\text{for } \mu^g \otimes \nu^g\text{-a.e. and } 0 < \rho < 1\Big\}. \tag{5.84}$$

Then the combined dual problem is defined as

$$\mathcal{D}(\mu, \nu) := \sup \Big\{\mathcal{D}_b(\mu^b, \nu^b) + \mathcal{D}_g(\mu^g, \nu^g) + \big[\nu - (\nu^b + \nu^g)\big](\mathbb{R}^d)\Big\}, \tag{5.85}$$

where the supremum is taken over a set

$$\Big\{(\mu^b, \nu^b, \mu^g, \nu^g) \in \mathcal{M}(\mathbb{R}^d)^{\otimes 4} : \mu^b + \mu^g = \mu,\ \nu^b + \nu^g \leqslant \nu,\ \mu^b(\mathbb{R}^d) = \nu^b(\mathbb{R}^d)\Big\}. \tag{5.86}$$

The left-over term $[\nu - (\nu^b + \nu^g)](\mathbb{R}^d)$ is the mass which only grows in place.

Proposition 74. $\mathcal{D}(\mu, \nu) \leqslant \mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu)$.

Proof. For every α^b concentrated on $\{0 < r_0 = r_1\}$ satisfying $\mathfrak{h}^2_1 \alpha^b = \mu^b$ and $\mathfrak{h}^2_2 \alpha^b = \nu^b$, and $(\phi_b, \psi_b) \in I_b$,

$$\begin{aligned}\int \phi_b \, d\mu^b + \int \psi_b \, d\nu^b &= \int \big(\phi_b \cdot r_0^2 + \psi_b \cdot r_1^2\big) \, d\alpha^b = \int \big(\phi_b + \psi_b\big) \cdot r_1^2 \, d\alpha^b \\ &\leqslant \int c(x, 1, y, 1) \cdot r_1^2 \, d\alpha^b = \int c(x, r_0, y, r_1) \, d\alpha^b,\end{aligned}$$

which implies that

$$\mathcal{D}_b(\mu^b, \nu^b) \leqslant \int c(x, r_0, y, r_1) \, d\alpha^b.$$

For every α^g concentrated on $\{0 < r_0 = r_1\}$ satisfying $\mathfrak{h}^2_1 \alpha^g = \mu^g$ and $\mathfrak{h}^2_2 \alpha^g = \nu^g$, and $(\phi_g, \psi_g) \in I_g$,

$$\begin{aligned}\int \phi_g \, d\mu^g + \int \psi_g \, d\nu^g &= \int \big(\phi_g \cdot r_0^2 + \psi_g \cdot r_1^2\big) \, d\alpha^g = \int \Big[\phi_g \cdot (\frac{r_0}{r_1})^2 + \psi_g\Big] \cdot r_1^2 \, d\alpha^g \\ &\leqslant \int c(x, \frac{r_0}{r_1}, y, 1) \cdot r_1^2 \, d\alpha^g = \int c(x, r_0, y, r_1) \, d\alpha^g,\end{aligned}$$

which implies that

$$\mathcal{D}_g(\mu^g, \nu^g) \leqslant \int c(x, r_0, y, r_1)\, d\alpha^g.$$

For every α^o concentrated on $\{r_0 = 0\}$ satisfying $\mathfrak{h}_2^2 \alpha^o = \nu - (\nu^b + \nu^g)$,

$$\big[\nu - (\nu^b + \nu^g)\big](\mathbb{R}^d) = \int r_1^2\, d\alpha^o = \int c(x, r_0, y, r_1)\, d\alpha^o.$$

By combining all the information, we obtain

$$\begin{aligned}\mathcal{D}_b(\mu^b, \nu^b) + \mathcal{D}_g(\mu^g, \nu^g) + \big[\nu - (\nu^b + \nu^g)\big](\mathbb{R}^d) &\leqslant \int c(x, r_0, y, r_1)\, d(\alpha^b + \alpha^g + \alpha^o) \\ &= \int c(x, r_0, y, r_1)\, d\alpha.\end{aligned}$$

for all $\alpha \in \mathcal{M}(\mathfrak{C} \times \mathfrak{C})$ satisfying $\mathfrak{h}_1^2 \alpha = \mu$ and $\mathfrak{h}_2^2 \alpha = \nu$. Taking the supremum for the left and minimum for the right, the desired inequality $\mathcal{D}(\mu, \nu) \leqslant \mathbf{nHK}^2_{\mathfrak{C}}(\mu, \nu)$ will hold. □

Theorem 75 (Main Theorem, Duality)**.** For any $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$, assume that for some $R > 0$, $spt(\mu), spt(\nu) \subset B_R(0)$. Then

$$\mathcal{D}(\mu, \nu) = \mathbf{nHK}^2_{\mathfrak{C}}(\mu, \nu). \tag{Duality}$$

The Duality will be proved in Section 5.4.2. Before that, we will discuss the dual pairs of the dual problem in the next section.

5.4.1 Dual Pairs

In this subsection, we are given an optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{nHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ with $\int c\, d\hat{\alpha} < \infty$ and $\hat{\alpha}((\mathfrak{o}, \mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{r_1,2}\hat{\alpha}$ and by Subsection 5.3, the support of $\tilde{\alpha}$ is $\Gamma = \Gamma_0 \cup \Gamma_= \cup \Gamma_>$. Then we can decompose $\tilde{\alpha}$ into $\tilde{\alpha} = \alpha^b + \alpha^g + \alpha^o$ where

$$\alpha^b = \tilde{\alpha}\mathbb{1}_{\Gamma_=}, \qquad \alpha^g = \tilde{\alpha}\mathbb{1}_{\Gamma_>}, \qquad \alpha^o = \tilde{\alpha}\mathbb{1}_{\Gamma_0}.$$

Let $(\mu^b, \nu^b) = (\mathfrak{h}_1^2 \alpha^b, \mathfrak{h}_2^2 \alpha^b)$ and $(\mu^g, \nu^g) = (\mathfrak{h}_1^2 \alpha^g, \mathfrak{h}_2^2 \alpha^g)$.

5.4.1.1 Dual Pairs of Dual Problem for the Transport

In this subsection, our goal is to construct ϕ_b, ψ_b such that

$$\phi_b(x) + \psi_b(y) \leqslant c(x, 1, y, 1) \quad \forall x, y \in \mathbb{R}^d,$$

and the equality holds on $\Gamma_=$. We will generalize the construction of dual pairs for Dual problems in [ABS21, Lecture 3]. If ϕ_b, ψ_b satisfy the conditions, then for any $x \in \mathbb{R}^d$ and $([x', 1], [y', 1]) \in \Gamma_=$,

$$\begin{aligned} \phi_b(x) - \phi_b(x') &\leqslant c(x, 1, y', 1) - \psi_b(y') - [c(x', 1, y', 1) - \psi_b(y')] \\ &= c(x, 1, y', 1) - c(x', 1, y', 1). \end{aligned}$$

Fix $([x_0, 1], [y_0, 1]) \in \Gamma_=$ with $\phi_b(x_0) = 0$ and take $\{([x_j, 1], [y_j, 1])\}_{j=1}^N \in \Gamma_=$ for some $N \geqslant 1$, we can observe that

$$\begin{aligned} \phi_b(x) &= \phi_b(x) - \phi_b(x_N) + \phi_b(x_N) - \phi_b(x_{N-1}) + \ldots + \phi_b(x_1) - \phi_b(x_0) + \phi_b(x_0) \\ &\leqslant c(x, 1, y_N, 1) - c(x_N, 1, y_N, 1) + \ldots + c(x_1, 1, y_0, 1) - c(x_0, 1, y_0, 1). \end{aligned} \tag{5.87}$$

Inspiring from this upper bound, we will construct ϕ_b, ψ_b in the following theorem

Theorem 76. Fix $([x_0, 1], [y_0, 1]) \in \Gamma_=$, we define $\phi_b : \mathbb{R}^d \to \mathbb{R}^d$ as

$$\phi_b(x) := \inf \left\{ c(x, 1, y_N, 1) - c(x_N, 1, y_N, 1) + \ldots + c(x_1, 1, y_0, 1) - c(x_0, 1, y_0, 1) \right\}, \tag{5.88}$$

where the infimum is taken over all $N \geqslant 1$ and $\{([x_j, 1], [y_j, 1])\}_{j=1}^N \in \Gamma_=$ and we define $\psi_b : \mathbb{R}^d \to \mathbb{R}^d$ as

$$\psi_b(y) := \inf \left\{ c(x, 1, y, 1) - \phi_b(x) : x \in \mathbb{R}^d \right\}. \tag{5.89}$$

Then $\phi_b(x_0) = 0$, so ϕ_b is well-defined. Moreover,

$$\phi_b(x) + \psi_b(y) \leqslant c(x, 1, y, 1) \quad \forall x, y \in \mathbb{R}^d. \tag{5.90}$$

and the equality holds on $\Gamma_=$.

Proof. Choosing $N = 1$ and $(x_1, y_1) = (x_0, y_0)$, we have $\phi_b(x_0) \leqslant 0$. By Theorem 73, we have $\phi_b(x_0) \geqslant 0$. If we minimize (x_N, y_N) first, we can observe that ϕ_b in (5.88) can be rewritten as

$$\phi_b(x) = \inf_{([x_N,1],[y_N,1])} \left\{ c(x,1,y_N,1) - c(x_N,1,y_N,1) + \phi(x_N) \right\}. \tag{5.91}$$

The inequality in (5.90) is held by the construction of ψ_b. Now, take $([x',1],[y',1]) \in \Gamma_=$. From ϕ_b in (5.91), we obtain

$$c(x,1,y',1) - \phi_b(x) \geqslant c(x',1,y',1) - \phi_b(x'). \tag{5.92}$$

Take the infimum w.r.t x in the left side, we obtain $\psi_b(y') \geqslant c(x',1,y',1) - \phi_b(x')$. The other direction is obtained when we choose $x = x'$. Therefore, $\phi_b(x') + \psi_b(y') = c(x',1,y',1)$. $\square$

Corollary 77. With ϕ_b, ψ_b defined in (5.88) and (5.89) resp, if there exists $R > 0$ such that $spt(\mu^b), spt(\nu^b) \subset B_R(0)$, then ϕ_b, ψ_b are Lipschitz and bounded. Thus, by Theorem 76, $(\phi_b, \psi_b) \in I_b$ where I_b defined in (5.82) such that $\phi_b(x) + \psi_b(y) = c(x,1,y,1)$ holds on $\Gamma_=$.

Proof. We will use the equivalent formulation of ϕ_b in (5.91) to show that ϕ_b is Lipschitz. For any $x, z \in \mathbb{R}^d$, we want to show that there exists $K > 0$ such that

$$|\phi_b(x) - \phi_b(z)| \leqslant K\|x - z\|.$$

Given $\varepsilon > 0$, there exists $([x_N,1],[y_N,1]) \in \Gamma_=$ such that

$$\phi_b(z) > c(z,1,y',1) - c(x',1,y',1) + \phi_b(x') - \varepsilon,$$
$$\phi_b(x) \leqslant c(x,1,y',1) - c(x',1,y',1) + \phi_b(x').$$

Since $c(\cdot,1,y',1)$ is uniformly Lipschitz within $B_R(0)$, then there exists $K > 0$ such that

$$\phi_b(x) - \phi_b(z) = c(x,1,y',1) - c(z,1,y',1) + \varepsilon \leqslant K\|x - z\| + \varepsilon.$$

Similarly for $\phi_b(z) - \phi_b(x)$. The same techniques can be applied to show that ψ_b is Lipschitz.

Now let's prove that ϕ_b, ψ_b are bounded. The ϕ_b is bounded above due to the boundedness of c, and $\psi_b(y) \leqslant c(x,1,y,1) - \phi_b(x) \leqslant \sup c$ is also uniformly upper bounded. According to [ABS21, Theorem 3.14], we have that

$$\phi_b(x) = \inf \left\{ c(x,1,y,1) - \psi_b(y) \right\} \geqslant \inf c - \sup \psi_b,$$

so ϕ_b is bounded from below. Similarly,

$$\psi_b(y) = \inf\left\{c(x,1,y,1) - \phi_b(x)\right\} \geqslant \inf c - \sup \phi_b,$$

so ψ_b is bounded from below. $\square$

5.4.1.2 Dual Pairs of Dual Problem for the Growth

In this subsection, our goal is to construct ϕ_g, ψ_g such that

$$\phi_g(x) + \psi_g(y) \leqslant c(x,\rho,y,1) \quad \forall x,y \in \mathbb{R}^d,\ \rho \in (0,1).$$

and the equality holds on $\Gamma_>$.

Theorem 78. Given any $([x,\rho],[y,1]) \in \Gamma_>$, we define

$$\phi_g(x) := \mathcal{F}(x,y,\rho) \quad \text{and} \quad \psi_g(y) := \mathcal{F}(y,x,\rho). \tag{5.93}$$

Then for any $x \in \mathrm{spt}(\mu^g), y \in \mathrm{spt}(\nu^g)$ and $\rho \in (0,1)$,

$$\phi_g(x)\cdot\rho^2 + \psi_g(y) \leqslant c(x,\rho,y,1)$$

and the equality holds on $\Gamma_>$.

Remark 5.94. Relied on the Theorem 69, ϕ_g and ψ_g are well-defined in $spt(\mu^g)$ and $spt(\nu^g)$ resp. Due to the definition of $\mathcal{F}$ and $\mathcal{G}$, the equality inherently holds on $\Gamma_>$.

Proof. Fix any $x_1 \in spt(\mu^g)$ and $y_2 \in spt(\nu^g)$, there exist

$$([x_1,\rho_1],[y_1,1]) \quad \text{and} \quad ([x_2,\rho_2],[y_2,1]) \in \Gamma_>.$$

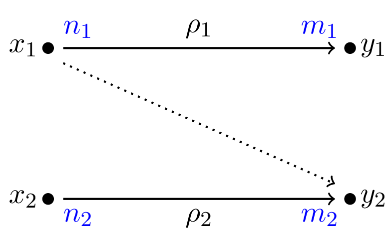
If $y_2 = y_1$, then by the Claim B.6

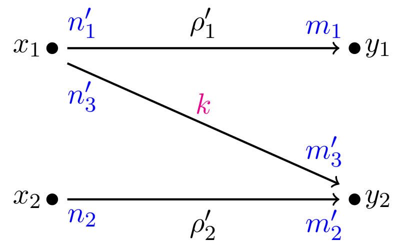
$$\phi_g(x_1)\cdot\rho^2 + \psi_g(y_2) \leqslant c(x_1,\rho,y_2,1) \quad \forall \rho \in (0,1).$$

If $y_2 \neq x_1$, we will prove the inequality by contradiction. Assume that there exists $k \in (0,1]$ such that

$$\phi_g(x_1)\cdot k^2 + \psi_g(y_2) > c(x_1,k,y_2,1) + \varepsilon.$$

Given an optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu, \nu)$ of $\mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu)$ and $\tilde{\alpha} := \mathrm{dil}_{r_1,2}\hat{\alpha}$, we want to construct a better plan α' s.t. $\int c\, d(\alpha' - \tilde{\alpha}) < 0$, which will contradict the optimality of $\tilde{\alpha}$. We will construct α' by rerouting the mass transported from x_1 to y_1. Let $m_i = \tilde{\alpha}(x_i, \rho_i, y_i, 1)$ and $n_i = \rho_i^2 m_i$.





The left figure is endowed with the original total costs as

$$\begin{aligned}\mathrm{cost}_l &= c(x_1, \sqrt{n_1}, y_1, \sqrt{m_1}) + c(x_2, \sqrt{n_2}, y_2, \sqrt{m_2}) \\ &= \phi(x_1)n_1 + \psi(y_1)m_1 + \phi(x_2)n_2 + \psi(y_2)m_2.\end{aligned}$$

In the right figure, we construct α' by rerouting the transport plan $\tilde{\alpha}$ where we

- transport from x_1 with new mass n_3' to y_2 with new mass $m_3' = k^{-2}n_3'$.
- transport from x_1 with new mass $n_1' = n_1 - n_3'$ to y_1 with mass m_1 s.t. $n_1' \leqslant m_1$.
- transport from x_2 with mass n_2 to y_2 with mass $m_2' = n_2 - m_3'$ s.t. $n_2 \leqslant m_2'$.

Let $\rho_1' = \sqrt{\frac{n_1'}{m_1}}$ and $\rho_2' = \sqrt{\frac{n_2}{m_2'}}$. Then the right figure is endowed with the total costs as

$$\begin{aligned}\mathrm{cost}_r &= c(x_1, \sqrt{n_1'}, y_1, \sqrt{m_1}) + c(x_2, \sqrt{n_2}, y_2, \sqrt{m_2'}) + c(x_2, \sqrt{n_3'}, y_1, \sqrt{m_3'}) \\ &< c(x_1, \sqrt{n_1'}, y_1, \sqrt{m_1}) + c(x_2, \sqrt{n_2}, y_2, \sqrt{m_2'}) + \big[\phi(x_2)n_3' + \psi(y_1)m_3' - \varepsilon m_3'\big].\end{aligned}$$

Then

$$\begin{aligned}\mathrm{cost}_r - \mathrm{cost}_l < \big[&c(x_1, \sqrt{n_1'}, y_1, \sqrt{m_1}) - \phi(x_1)n_1' - \psi(y_1)m_1\big] + \\ &\big[c(x_2, \sqrt{n_2}, y_2, \sqrt{m_2'}) - \phi(x_2)n_2 - \psi(y_2)m_2'\big] - \varepsilon m_3'.\end{aligned}$$

Now, let's find m_3' so that $\mathrm{cost}_r - \mathrm{cost}_l < 0$.

- Define $f(\rho) := c(x_1, \rho, y_1, 1) - \phi(x_1)\rho^2$. Then $f(\rho_1) = \psi(y_1)$ and by Claim B.6, we have $f(\rho) \geqslant f(\rho_1)$, and f is continuous in $\rho \in [0,1]$ so it is continuous in ρ^2. Since

$$\lim_{\rho\to\rho_1} \frac{f(\rho) - f(\rho_1)}{(\rho^2 - \rho_1^2)^2} = \lim_{\rho\to\rho_1} \frac{f'(\rho)}{4(\rho^2 - \rho_1^2)\rho} = \frac{f''(\rho_1)}{8\rho_1^2} < \infty,$$

then there exists $K_1 > 0$ s.t.

$$f(\rho_1') - f(\rho_1) \leqslant K_1\big[(\rho_1')^2 - \rho_1^2\big]^2.$$

- Define $g(\rho) := c(x_2, \rho, y_2, 1) - \phi(x_2)\rho^2$ where $g(\rho_2) = \psi(y_2)$. Similarly, there exists $K_2 > 0$ s.t.

$$g(\rho_2') - g(\rho_2) \leqslant K_2\big[(\rho_2')^2 - \rho_2^2\big]^2.$$

Therefore, we have

$$\begin{aligned}
\text{cost}_r - \text{cost}_l &< \big[f(\rho_1') - f(\rho_1)\big]m_1 + \big[g(\rho_2') - g(\rho_2)\big]m_2' - \varepsilon m_3' \\
&\leqslant K_1\big[(\rho_1')^2 - \rho_1^2\big]^2 m_1 + K_2\big[(\rho_2')^2 - \rho_2^2\big]^2 m_2' - \varepsilon m_3' \\
&< \big[m_3'\big(K_1 m_1^{-1} + K_2 m_2^{-1}\big) - \varepsilon\big]m_3' \\
&< -\frac{1}{2}\varepsilon m_3' < 0
\end{aligned}$$

if $m_3' < \varepsilon\big(K_1 m_1^{-1} + K_2 m_2^{-1}\big)^{-1}$. Pick small $\varepsilon > 0$, m_3' will satisfy the constraint and

$$\int c\, d(\alpha' - \hat{\alpha}) = \text{cost}_r - \text{cost}_l < -\frac{1}{2}\varepsilon < 0,$$

so we have the desired plan α', which is better than $\tilde{\alpha}$. Therefore,

$$\phi_g(x_1) \cdot \rho^2 + \psi_g(y_2) \leqslant c(x_1, \rho, y_2, 1) \quad \forall \rho \in (0,1).$$

□

We want to extend the definition of ϕ_g and ψ_g to $\mathbb{R}^d$ so that ϕ_g and ψ_g can be Lipschitz.

Definition 5.95. For any $x \in \mathbb{R}^d$,

$$\phi_g(x) := \inf_{\Gamma_>} \Big\{ c(x, 1, \tilde{y}, \tilde{r}) - c(\tilde{x}, 1, \tilde{y}, \tilde{r}) + \mathcal{F}(\tilde{x}, \tilde{y}, \tilde{\rho}) \Big\} \tag{5.96}$$

$$= \inf_{\Gamma_>} \Big\{ c(x, 1, \tilde{y}, \tilde{r}) - \mathcal{G}(\tilde{y}, \tilde{x}, \tilde{\rho}) \cdot \tilde{r}^2 \Big\}. \tag{5.97}$$

Remark 5.98. The second equality holds by the Claim 78

Claim 5.99. Given any $([\hat{x}, \hat{\rho}], [\hat{y}, 1]) \in \Gamma_>$, $\phi_g(\hat{x}) = \mathcal{F}(\hat{x}, \hat{y}, \hat{\rho})$.

Proof. It's trivial that $\phi_g(\hat{x}) \leqslant \mathcal{F}(\hat{x}, \hat{y}, \hat{\rho})$. On the other hand, by the Theorem 78.

$$c(\hat{x}, 1, \tilde{y}, \tilde{r}) - \mathcal{G}(\tilde{y}, \tilde{x}, \tilde{\rho}) \cdot \tilde{r}^2 \geqslant \mathcal{F}(\hat{x}, \hat{y}, \hat{\rho}) \quad \text{for all } ([\tilde{x}, \tilde{\rho}], [\tilde{y}, 1]) \in \Gamma_>.$$

□

Theorem 79. If there exists $R > 0$ such that $spt(\mu^g), spt(\nu^g) \subset B_R(0)$, then ϕ_g is Lipschitz.

Proof. For any $x, z \in \mathbb{R}^d$, we want to show that there exists $K > 0$ such that

$$|\phi_g(x) - \phi_g(z)| \leqslant K\|x - z\|.$$

By the definition of ϕ_g, for any $\varepsilon > 0$, there exists $([x', \rho'], [y', 1]) \in \Gamma_>$ such that

$$\phi_g(z) > c(z, 1, y', r') - \mathcal{G}(y', x', \rho') \cdot (r')^2 - \varepsilon,$$

$$\phi_g(x) \leqslant c(x, 1, y', r') - \mathcal{G}(y', x', \rho') \cdot (r')^2.$$

Since $c(\cdot, \rho, y', 1)$ is Lipschitz, then there exists $K > 0$ such that

$$\phi_g(x) - \phi_g(z) = c(x, 1, y', r') - c(z, 1, y', r') + \varepsilon \leqslant K\|x - z\| + \varepsilon.$$

Similarly for $\phi_g(z) - \phi_g(x)$. □

Definition 5.100. For any $y \in \mathbb{R}^d$,

$$\psi_g(y) := \inf\left\{c(x, \rho, y, 1) - \phi_g(x) \cdot \rho^2 : x \in \text{spt}(\mu^g), \rho \in (0, 1)\right\}. \tag{5.101}$$

Claim 5.102. Given any $([\hat{x}, \hat{\rho}], [\hat{y}, 1]) \in \Gamma_>$, $\psi_g(\hat{x}) = \mathcal{G}(\hat{y}, \hat{x}, \hat{\rho})$.

Proof. It's trivial that $\psi_g(\hat{y}) \leqslant \mathcal{G}(\hat{y}, \hat{x}, \hat{\rho})$. On the other hand, by the Theorem 78

$$c(x, \rho, \hat{y}, 1) - \phi_g(x) \cdot \rho^2 \geqslant \mathcal{G}(\hat{y}, \hat{x}, \hat{\rho}) \quad \forall x \in spt(\mu^g), \rho \in (0, 1).$$

□

Theorem 80. If there exists $R > 0$ such that $spt(\mu^g), spt(\nu^g) \subset B_R(0)$, then ψ_g is Lipschitz.

Proof. For any $y, z \in \mathbb{R}^d$, we want to show that there exists $K > 0$ such that

$$|\psi_g(z) - \psi_g(y)| \leqslant K\|z - y\|.$$

By the definition of ψ_g, for any $\varepsilon > 0$, there exists $\tilde{x} \in \mathbb{R}^d$ and $\tilde{\rho} \in (0, 1)$ such that

$$\psi_g(y) > c(\tilde{x}, \tilde{\rho}, y, 1) - \phi_g(\tilde{x}) \cdot \tilde{\rho}^2 - \varepsilon,$$

$$\psi_g(z) \leqslant c(\tilde{x}, \tilde{\rho}, z, 1) - \phi_g(\tilde{x}) \cdot \tilde{\rho}^2.$$

Since $c(\tilde{x}, \rho, \cdot, 1)$ is Lipschitz, then there exists $K > 0$ such that

$$\psi_g(z) - \psi_g(y) = c(\tilde{x}, \tilde{\rho}, z, 1) - c(\tilde{x}, \tilde{\rho}, y, 1) + \varepsilon \leqslant K\|z - y\| + \varepsilon.$$

Similarly for $\psi_g(y) - \psi_g(z)$. □

Corollary 81. When ϕ_g, ψ_g defined in (5.96) and (5.101), Theorem 78 still holds. In addition, if there exists $R > 0$ such that $spt(\mu^g), spt(\nu^g) \subset B_R(0)$, then ϕ_g, ψ_g are Lipschitz and bounded. Thus, $(\phi_g, \psi_g) \in I_g$ where I_g defined in (5.84) such that $\phi_g(x) \cdot \rho^2 + \psi_g(y) = c(x, \rho, y, 1)$ holds on $\Gamma_>$.

Proof. ϕ_g is upper bounded since it's bounded by a finite sequence of bounded cost functions. ψ_g is upper bounded since it's bounded by $\sup c$. From the definition of ϕ_g in (5.96) and the Claim 5.102, we have

$$\phi_g(x) = \inf_{\Gamma_>} \left\{ c(x, 1, \tilde{y}, \tilde{r}) - \psi_g(\tilde{y}) \cdot (\tilde{r})^2 \right\} \geqslant \inf c - \sup \psi_g.$$

so ϕ_g is lower bounded and similar for ψ_g. □

5.4.2 Duality

Theorem 82 (Main Theorem, Duality)**.** For any $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$, assume that for some $R > 0$, $spt(\mu), spt(\nu) \subset B_R(0)$. Then

$$\mathcal{D}(\mu, \nu) = \mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu). \qquad \text{(Duality)}$$

Proof. From Proposition 74, we have $\mathcal{D}(\mu,\nu) \leqslant \mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu)$. Thus, it suffices to find $(\mu^b,\nu^b,\mu^g,\nu^g)$ such that

$$\mathcal{D}_b(\mu^b,\nu^b) + \mathcal{D}_g(\mu^g,\nu^g) + \big[\nu - (\nu^b+\nu^g)\big](\mathbb{R}^d) = \mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu).$$

Given an optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu,\nu)$ of $\mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu)$ and $\tilde{\alpha} := \mathrm{dil}_{r_1,2}\hat{\alpha}$. Let $(\mu^b,\nu^b) = (\mathfrak{h}^2_1\alpha^b, \mathfrak{h}^2_2\alpha^b)$ and $(\mu^g,\nu^g) = (\mathfrak{h}^2_1\alpha^g, \mathfrak{h}^2_2\alpha^g)$ where $\alpha^b = \tilde{\alpha}\mathbb{1}_{\Gamma_=}$ and $\alpha^g = \tilde{\alpha}\mathbb{1}_{\Gamma_>}$. By Corollary 77, there exists $(\phi_b,\psi_b) \in I_b$ such that $\phi_b(x) + \psi_b(y) = c(x,1,y,1)$ on $\Gamma_=$. Thus,

$$\int \phi_b\, d\mu^b + \int \psi_b\, d\nu^b = \int_{\Gamma_=} \big(\phi_b + \psi_b\big)\, d\tilde{\alpha} = \int_{\Gamma_=} c(x,1,y,1)\, d\tilde{\alpha} = \int_{\Gamma_=} c(x,\rho,y,r_1)\, d\tilde{\alpha}.$$

By Corollary 81, there exists $(\phi_g,\psi_g) \in I_g$ where I_g defined in (5.84) such that $\phi_g(x)\cdot\rho^2 + \psi_g(y) = c(x,\rho,y,1)$ holds on $\Gamma_>$. Thus,

$$\int \phi_g\, d\mu^g + \int \psi_g\, d\nu^g = \int_{\Gamma_>} \big(\phi_g\cdot\rho^2 + \psi_g\big)\, d\tilde{\alpha} = \int_{\Gamma_>} c(x,\rho,y,r_1)\, d\tilde{\alpha}.$$

Moreover,

$$\int_{\Gamma_0} c\, d\tilde{\alpha} = \int_{\Gamma_0} r_1^2\, d\tilde{\alpha} = \left(\int_{\Gamma} - \int_{\Gamma_=\cup\Gamma_>}\right) r_1^2\, d\tilde{\alpha} = \big[\nu - (\nu^b+\nu^g)\big](\mathbb{R}^d).$$

Therefore,

$$\mathcal{D}_b(\mu^b,\nu^b) + \mathcal{D}_g(\mu^g,\nu^g) + \big[\nu - (\nu^b+\nu^g)\big](\mathbb{R}^d) = \int_{\Gamma_=\cup\Gamma_>\cup\Gamma_0} c\, d\tilde{\alpha} = \mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu).$$

□

5.4.3 Existence of Monge Maps

Theorem 83 (Rademacher, [ABS21])**.** Assume that $\Omega \subset \mathbb{R}^d$ is an open set and that $f : \Omega \to \mathbb{R}$ is locally Lipshictz. Then f is differentiable $\mathcal{L}^d$-a.e. in Ω.

Theorem 84. Given $\mu,\nu \in \mathcal{M}(\mathbb{R}^d)$. Assume that $\mu \ll \mathcal{L}^d$ and $spt(\mu), spt(\nu) \subset \Omega := B_R(0)$ for some $R > 0$. If $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu,\nu)$ with $\int c d\hat{\alpha} < \infty$ and $\hat{\alpha}((\mathfrak{o},\mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{\mathrm{r}_2,2}\hat{\alpha}$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$. By Subsection 5.3, the support of $\tilde{\alpha}$ is $\Gamma = \Gamma_0 \cup \Gamma_= \cup \Gamma_>$. Then there exists a Monge map $T_b : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\Gamma_= = \left\{([x, 1], [y, 1]) : y = T_b(x)\right\} \tag{5.103}$$

and there exist a Monge map $T_g : \mathbb{R}^d \to \mathbb{R}^d$ and a growth map $\varrho_g : \mathbb{R}^d \to (0, 1)$ such that

$$\Gamma_> = \left\{([x, \rho], [y, 1]) : y = T_g(x),\, \rho = \varrho_g(x)\right\}. \tag{5.104}$$

Remark 5.105. The set $\Gamma_=$ for balanced transfer and the set $\Gamma_>$ for unbalanced transfer both admit Monge maps respectively. However, there might exist x such that the mass was sent from x to two locations $T_b(x)$ and $T_g(x)$.

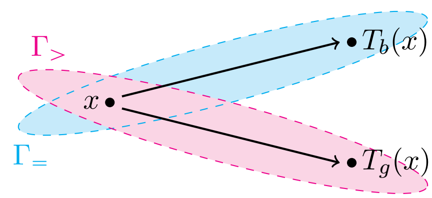


FIGURE 5.11: The support of an optimal transport plan $\tilde{\alpha}$

Proof of Theorem 84. If $\mu \ll \mathcal{L}^d$, then $\mu^b := \mathfrak{h}_1^2(\tilde{\alpha}\mathbb{1}_{\Gamma_=}) \ll \mathcal{L}^d$ and $\mu^g := \mathfrak{h}_1^2(\tilde{\alpha}\mathbb{1}_{\Gamma_>}) \ll \mathcal{L}^d$.

1. By Corollary 77, ϕ_b, ψ_b defined in (5.88) and (5.89) are Lipschitz and bounded. By Rademacher Theorem 83, ϕ_b is differentiable μ^b-a.e. in Ω. Take $x' \in \Omega$ where $\nabla_x \phi_b(x')$ exists. For any $([x, 1], [y, 1]) \in \Gamma_=$, $c(x', 1, y, 1) - \phi_b(x')$ attains the minimum at $x' = x$. By differentiation, we obtain

$$\nabla_x[c(x, 1, y, 1) - \phi_b(x)] = 0.$$

For simplicity, ∇ represents the gradient w.r.t x. Since $c(x, 1, y, 1) = \theta^2$, then we have

$$2\theta\nabla\theta - \nabla\phi_b = 0.$$

This equation provides

$$y = x + \theta\nabla\theta = x + \frac{1}{2}\nabla\phi_b := T_b(x).$$

Therefore,

$$\Gamma_= = \Big\{([x,1],[y,1]) : y = T_b(x)\Big\}.$$

2. By Corollary 81, ϕ_g, ψ_g defined in (5.96) and (5.101) are Lipschitz and bounded such that for any $([x,\rho],[y,1]) \in \Gamma_>$, $\phi_g(x) = \mathcal{F}(x,y,\rho)$. By Rademacher Theorem 83, ϕ_g is differentiable μ^g-a.e. in Ω. Now, take $x' \in \Omega$ where $\nabla_x \phi_g(x')$ exists. For any $([x,1],[y,1]) \in \Gamma_>$, $c(x',\rho,y,1) - \phi_g(x')\rho^2$ attains the minimum at $x' = x$. By differentiation, we obtain

$$\nabla_x\big[c(x,\rho,y,1) - \phi_g(x)\cdot\rho^2\big] = 0.$$

For simplicity, ∇ represents the gradient w.r.t x. Thus, given x, our goal is to find y, ρ satisfying $\phi_g(x) = \mathcal{F}(x,y,\rho)$ and $\nabla\phi_g(x) = \rho^{-2}\nabla c(x,\rho,y,1)$. Let $\mathcal{P}^1 : \mathfrak{C}\times\mathfrak{C} \to X$ be the projection map, $\mathcal{P}^1([x,\rho],[y,1]) = x$.

- For $x \in \mathcal{P}^1(\Gamma_> \cap \{\theta \leqslant \beta\})$, solving

  $$\phi_g = 1 - \rho^{-1}\cos(\theta) \quad \text{and} \quad \nabla\phi_g = 2\rho^{-1}\sin(\theta)\nabla\theta,$$

  we have that

  $$\mathbf{tan}(y-x) := \tan(\theta)\nabla\theta = \frac{\sin(\theta)\nabla\theta}{\cos(\theta)} = \frac{\nabla\phi_g}{2(1-\phi_g)}.$$

  Hence,

  $$y = x - \mathbf{arctan}\left(\frac{\nabla\phi_g}{2(\phi_g - 1)}\right) := T_g(x) \quad \text{and} \quad \varrho_g(x) := \frac{\cos(d(x,T_g(x)))}{1-\phi_g}.$$

- For $x \in \mathcal{P}^1(\Gamma_> \cap \{\theta > \beta\})$, solving

  $$\phi_g = \big(\theta - \beta + \frac{\gamma}{\rho}\big)\cdot\big(\theta-\beta\big) \quad \text{and} \quad \nabla\phi_g = 2\big(\theta-\beta+\frac{\gamma}{\rho}\big)\nabla\theta.$$

  Divide them, we can observe that

  $$\nabla\theta = \big(\theta-\beta\big)\frac{\nabla\phi_g}{2\phi_g} \implies \big(\theta-\beta\big)^2 = \frac{4\phi_g^2}{|\nabla\phi_g|^2} \implies \theta - \beta = \frac{2\phi_g}{|\nabla\phi_g|}.$$

  Substitute that back for ϕ_g, we have that

  $$\phi_g = \big(\frac{2\phi_g}{|\nabla\phi_g|} + \frac{\gamma}{\rho}\big)\cdot\frac{2\phi_g}{|\nabla\phi_g|}.$$

Solving this for ρ, we obtain that

$$\rho = \frac{1}{\sqrt{1 + (\frac{|\nabla \phi_g|}{2} - \frac{2\phi_g}{|\nabla \phi_g|})^2}} := \varrho_g(x).$$

Further, we can solve for y such that

$$y = x + \theta \nabla \theta = x + \left(\beta(\varrho_g(x)) + \frac{2\phi_g}{|\nabla \phi_g|} \right) \cdot \frac{\nabla \phi_g}{|\nabla \phi_g|} := T_g(x).$$

Therefore,

$$\Gamma_> = \left\{ ([x, \rho], [y, 1]) : y = T_g(x),\ \rho = \varrho_g(x) \right\}.$$

□

5.4.4 Uniqueness of the Optimal Transport Plan

Proposition 85. Given $\mu, \nu \in \mathcal{M}(\mathbb{R}^d)$. Assume that $\mu \ll \mathcal{L}^d$ and $spt(\mu), spt(\nu) \subset \Omega := B_R(0)$ for some $R > 0$. If $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu, \nu)$ with $\int c\, d\hat{\alpha} < \infty$ and $\hat{\alpha}((\mathfrak{o}, \mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{\mathsf{r}_2, 2}\hat{\alpha}$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$, then $\tilde{\alpha}$ is uniquely determined.

Proof. Let $\mathcal{P}^i : \mathfrak{C} \times \mathfrak{C} \to X$ be the projection maps, $i \in \{1, 2\}$ such that $\mathcal{P}^1([x, \rho], [y, 1]) = x$ and $\mathcal{P}^2([x, \rho], [y, 1]) = y$. Given two optimal solutions $\hat{\alpha}_1, \hat{\alpha}_2$ of $\mathbf{mHK}^2_{\mathfrak{C}}(\mu, \nu)$ and $\tilde{\alpha}_i := \mathrm{dil}_{\mathsf{r}_2, 2}\hat{\alpha}_i$ with their supports $\Gamma_i := spt(\tilde{\alpha}_i) = \Gamma^i_0 \cup \Gamma^i_= \cup \Gamma^i_>$. By Theorem 84, there exist $T^i_b : \mathbb{R}^d \to \mathbb{R}^d$ and $T^i_g : \mathbb{R}^d \to \mathbb{R}^d$, $\rho^i_g : \mathbb{R}^d \to (0, 1)$ such that

$$\Gamma^i_= = \left\{ ([x, 1], [y, 1]) : y = T^i_b(x) \right\},$$

$$\Gamma^i_> = \left\{ ([x, \rho], [y, 1]) : y = T^i_g(x), \rho = \rho^i_g(x) \right\}.$$

We can construct a new optimal transport plan $\alpha' = \frac{1}{2}\tilde{\alpha}_1 + \frac{1}{2}\tilde{\alpha}_2$ such that

$$\int c(x, \rho, y, 1)\, d\alpha' = \frac{1}{2} \int c(x, \rho, y, 1)\, d\tilde{\alpha}_1 + \frac{1}{2} \int c(x, \rho, y, 1)\, d\tilde{\alpha}_2 = \int c(x, \rho, y, 1)\, d\tilde{\alpha}_1$$

and its support is $\Gamma := spt(\alpha') = \Gamma_= \cup \Gamma_> \cup \Gamma_0$ where

1. $\Gamma_0 = \Gamma_0^1 \cup \Gamma_0^2$.
2. $\Gamma_= = \Gamma_=^1 \cup \Gamma_=^2$. By Theorem 84, there exists $T_b : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\Gamma_= = \left\{([x,1],[y,1]) : y = T_b(x), x \in \mathcal{P}^1(\Gamma_=^1) \cup \mathcal{P}^1(\Gamma_=^2)\right\}.$$

   Therefore, we have

$$T_b(x) = \begin{cases} T_b^1(x) & \text{if } x \in \mathcal{P}^1(\Gamma_=^1) \\ T_b^2(x) & \text{if } x \in \mathcal{P}^1(\Gamma_=^2) \end{cases}$$

   and by the one-to-one property of T_b, we have that T_b^1 and T_b^2 coincide within $\mathcal{P}^1(\Gamma_=^1) \cap \mathcal{P}^1(\Gamma_=^2)$.
3. $\Gamma_> = \Gamma_>^1 \cup \Gamma_>^2$. By Theorem 84, there exists $T_g : \mathbb{R}^d \to \mathbb{R}^d$ and $\rho_g : \mathbb{R}^d \to (0,1)$ such that

$$\Gamma_> = \left\{([x,\rho],[y,1]) : y = T_g(x), \rho = \rho_g(x)\right\}.$$

   Therefore, we have

$$T_g(x) = \begin{cases} T_g^1(x) & \text{if } x \in \mathcal{P}^1(\Gamma_>^1) \\ T_g^2(x) & \text{if } x \in \mathcal{P}^1(\Gamma_>^2) \end{cases} \quad \text{and} \quad \rho_g(x) = \begin{cases} \rho_g^1(x) & \text{if } x \in \mathcal{P}^1(\Gamma_>^1) \\ \rho_g^2(x) & \text{if } x \in \mathcal{P}^1(\Gamma_>^2) \end{cases}$$

   and by the one-to-one property of (T_g, ρ_g), we have that T_g^1, T_g^2 coincide and ρ_g^1, ρ_g^2 coincide within $\mathcal{P}^1(\Gamma_>^1) \cap \mathcal{P}^1(\Gamma_>^2)$.

Now, let us prove that $\Gamma_=^1 = \Gamma_=^2$. Pick $x \in \mathcal{P}^1(\Gamma_=^1)$ and suppose that $x \notin \mathcal{P}^1(\Gamma_=^2)$. Then $T_b(x) \in \mathcal{P}^2(\Gamma_=^1) \subseteq \mathcal{P}^2(\Gamma_=)$. By Corollary 71,

$$\mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_>) = \varnothing = \mathcal{P}^2(\Gamma_=) \cap \mathcal{P}^2(\Gamma_0),$$

so $T_b(x) \notin \mathcal{P}^2(\Gamma_> \cup \Gamma_0)$. Since $\Gamma_>^2 \cup \Gamma_0^2 \subseteq \Gamma_> \cup \Gamma_0$, then $T_b(x) \notin \mathcal{P}^2(\Gamma_>^2 \cup \Gamma_0^2)$. Hence, $T_b(x) \in \mathcal{P}^2(\Gamma_=^2)$ and there exists $z \in \mathcal{P}^1(\Gamma_=^2)$ such that $T_b(x) = T_b(z)$. By the uniqueness of T_b, $x = z$, which gives a contradiction. Therefore, $x \in \mathcal{P}^1(\Gamma_=^2)$ and $\mathcal{P}^1(\Gamma_=^1) \subseteq \mathcal{P}^1(\Gamma_=^2)$. Similarly, we can prove that $\mathcal{P}^1(\Gamma_=^2) \subseteq \mathcal{P}^1(\Gamma_=^1)$. Therefore, $\mathcal{P}^1(\Gamma_=^1) = \mathcal{P}^1(\Gamma_=^2)$. Since the target is determined by x via the monge map T_b, then $\Gamma_=^1 = \Gamma_=^2$.

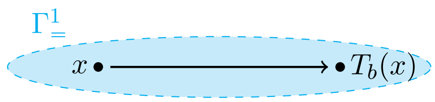


FIGURE 5.12: Transport plan $\tilde{\alpha}_1$

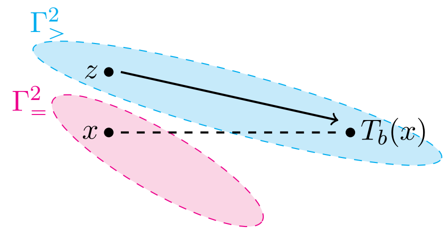


FIGURE 5.13: Transport plan $\tilde{\alpha}_2$

Now, let us prove that $\Gamma^1_> = \Gamma^2_>$. Pick $x \in \mathcal{P}^1(\Gamma^1_>)$ and suppose that $x \notin \mathcal{P}^1(\Gamma^2_>)$. Then $x \in \mathcal{P}^1(\Gamma^2_=)$ and by the previous statement, $x \in \mathcal{P}^1(\Gamma^1_=)$ as well. Observe that $T_b(x) \in \mathcal{P}^2(\Gamma_=)$ and by Corollary 70, $T_b(x) \notin \mathcal{P}^2(\Gamma_> \cup \Gamma_0)$. Hence, $T_b(x) \in \mathcal{P}^2(\Gamma^2_=)$ but $T_b(x) \notin \mathcal{P}^2(\Gamma^2_> \cup \Gamma^2_=)$. However,

$$\begin{aligned}\tilde{\alpha}_2(x, 1, T_b(x), 1) = \tilde{\alpha}_1(x, 1, T_b(x), 1) &< \tilde{\alpha}_1(x, 1, T_b(x), 1) + \rho_g(x)^2 \cdot \tilde{\alpha}_1(x, \rho_g(x), T_g(x), 1) \\ &= \tilde{\alpha}_2(x, 1, T_b(x), 1)\end{aligned}$$

gives a contradiction. Hence, $x \in \mathcal{P}^1(\Gamma^2_>)$ and $\mathcal{P}^1(\Gamma^1_>) \subseteq \mathcal{P}^1(\Gamma^2_>)$. Similarly, we can prove that $\mathcal{P}^1(\Gamma^2_>) \subseteq \mathcal{P}^1(\Gamma^1_>)$. Therefore, $\mathcal{P}^1(\Gamma^1_>) = \mathcal{P}^1(\Gamma^2_>)$ so $\Gamma^1_> = \Gamma^2_>$.

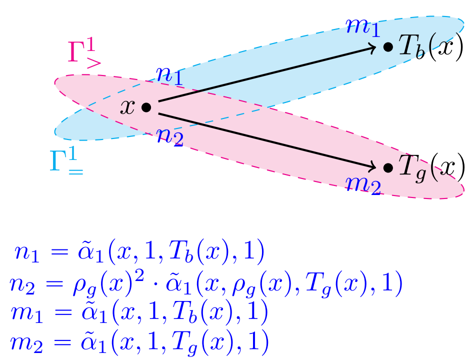


FIGURE 5.14: Transport plan $\tilde{\alpha}_1$

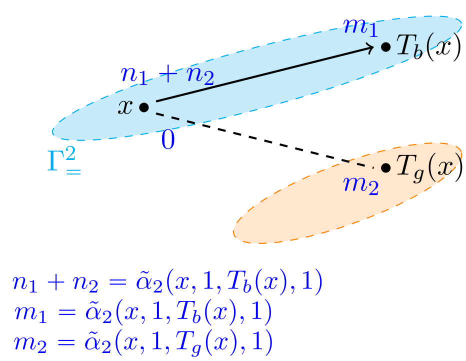


FIGURE 5.15: Transport plan $\tilde{\alpha}_2$

From the previous statement, we obtain $\mathcal{P}^2(\Gamma^1_= \cup \Gamma^1_>) = \mathcal{P}^2(\Gamma^2_= \cup \Gamma^2_>)$. Since $\mathcal{P}^2(\Gamma^1) = \mathcal{P}^2(\Gamma^2) = spt(\mu)$, then $\Gamma^1_0 = \Gamma^2_0$. Therefore, the optimal transport plan is uniquely determined after the scaling, i.e. $\tilde{\alpha}$ is unique. $\square$

5.5 Dynamic Formulation

Let $e_t : C(I;\mathfrak{C}) \to \mathfrak{C}$ be an evaluation map s.t. $e_t(\gamma) = \gamma(t) = [x(t), r(t)]$.

Definition 5.106. The **Dynamic Formulation** of the modified Wasserstein distance $W_{\tilde{d}_{\mathfrak{C}}}$ in $\mathcal{M}(\mathfrak{C})$ between two measures $\nu_1, \nu_2 \in \mathcal{M}(\mathfrak{C})$ is given by

$$Dyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2) := \min\left\{ \int_{C(I;\mathfrak{C})} \mathcal{A}_2(\gamma)\, d\eta(\gamma) : \eta \in \mathcal{P}(C(I;\mathfrak{C})), \right.$$

$$\left. (e_0)_{\#}\eta = \nu_1,\ (e_1)_{\#}\eta = \nu_2 \right\}, \tag{5.107}$$

where $\tilde{\mathcal{A}}_2 : C(I;\mathfrak{C}) \to [0,\infty]$ is the action of a curve defined as

$$\tilde{\mathcal{A}}_2(\gamma) := \begin{cases} \int_0^1 |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, dt & \text{if } \gamma \in AC(I;(\mathfrak{C},\tilde{d}_{\mathfrak{C}})), \\ +\infty & \text{else.} \end{cases} \tag{5.108}$$

Denote $OptDyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2)$ by the set of optimal dynamic plans for the $Dyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2)$.

Theorem 86. If (X,d) is a Polish and geodesic metric space, then

$$Dyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2) = W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1,\nu_2), \quad \text{where } W_{\tilde{d}_{\mathfrak{C}}} \text{ is defined in (5.29).} \tag{5.109}$$

Moreover, $\eta \in OptDyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2)$ if and only if $(e_0,e_1)_{\#}\eta \in Opt_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2)$ and η is supported in a set

$$\left\{ \gamma \in C(I;\mathfrak{C}) : \gamma = \widetilde{\Sigma}(\gamma(0),\gamma(1)),\ \text{where } \widetilde{\Sigma} \text{ defined in (5.21)} \right\}, \tag{5.110}$$

Proof. If α is an optimal transport plan of $W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1,\nu_2)$, then we can choose $\eta := \widetilde{\Sigma}_{\#}\alpha \in \mathcal{P}(C(I;\mathfrak{C}))$ such that η is concentrated on a set defined in (5.110) and

$$\int \tilde{\mathcal{A}}_2(\gamma)\, d\eta(\gamma) = \int \tilde{d}^2_{\mathfrak{C}}(\gamma(0),\gamma(1))\, d\eta(\gamma) = \int_{\mathfrak{C}\times\mathfrak{C}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0,\mathfrak{y}_1)\, d\alpha = W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1,\nu_2).$$

On the other hand, if η is an optimal dynamic plan of $Dyn_{W_{\tilde{d}_{\mathfrak{C}}}}(\nu_1,\nu_2)$, let $\alpha = (e_0,e_1)_{\#}\eta$. Then α satisfies the marginal constraints $p^i_{\#}\alpha = \nu_i$ and

$$\int_{C(I;\mathfrak{C})} \tilde{\mathcal{A}}_2(\gamma)\, d\eta(\gamma) \geqslant \int_{C(I;\mathfrak{C})} \tilde{d}^2_{\mathfrak{C}}(\gamma(0),\gamma(1))\, d\eta(\gamma)$$

$$= \int_{\mathfrak{C}\times\mathfrak{C}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0,\mathfrak{y}_1)\, d(e_0,e_1)_{\#}\eta \geqslant W^2_{\tilde{d}_{\mathfrak{C}}}(\nu_1,\nu_2).$$

The first inequality is an equality if and only if η is supported on a set in (5.110) and the second inequality is an equality if and only if $(e_0, e_1)_{\#}\eta \in Opt_{W_{\tilde{d}_{\mathfrak{C}}}}(\mu, \nu)$. □

Definition 5.111. With equality in Proposition (53), we can establish the **Dynamic Formulation** of the modified Hellinger Kantorvich problem $\mathbf{mHK}_{\mathfrak{C}}$ between two measures $\mu_1, \mu_2 \in \mathcal{M}(X)$ as

$$Dyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2) := \min\left\{ \int_{C(I;\mathfrak{C})} \tilde{\mathcal{A}}_2(\gamma)\, d\eta(\gamma) : \eta \in \mathcal{P}(C(I;\mathfrak{C})), \right.$$
$$\left. \mathfrak{h}^2 \circ (e_0)_{\#}\eta = \mu_1,\ \mathfrak{h}^2 \circ (e_1)_{\#}\eta = \mu_2 \right\}, \tag{5.112}$$

We denote $OptDyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2)$ by a set of optimal dynamic plans for $Dyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2)$.

Corollary 87. If (X, d) is a Polish and geodesic metric space, then

$$Dyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2) = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_1, \mu_2), \quad \text{where } \mathbf{mHK}_{\mathfrak{C}} \text{ is defined in (5.25).} \tag{5.113}$$

Moreover, $\eta \in OptDyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2)$ if and only if $(e_0, e_1)_{\#}\eta \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_1, \mu_2)$ and η is supported on a set in (5.110).

5.6 Absolutely Continuous Curves in a Radon Space

Corresponding to the Theorem 12, the theorem below shows that every absolutely continuous curve $\mu : I \to (\mathcal{M}(X), \mathbf{mHK}_{\mathfrak{C}})$ can be written via a dynamic plan η as $\mu_t = \mathfrak{h}^2_t\eta := \mathfrak{h}^2 \circ (e_t)_{\#}\eta$.

Theorem 88. Let (X, d) be a complete and separable metric space. Let $(\mu_t)_{t\in I}$ be a curve in $AC^2(I; (\mathcal{M}(X), \mathbf{mHK}_{\mathfrak{C}})$ with

$$\Theta := \sqrt{\mu_0(X)} + \int_0^1 |\mu'|_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt. \tag{5.114}$$

Then there exists a dynamic plan $\eta \in \mathcal{P}(AC^2(I; \mathfrak{C}))$ such that

1. $\nu_t = (e_t)_{\#}\eta$ is concentrated on $\mathfrak{C}[\Theta]$ for every $t \in I$.

2. $\mu_t = \mathfrak{h}_t^2\eta = \mathfrak{h}^2\nu_t$ in I.

3. $|\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) = |\nu'|^2_{W_{\tilde{d}_{\mathfrak{C}}}}(t) = \int_{C(I;\mathfrak{C})} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma)$ for $\mathcal{L}^1$-a.e. $t \in (0,1)$.

Proposition 89. Given $\mu_0, \mu_1 \in \mathcal{M}_2(\mathbb{R}^d)$ with compact supports. Assume that $\mu_0 \ll \mathcal{L}^d$. Then by Proposition 85, there exists a unique $\tilde{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$. Thus, there exist

(i) an optimal dynamic plan $\eta \in OptDyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$, denote by $\eta = \widetilde{\Sigma}_{\#}\tilde{\alpha}$, where $\widetilde{\Sigma}$ is defined in (5.21), and

(ii) a curve $(\mu_t)_{t\in I}$ in $C(I; \mathcal{P}_2(\mathbb{R}^d))$ joining μ_0 and μ_1, given by $\mathfrak{h}_t^2\eta$, such that

$$\mu \in AC(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}})), \quad \text{and} \quad \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) = \int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt \tag{5.115}$$

meaning that $(\mu_t)_{t\in I}$ is the minimizing curve of $\mathbf{mHK}_{\mathfrak{C}}(\mu_0, \mu_1)$.

Proof. Given any optimal transport plan $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ with $\int c\, d\hat{\alpha} < \infty$ and $\alpha((\mathfrak{o}, \mathfrak{o})) = 0$. Let $\tilde{\alpha} := \mathrm{dil}_{\mathsf{r}_2, 2}\hat{\alpha}$, where r_2 is defined in (3.10) such that $r_2([x, r_0], [y, r_1]) = r_1$. Due to Corollary 87, we can verify that $\eta := \widetilde{\Sigma}_{\#}\tilde{\alpha} \in OptDyn_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ by

$$\int_{spt(\eta)} \tilde{\mathcal{A}}_2(\gamma)\, d\eta(\gamma) = \int_{spt(\eta)} \tilde{d}^2_{\mathfrak{C}}(\gamma(0), \gamma(1))\, d\eta(\gamma) = \int_{\mathfrak{C}\times\mathfrak{C}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d\tilde{\alpha} = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1).$$

Let $\mu_t = \mathfrak{h}_t^2\eta$. Since

$$\begin{aligned}
\mathbf{mHK}_{\mathfrak{C}}(\mu_t, \mu_{t+h}) = \mathbf{mHK}_{\mathfrak{C}}(\mathfrak{h}_t^2\eta, \mathfrak{h}_{t+h}^2\eta) &\leqslant \int \tilde{d}_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d(e_t, e_{t+h})_{\#}\eta(\mathfrak{y}_0, \mathfrak{y}_1) \\
&= \int_{spt(\eta)} \tilde{d}_{\mathfrak{C}}(e_t(\gamma), e_{t+h}(\gamma))\, d\eta(\gamma) \leqslant \int_{spt(\eta)} \int_t^{t+h} |\gamma'|_{\tilde{d}_{\mathfrak{C}}}(s)\, ds\, d\eta(\gamma) \\
&\leqslant \int_t^{t+h} \left| \int_{spt(\eta)} |\gamma'|_{\tilde{d}_{\mathfrak{C}}}(s)\, d\eta(\gamma) \right| ds.
\end{aligned}$$

and by Jensen's inequality,

$$\int_0^1 \int_{spt(\eta)} |\gamma'|_{\tilde{d}_{\mathfrak{C}}}(s)\, d\eta(\gamma)\, ds \leqslant \left(\int_0^1 \int_{spt(\eta)} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(s)\, d\eta(\gamma)\, ds \right)^{\frac{1}{2}} = \mathbf{mHK}_{\mathfrak{C}}(\mu_0, \mu_1) < +\infty,$$

then $\mu \in AC(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}}))$. Moreover, since

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_t, \mu_{t+h}) = \mathbf{mHK}^2_{\mathfrak{C}}(\mathfrak{h}^2_t \eta, \mathfrak{h}^2_{t+h}\eta) \leqslant \int \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d(e_t, e_{t+h})_{\#}\eta(\gamma)$$

$$= \int_{spt(\eta)} \tilde{d}^2_{\mathfrak{C}}(e_t(\gamma), e_{t+h}(\gamma))\, d\eta(\gamma) = \int_{spt(\eta)} \left(\int_t^{t+h} |\gamma'|_{\tilde{d}_{\mathfrak{C}}}(s)\, ds \right)^2 d\eta(\gamma).$$

then we can compute

$$|\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) = \lim_{h \to 0} \frac{\mathbf{mHK}^2_{\mathfrak{C}}(\mu_t, \mu_{t+h})}{h^2} \leqslant \int_{spt(\eta)} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma).$$

Therefore,

$$\int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt \leqslant \int_0^1 \int |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma) = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1).$$

By the absolute continuity of μ_t, we have

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) \leqslant \int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt.$$

Therefore,

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) = \int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt.$$

□

Corollary 90. Continuing with Proposition 89, let $(\mu_0^b, \mu_1^b) := (\mathfrak{h}^2_1 \alpha^b, \mathfrak{h}^2_2 \alpha^b)$ and $(\mu_0^g, \mu_1^g) := (\mathfrak{h}^2_1 \alpha^g, \mathfrak{h}^2_2 \alpha^g)$ where $\alpha^b = \tilde{\alpha} \mathbb{1}_{\Gamma_=}$ and $\alpha^g = \tilde{\alpha} \mathbb{1}_{\Gamma_>}$. Then there exist $X_t^b : \mathbb{R}^d \to \mathbb{R}^d$, $X_t^g : \mathbb{R}^d \to \mathbb{R}^d$, and $\lambda_t^g : \mathbb{R}^d \to (1, \infty)$ such that

$$\mu_t = \mathfrak{h}^2_t(\widetilde{\Sigma}_{\#}\tilde{\alpha}) = \hat{\mu}_t + t^2 \cdot \left[\mu_1 - \hat{\mu}_1\right], \quad \text{where } \hat{\mu}_t := (X_t^b)_{\#}\mu_0^b + (X_t^g)_{\#}(\lambda_t^g \mu_0^g). \tag{5.116}$$

Proof. Now, let us prove the explicit formulation of μ_t. By Theorem 84, there exists a Monge map $T_b : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\Gamma_= = \left\{ ([x, 1], [y, 1]) : y = T_b(x) \right\},$$

and there exist a Monge map $T_g : \mathbb{R}^d \to \mathbb{R}^d$ and a growth map $\varrho_g : \mathbb{R}^d \to (0,1)$ such that

$$\Gamma_> = \Big\{([x,\rho],[y,1]) : y = T_g(x),\ \rho = \varrho_g(x)\Big\}.$$

Then we can define

$$X_t^b(x) := \mathsf{x} \circ e_t \circ \widetilde{\Sigma}([x,1],[y,1]) = \mathsf{x} \circ e_t([x,1],[T_b(x),1])$$

such that $X_0^b(x) = x$ and $X_1^b(x) = T_b(x)$. Therefore, from Subsection 5.1.1 we obtain

$$X_t^b(x) = (1-t)x + tT_b(x).$$

Moreover, let

$$\tilde{t} := \frac{d(x,T_g(x)) - \cos^{-1}(\rho_g(x))}{d(x,T_g(x)) - \cos^{-1}(\rho_g(x)) + \sqrt{r_g(x)^2 - 1}},$$

where $r_g : \mathbb{R}^d \to (1,\infty)$ such that $r_g \circ \rho_g = id$. Then we can define

$$\begin{aligned}
X_t^g &:= \mathsf{x} \circ e_t \circ \tilde{\Sigma}([x,\rho],[y,1]) = \mathsf{x} \circ e_t([x,\rho_g(x)],[T_g(x),1]),\\
\rho_t^g &:= \mathsf{r} \circ e_t \circ \tilde{\Sigma}([x,\rho],[y,1]) = \mathsf{r} \circ e_t([x,\rho_g(x)],[T_g(x),1]),\\
\lambda_t^g &:= (\rho_t^g(x) \cdot r_g(x))^2,
\end{aligned}$$

such that $X_0^g(x) = x, X_1^g(x) = T_g(x)$ and $\lambda_0^g(x) = 1, \lambda_1^g(x) = r_g(x)^2$. Therefore, from Subsection 5.1.1,

1. If $t \in \big[0,\tilde{t}\big]$,

$$\begin{aligned}
\lambda_t^g(x) &:= 1,\\
X_t^g(x) &:= x + t\Bigg[d(x,T_g(x)) - \cos^{-1}(\rho_g(x)) + \sqrt{r_g(x)^2 - 1}\Bigg].
\end{aligned}$$

2. If $t \in (\tilde{t},1]$, let $s = \frac{t-\tilde{t}}{1-\tilde{t}} \in (0,1]$, then

$$\begin{aligned}
\lambda_t^g(x) &:= (1-s^2) + s^2 r_g(x)^2,\\
X_t^g(x) &:= x + \frac{T_g(x) - x}{d(x,T_g(x))} \cdot \left[d(x,T_g(x)) - \cos^{-1}(\rho_g(x)) + \cos^{-1}\left(\frac{1}{\sqrt{\lambda_t^g(x)}}\right)\right].
\end{aligned}$$

We can verify that $\mu_1^b = (X_1^b)_\# \mu_0^b$ and $\mu_1^g = (X_1^g)_\# (\lambda_1^g \mu_0^g)$.

1. $\mu_1^b = (X_1^b)_\# \mu_0^b$ since for any $\varphi \in \mathcal{B}(\mathbb{R}^d)$,

$$\int \varphi(y)\, d\mu_1^b = \int \varphi(y)\, d(\mathfrak{h}_2^2 \alpha^b) = \int \varphi(\mathsf{x}_2([x,1],[T_b(x),1])) \cdot \mathsf{r}_2([x,1],[T_b(x),1])^2\, d\alpha^b$$
$$= \int \varphi(T_b(x))\, d\alpha^b$$
$$= \int \varphi \circ T_b(\mathsf{x}_1([x,1],[T_b(x),1])) \cdot \mathsf{r}_1([x,1],[T_b(x),1])^2\, d\alpha^b$$
$$= \int \varphi(T_b(x))\, d(\mathfrak{h}_1^2 \alpha^b) = \int \varphi(X_1^b(x))\, d\mu_0^b = \int \varphi(y)\, d[(X_1^b)_\# \mu_0^b].$$

2. $\mu_1^g = (X_1^g)_\# (\lambda_1^g \mu_0^g)$ since for any $\varphi \in \mathcal{B}(\mathbb{R}^d)$,

$$\int \varphi(y)\, d\mu_1^g = \int \varphi(y)\, d(\mathfrak{h}_2^2 \alpha^g)$$
$$= \int \varphi(\mathsf{x}_2([x,\rho_g(x)],[T_g(x),1])) \cdot \mathsf{r}_2([x,\rho_g(x)],[T_g(x),1])^2\, d\alpha^g$$
$$= \int \varphi(T_g(x))\, d\alpha^g = \int \varphi(T_g(x)) \cdot r_g(x)^2 \cdot \rho_g(x)^2\, d\alpha^g$$
$$= \int \varphi \circ T_g(\mathsf{x}_1([x,\rho_g(x)],[T_g(x),1])) \cdot r_g(x)^2 \cdot \mathsf{r}_1([x,\rho_g(x)],[T_g(x),1])^2\, d\alpha^g$$
$$= \int \varphi(T_g(x)) \cdot r_g(x)^2\, d(\mathfrak{h}_1^2 \alpha^g) = \int \varphi(X_1^g(x)) \cdot \lambda_1^g(x)\, d\mu_0^g$$
$$= \int \varphi(y)\, d[(X_1^g)_\# (\lambda_1^g \mu_0^g)].$$

Now, let us show that

$$\mu_t = \mathfrak{h}_t^2(\widetilde{\Sigma}_\# \tilde{\alpha}) = \hat{\mu}_t + t^2 \cdot \big[\mu_1 - \hat{\mu}_1\big], \quad \text{where } \hat{\mu}_t := (X_t^b)_\# \mu_0^b + (X_t^g)_\# (\lambda_t^g \mu_0^g).$$

1. For any $\phi \in \mathcal{B}(\mathbb{R}^d)$,

$$\int \phi(x)\, d[\mathfrak{h}_t^2 \widetilde{\Sigma}_\# (\tilde{\alpha} \mathbb{1}_{\Gamma_=})] = \int_{\Gamma_=} \phi(\mathsf{x} \circ e_t(\widetilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))) \cdot \mathsf{r} \circ e_t(\widetilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))^2\, d\tilde{\alpha}$$
$$= \int_{\Gamma_=} \phi(\mathsf{x}([X_t^b(x),1]) \cdot \mathsf{r}([X_t^b(x),1])^2\, d\tilde{\alpha}([x,1],[T_b(x),1])$$
$$= \int_{\Gamma_=} \phi(X_t^b(x))\, d\tilde{\alpha} = \int \phi(X_t^b(x))\, d\mu_0^b = \int \phi(x)\, d[(X_t^b)_\# \mu_0^b].$$

2. For any $\phi \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
&\int \phi(x)\, d[\mathfrak{h}_t^2 \widetilde{\Sigma}_\#(\tilde{\alpha}\mathbb{1}_{\Gamma_>})] \\
&\qquad = \int_{\Gamma_>} \phi(\mathsf{x} \circ e_t(\widetilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))) \cdot \mathsf{r} \circ e_t(\widetilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))^2\, d\tilde{\alpha} \\
&\qquad = \int_{\Gamma_>} \phi(\mathsf{x}([X_t^g(x), \rho_t^g(x)])) \cdot \mathsf{r}([X_t^g(x), \rho_t^g(x)])^2\, d\tilde{\alpha}([x, \rho_g(x)], [T_g(x), 1]) \\
&\qquad = \int_{\Gamma_>} \phi(X_t^g(x)) \cdot (\rho_t^g(x))^2\, d\tilde{\alpha} = \int_{\Gamma_>} \phi(X_t^g(x)) \cdot \lambda_t^g(x) \cdot (\rho_t^g(x))^2\, d\tilde{\alpha} \\
&\qquad = \int \phi(X_t^g(x)) \cdot \lambda_t^g(x)\, d\mu_0^g = \int \phi(x)\, d[(X_t^g)_\#(\lambda_t^g \mu_0^g)].
\end{aligned}$$

3. For any $\phi \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\int \phi(x)\, d[\mathfrak{h}_t^2 \widetilde{\Sigma}_\#(\tilde{\alpha}\mathbb{1}_{\Gamma_0})](x) &= \int_{\Gamma_0} \phi(\mathsf{x}(\mathfrak{o}, [y, t])) \cdot \mathsf{r}(\mathfrak{o}, [y, t])^2\, d\tilde{\alpha}(\mathfrak{o}, [y, 1]) \\
&= \int_{\Gamma_0} \phi(y) \cdot t^2\, d\tilde{\alpha}(\mathfrak{o}, [y, 1]) = \int \phi(y)\, d(t^2[\mu_1 - \hat{\mu}_1])(y) \\
&= \int \phi(x)\, d(t^2[\mu_1 - \hat{\mu}_1])(x).
\end{aligned}$$

□

5.7 Equivalent Characterization of mHK Problem

Theorem 91 (Disintegration Theorem, [ABS21, p.16])**.** Let Z, X be Polish spaces, $\sigma \in \mathcal{M}_+(Z)$, $f : Z \to X$ a Borel function and set $\theta := f_\#\sigma \in \mathcal{M}_+(X)$. Then there exists a family $\{\sigma_x\}_{x \in X} \subset \mathcal{P}(Z)$ such that

(i) $x \mapsto \sigma_x$ is Borel, i.e. $\sigma_x(A)$ is Borel for all $A \in \mathcal{B}(Z)$.

(ii) $\sigma = \int_X \sigma_x\, d\theta$, i.e. $\sigma(A) = \int_X \sigma_x(A)\, d\theta(x)$ for all $A \in \mathcal{B}(Z)$.

(iii) σ_x is concentrated on $f^{-1}(x)$ for θ-a.e. $x \in X$.

Any other family $\{\sigma'_x\}_{x \in X} \subset \mathcal{P}(Z)$ with these properties satisfies $\sigma'_x = \sigma_x$ for θ-a.e. $x \in X$.

Proposition 92 (Main Proposition, Existence of $\mathbf{mHK}^2$)**.** We have the equivalence between the $\mathbf{mHK}^2$ problem in (4.2) and the $\mathbf{mHK}^2_{\mathfrak{C}}$ problem in (5.25), i.e. $\mathbf{mHK}^2 = \mathbf{mHK}^2_{\mathfrak{C}}$. Moreover, the minimum in $\mathbf{mHK}^2$ is attained.

Fix the time interval $I = [0,1]$. Let $\tilde{\Sigma}$ be defined in (5.21) and the evaluation map e_t be defined in Section 3.4 such that

$$(\mathfrak{y}_0, \mathfrak{y}_1) \xmapsto{\tilde{\Sigma}} \bar{\mathfrak{y}} \xmapsto{e_t} \bar{\mathfrak{y}}(t) \quad \text{where} \quad \bar{\mathfrak{y}}(t) \xmapsto{\mathsf{x}} \bar{x}(t) \quad \text{and} \quad \bar{\mathfrak{y}}(t) \xmapsto{\mathsf{r}} \bar{r}(t).$$

Furthermore, we can define $\mathsf{X}_t : C(I;\mathfrak{C}) \to \mathbb{R}^d$ and $\mathsf{R}_t : C(I;\mathfrak{C}) \to \mathbb{R}_+$ by $\mathsf{X}_t := \mathsf{x} \circ e_t$ and $\mathsf{R}_t := \mathsf{r} \circ e_t$ respectively such that

$$\mathfrak{y} \xmapsto{X_t} x(t) \quad \text{and} \quad \mathfrak{y} \xmapsto{R_t} r(t).$$

Lemma 93. Given a dynamic plan $\eta \in \mathcal{P}(C(I;\mathfrak{C}))$. For almost every $t \in I$, there exist $\mu_t : \mathbb{R}^d \to \mathcal{M}(\mathbb{R}^d)$, $v_t : \mathbb{R}^d \to \mathbb{R}^d$, and $w_t : \mathbb{R}^d \to \mathbb{R}_+$ such that

$$v_t, w_t \in L^2(\mu_t) \quad \text{and} \quad \int_{\mathbb{R}^d} |(v_t, \frac{1}{2}w_t)|^2 \, d\mu_t \leqslant \int_{C(I,\mathfrak{C})} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t) \, d\eta(\gamma) \quad \mathcal{L}^1\text{-a.e. } t \in (0,1). \tag{5.117}$$

Moreover, (μ, v, w) satisfies the constraints in (4.3).

Proof. For $t \in I$, we can define a scaled dynamic plan $\sigma_t := R_t^2\eta$, its projection $\mu_t := (\mathsf{X}_t)_\# \sigma_t \in \mathcal{M}(\mathbb{R}^d)$, and $u_t : C(I;\mathfrak{C}) \to \mathbb{R}^d \times \mathbb{R}_+$ by

$$u_t(\gamma) := \begin{cases} \left(\dot{x}(t), \frac{\dot{r}(t)}{r(t)}\right) & \text{if } \gamma \in AC(I;(\mathfrak{C}, \tilde{d}_{\mathfrak{C}})), \\ (0,0) & \text{else.} \end{cases} \tag{5.118}$$

For each $t \in I$, applying the disintegration theorem 91 with respect to x, there exists a family $\{\sigma_t^x\}_{x \in X} \subset \mathcal{P}(C(I;\mathfrak{C}))$ such that $\sigma_t = \int \sigma_t^x \, d\mu_t$. Then we can construct v_t, w_t such that

$$(v_t, \frac{1}{2}w_t) = \int_{C(I;\mathfrak{C})} |u_t(\gamma)|^2 \, d\sigma_t^x(\gamma). \tag{5.119}$$

Since $\dot{r}(t) \geqslant 0$, then $w \geqslant 0$. By Jensen's inequality, we have

$$\int_{\mathbb{R}^d} |(v_t, \frac{1}{2}w_t)|^2 \, d\mu_t \leqslant \int_{\mathbb{R}^d} \int_{C(I;\mathfrak{C})} |u_t(\gamma)|^2 \, d\sigma_t^x(\gamma) \, d\mu_t(x) = \int_{C(I;\mathfrak{C})} |u_t(\gamma)|^2 \, d\sigma_t(\gamma)$$

$$= \int_{C(I;\mathfrak{C})} |u_t(\gamma) \cdot R_t(\gamma)|^2 \, d\eta(\gamma) = \int_{AC(I;(\mathfrak{C},\tilde{d}_{\mathfrak{C}}))} r(t)^2 |\dot{x}(t)|^2 + |\dot{r}(t)|^2 \, d\eta(\gamma)$$

$$= \int_{C(I;\mathfrak{C})} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t) \, d\eta.$$

so $v_t, w_t \in L^2(\mu_t)$. We can show that (μ, v, w) satisfies the continuity equation in (4.3). For any $\phi \in C_c^\infty(X)$,

$$\frac{d}{dt} \int_X \phi(x) \, d\mu_t = \frac{d}{dt} \int \phi(X_t) \cdot R_t^2 \, d\eta = \frac{d}{dt} \int \phi(\bar{x}(t)) \cdot \bar{r}(t)^2 \, d\eta$$

$$= \int \Big[\nabla_x \phi(\bar{x}(t)) \cdot \dot{\bar{x}}(t) \, \bar{r}(t)^2 + 2\phi(\bar{x}(t)) \, \bar{r}(t) \, \dot{\bar{r}}(t) \Big] d\eta$$

$$= \int (\nabla_x \phi(X_t), 2\phi(X_t)) \cdot u_t \, d\sigma_t = \int \Big[\int (\nabla_x \phi(X_t), 2\phi(X_t)) \cdot u_t \, d\sigma_t^x \Big] \, d\mu_t$$

$$= \int (\nabla_x \phi, 2\phi) \cdot (v_t, \frac{1}{2}w_t) \, d\mu_t = \int \nabla_x \phi \cdot v_t + \phi w_t \, d\mu_t.$$

We have the desired result from Proposition 4. □

Proof of $\leqslant$ in Proposition 92. Let us prove $\mathbf{mHK}^2 \leqslant \mathbf{mHK}^2_{\mathfrak{C}}$. Assume that $\hat{\alpha}$ is an optimal solution of $\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1)$. Then by Proposition 89, we can construct a optimal dynamic plan η by $\eta := \widetilde{\Sigma}_\# \hat{\alpha} \in \mathcal{P}(C(I; \mathfrak{C}))$. By Lemma 93, there exists (μ, v, w) satisfying the constraints in (4.3) such that $v_t, w_t \in L^2(\mu_t)$ and

$$\int_X |(v_t, \frac{1}{2}w_t)|^2 \, d\mu_t \leqslant \int |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t) \, d\eta(\gamma) = \int \Big(\bar{r}^2(t) |\dot{\bar{x}}|_d^2(t) + |\dot{\bar{r}}|^2(t) \Big) \, d\hat{\alpha}.$$

Therefore,

$$\int_0^1 \int_X |(v_t, \frac{1}{2}w_t)|^2 \, d\mu_t \, dt \leqslant \int_0^1 \int \Big(\bar{r}^2(t) |\dot{\bar{x}}|_d^2(t) + |\dot{\bar{r}}|^2(t) \Big) \, d\hat{\alpha} \, dt = \int_{\mathfrak{C}^{\otimes 2}} \tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \, d\hat{\alpha}.$$

□

Lemma 94. Let (ρ_ε) be a sequence of mollifiers (e.g. $\rho_\varepsilon(x) = \frac{1}{\sqrt{(2\pi\varepsilon^d)}} e^{-\frac{|x|^2}{2\varepsilon}}$), then

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu) = \lim_{\varepsilon\to 0} \mathbf{mHK}^2_{\mathfrak{C}}(\mu * \rho_\varepsilon, \nu * \rho_\varepsilon).$$

Proof of Proposition 94 . By previous proposition, there exist $\alpha \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu,\nu)$ and $\alpha_\varepsilon \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu * \rho_\varepsilon, \nu * \rho_\varepsilon)$ such that

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu) = \int_{\mathfrak{C}[R]} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0,\mathfrak{y}_1)\, d\alpha \quad \text{and} \quad \mathbf{mHK}^2_{\mathfrak{C}}(\mu * \rho_\varepsilon, \nu * \rho_\varepsilon) = \int_{\mathfrak{C}[R]} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0,\mathfrak{y}_1)\, d\alpha_\varepsilon.$$

Suppose that α_ε does not converge to α w.r.t the weak topology. For any $\varphi \in C_b(X)$, there exists $\varepsilon > 0$ such that for all $\delta > 0$, there exists an α_ε such that

$$\left| \int_X \varphi(x)\, d\big(\mu - \mu * \rho_\varepsilon\big) \right| = \left| \int_{\mathfrak{C}[R]} \varphi(\mathsf{x}_1) \cdot \mathsf{r}_1^2\, d\big(\alpha - \alpha_\varepsilon\big) \right| > \delta.$$

As $\varepsilon \to 0$, since $\mu * \rho_\varepsilon \to \mu$ weakly, the leftmost term goes to 0, which gives a contradiction. Therefore,

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu,\nu) = \lim_{\varepsilon\to 0} \int_{\mathfrak{C}[R]} \tilde{d}^2_{\mathfrak{C}}\, d\alpha_\varepsilon = \lim_{\varepsilon\to 0} \mathbf{mHK}^2_{\mathfrak{C}}(\mu * \rho_\varepsilon, \nu * \rho_\varepsilon).$$

□

Proof of $\geqslant$ ***in Proposition 92***. Now, let us prove that $\mathbf{mHK}^2 \geqslant \mathbf{mHK}^2_{\mathfrak{C}}$. Given $\delta > 0$, suppose that $(\tilde{\mu}_t, \tilde{v}_t, \tilde{w}_t)$ is a triplet such that

$$\mathcal{J}(\tilde{\mu}_t, \tilde{v}_t, \tilde{w}_t) \leqslant \mathbf{mHK}^2(\mu_0,\mu_1) + \delta.$$

1. If $(\tilde{v}_t, \tilde{w}_t)$ satisfies conditions in the first position of (5), then there exists X_t and λ_t such that $\tilde{\mu}_t = (X_t)_\#(\lambda_t\mu_0)$ given in (5). Therefore, by Jensen's inequality,

$$\begin{aligned}\mathcal{J}(\tilde{\mu}_t, \tilde{v}_t, \tilde{w}_t) &= \int_0^1 \int_X \big|\big(\tilde{v}_t, \frac{1}{2}\tilde{w}_t\big)\big|^2\, d\tilde{\mu}_t\, dt = \int_0^1 \int_X \big|\big(\tilde{v}_t(X_t), \frac{1}{2}\tilde{w}_t(X_t)\big)\big|^2\, \lambda_t\, d\mu_0\, dt \\ &= \int_0^1 \int_X \big|\big(\frac{d}{dt}X_t, \frac{1}{2}\frac{d}{dt}\log(\lambda_t)\big)\big|^2\, \lambda_t\, d\mu_0\, dt \\ &= \int_0^1 \int_X \big|\big(\sqrt{\lambda_t}\frac{d}{dt}X_t, \frac{d}{dt}\sqrt{\lambda_t}\big)\big|^2 d\mu_0\, dt \\ &\geqslant \int_X \left[\int_0^1 \big|\big(\sqrt{\lambda_t}\frac{d}{dt}X_t, \frac{d}{dt}\sqrt{\lambda_t}\big)\big|\, dt \right]^2 d\mu_0.\end{aligned}$$

Since $\frac{\dot{\lambda}_t}{\lambda_t} = \tilde{w}_t(X_t) \geqslant 0$, then $\dot{\lambda}_t \geqslant 0$. Thus, by the definition of $\tilde{d}_{\mathfrak{C}}$ in (5.3),

$$\tilde{d}_{\mathfrak{C}}([X_0(x_0), \sqrt{\lambda_0(x_0)}], [X_1(x_0), \sqrt{\lambda_1(x_0)}]) \leqslant \int_0^1 \left|\left(\sqrt{\lambda_t}\frac{d}{dt}X_t, \frac{d}{dt}\sqrt{\lambda_t}\right)\right| dt. \tag{5.120}$$

Defining $\alpha = ([X_0, \sqrt{\lambda_0}], [X_1, \sqrt{\lambda_1}])_{\#}\mu_0$, we would have that

$$\begin{aligned}\int_{\mathfrak{C}^{\otimes 2}} \tilde{d}^2_{\mathfrak{C}}(\mathfrak{y}_0, \mathfrak{y}_1)\, d\alpha &= \int_X \tilde{d}^2_{\mathfrak{C}}\big([X_0(x_0), \sqrt{\lambda_0(x_0)}], [X_1(x_0), \sqrt{\lambda_1(x_0)}]\big)\, d\mu_0 \\ &\leqslant \mathcal{J}(\tilde{\mu}_t, \tilde{v}_t, \tilde{w}_t) \leqslant \mathbf{mHK}^2(\mu_0, \mu_1) + \delta.\end{aligned}$$

Moreover, α satisfies the homogeneous marginals $\mathfrak{h}^2_1\alpha = \mu_0$ and $\mathfrak{h}^2_2\alpha = \mu_1$ hold. More specifically,

(a) for any Borel $\phi \in \mathcal{B}(X)$,

$$\begin{aligned}\int_X \phi(x)\, d\mathfrak{h}^2_1\alpha &= \int_{\mathfrak{C}^{\otimes 2}} \phi(\mathsf{x}_1(\mathfrak{y})) \cdot \mathsf{r}^2_1(\mathfrak{y})\, d\alpha = \int_X \phi(X_0(x)) \cdot \lambda_0(x)\, d\mu_0 \\ &= \int_X \phi(x)\, d[(X_0)_{\#}(\lambda_0\mu_0)] = \int_X \phi(x_0)\, d\mu_0.\end{aligned}$$

(b) for any Borel $\varphi \in \mathcal{B}(X)$,

$$\begin{aligned}\int_X \varphi(y)\, d(\mathfrak{h}^2_2\alpha) &= \int_{\mathfrak{C}^{\otimes 2}} \varphi(\mathsf{x}_2(\mathfrak{y})) \cdot \mathsf{r}^2_2(\mathfrak{y})\, d\alpha = \int_X \varphi(X_1(x)) \cdot \lambda_1(x)\, d\mu_0 \\ &= \int_X \varphi(x)\, d[(X_1)_{\#}(\lambda_1\mu_0)] = \int \varphi(y)\, d\mu_1.\end{aligned}$$

2. If $\tilde{v}_t, \tilde{w}_t$ satisfies conditions in the second position of (5), then there exists $(\mu^\varepsilon_t, v^\varepsilon_t, w^\varepsilon_t) \in \mathcal{CE}^+(\mu_0 * \rho_\varepsilon, \mu_1 * \rho_\varepsilon)$ such that

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0 * \rho_\varepsilon, \mu_1 * \rho_\varepsilon) \leqslant \mathcal{J}(\mu^\varepsilon_t, v^\varepsilon_t, w^\varepsilon_t) \leqslant \mathcal{J}(\tilde{\mu}_t, \tilde{v}_t, \tilde{w}_t) \leqslant \mathbf{mHK}^2(\mu_0, \mu_1) + \delta.$$

Take $\varepsilon \to 0$, by Lemma 94, we obtain

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) \leqslant \mathbf{mHK}^2(\mu_0, \mu_1) + \delta.$$

□

Proof of Proposition 92. From the previous proofs, we show that $\mathbf{mHK}^2 = \mathbf{mHK}^2_{\mathfrak{C}}$. Now, let us prove that the minimum in $\mathbf{mHK}^2$ is attained. Show in **Proof of $\leqslant$ in Proposition 92**, given an optimal solution $\hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$, there exist (μ, v, w) satisfying the continuity equation with reaction in (4.3) such that $v_t, w_t \in L^2(X; \mu_t)$ and

$$\mathbf{mHK}^2(\mu_0, \mu_1) \leqslant \int_0^1 \int_{\mathbb{R}^d} |(v_t, \frac{1}{2}w_t)|^2 \, d\mu_t \, dt \leqslant \int_{\mathfrak{C}^{\otimes 2}} \tilde{d}_{\mathfrak{C}}^2(\mathfrak{y}_0, \mathfrak{y}_1) \, d\hat{\alpha} = \mathbf{mHK}_{\mathfrak{C}}^2(\mu_0, \mu_1).$$

Since $\mathbf{mHK}^2 = \mathbf{mHK}_{\mathfrak{C}}^2$, then $\mathbf{mHK}^2(\mu_0, \mu_1)$ attains the minimum at (μ, v, w). □

5.7.1 Correspondence Between Absolutely Continuous Curves and Solutions to the Continuity Equation

Lemma 95 ([ABS21, p.209])**.** Let $\gamma : I \to X$ be a curve with values in a metric space (X, d), satisfying

$$d^2(\gamma(s), \gamma(t)) \leqslant (t - s) \int_s^t g(r)^2 \, dr \quad \forall s, t \in I \text{ with } s \leqslant t \tag{5.121}$$

for some $g \in L^2(0, 1)$. Then $\gamma \in AC^2(I; X)$ and $|\gamma'| \leqslant g$ holds $\mathcal{L}^1$-a.e. in $(0, 1)$.

Theorem 96. Let $(\mu_t)_{t \in I}$ be a curve in $C(I; \mathcal{M}(\mathbb{R}^d))$. If μ is a solution to the continuity equation with reaction in (4.3) induced by a vector field $(v_t)_{t \in I}$ and a scalar field $(w_t)_{t \in I}$, then $\mu \in AC^2(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}}))$ and

$$|\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) \leqslant \int_{\mathbb{R}^d} \left(|v_t|^2 + \frac{1}{4}|w_t|^2 \right) d\mu_t \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, 1). \tag{5.122}$$

Proof. By Proposition 92 and applying a rescaled version of the Benamou-Brenier formula under a constraint (Proposition 35), we obtain

$$\begin{aligned} \mathbf{mHK}_{\mathfrak{C}}^2(\mu_t, \mu_{t+h}) &= \mathbf{mHK}^2(\mu_t, \mu_{t+h}) \\ &\leqslant h \int_t^{t+h} \int \left(|v(s, x)|^2 + \frac{1}{4}|w(s, x)|^2 \right) \mu(s, dx) \, ds \quad \forall h > 0. \end{aligned}$$

Then by Lemma 95, we have the desired result. □

Theorem 97. Given $\mu \in AC^2(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}}))$, there exist a velocity field $(v_t)_{t \in I}$ and a scalar field $(w_t)_{t \in I}$ such that μ_t solves the associated continuity equation in (4.3) and

$$v_t, w_t \in L^2(\mu_t) \quad \text{and} \quad \int \left(|v_t|^2 + \frac{1}{4}|w_t|^2 \right) d\mu_t = |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) \quad \text{for } \mathcal{L}^1\text{-a.e. } t \in (0, 1). \tag{5.123}$$

Proof. By Theorem 88, if $\mu \in AC^2(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}}))$, then there exists a dynamic plan $\eta \in \mathcal{P}(AC^2(I; \mathfrak{C}))$ such that

1. $\mu_t = \mathfrak{h}_t^2 \eta$ in I.
2. $|\mu_t'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) = \int_{C(I;\mathfrak{C})} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma)$ for $\mathcal{L}^1$-a.e. $t \in (0,1)$.

We have that $\mu_t = (\mathsf{X}_t)_\#(R_t^2 \eta) \in \mathcal{M}(\mathbb{R}^d)$ since for any $\varphi \in \mathcal{B}_b(\mathbb{R}^d)$,

$$\int \varphi(x)\, d\mu_t = \int \varphi(x)\, d(\mathfrak{h}_t^2 \eta) = \int \varphi(\mathsf{x}(e_t(\gamma))) \cdot \mathsf{r}(e_t(\gamma))^2\, d\eta(\gamma) = \int \varphi(x)\, d[(\mathsf{X}_t)_\#(R_t^2 \eta)].$$

By Lemma 93, there exist v, w satisfying the continuity equation with reaction in (4.3) such that $v_t, w_t \in L^2(\mu_t)$ and

$$\int |(v_t, \frac{1}{2} w_t)|^2\, d\mu_t \leqslant \int |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma) = |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) \leqslant \int |(v_t, \frac{1}{2} w_t)|^2\, d\mu_t,$$

where the last inequality holds thanks to Theorem 96. □

Proposition 98. Given $\mu_0, \mu_1 \in \mathcal{M}_2(\mathbb{R}^d)$ with compact supports. Assume that $\mu_0 \ll \mathcal{L}^d$. If (μ, v, w) is the optimal solution of $\mathbf{mHK}^2(\mu_0, \mu_1)$, then there exists $\tilde{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ such that $\tilde{\alpha} = \mathrm{dil}_{\vartheta,2} \tilde{\alpha}$ and

$$\mu_t = \mathfrak{h}_t^2 \widetilde{\Sigma}_\# \tilde{\alpha}, \tag{5.124}$$

where $\tilde{\Sigma}$ is given in (5.21).

Proof. If (μ, v, w) is the optimal solution of $\mathbf{mHK}^2(\mu_0, \mu_1)$ solving the associated continuity equation, then by Theorem 96, $\mu \in AC^2(I; (\mathcal{M}(\mathbb{R}^d), \mathbf{mHK}_{\mathfrak{C}}))$ and

$$\int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt \leqslant \int_0^1 \int_{\mathbb{R}^d} \left(|v(t,x)|^2 + \frac{1}{4} |w(t,x)|^2 \right) \mu(t, dx)\, dt = \mathbf{mHK}^2(\mu_0, \mu_1).$$

Thanks to Proposition 92, $\mathbf{mHK}^2 = \mathbf{mHK}^2_{\mathfrak{C}}$, and due to the definition of the metric derivative of $\mathbf{mHK}_{\mathfrak{C}}$, we obtain

$$\int |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt \leqslant \mathbf{mHK}^2(\mu_0, \mu_1) = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) \leqslant \int |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt.$$

which implies that

$$\mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1) = \int |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt.$$

Moreover, by Theorem 88, there exists η such that $\mu_t = \mathfrak{h}_t^2 \eta$ and

$$|\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t) = \int_{C(I;\mathfrak{C})} |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, d\eta(\gamma).$$

Therefore,

$$\int_{C(I;\mathfrak{C})} \int_0^1 |\gamma'|^2_{\tilde{d}_{\mathfrak{C}}}(t)\, dt\, d\eta(\gamma) = \int_0^1 |\mu'|^2_{\mathbf{mHK}_{\mathfrak{C}}}(t)\, dt = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1).$$

so that η is an optimal dynamic plan. By Corollary 87, $\alpha := (e_0, e_1)_{\#}\eta \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$. Let $\tilde{\alpha} = dil_{r_1,2}\alpha$ be a scaled optimal transport plan. Due to Proposition 89, there exists a Radon measure $\tilde{\mu}_t = \mathfrak{h}_t^2 \tilde{\Sigma}_{\#}\tilde{\alpha}$ such that

$$\int_0^1 |\tilde{\mu}'|^2(t)\, dt = \mathbf{mHK}^2_{\mathfrak{C}}(\mu_0, \mu_1).$$

We can show that $\mu_t = \tilde{\mu}_t$: for any $\phi \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\int \phi(x)\, d\mu_t = \int \phi(x)\, d(\mathfrak{h}_t^2 \eta) &= \int \phi(\mathsf{x} \circ e_t(\gamma)) \cdot \left[\mathsf{r} \circ e_t(\gamma)\right]^2 d\eta(\gamma) \\
&= \int [\phi(\mathsf{x}) \cdot \mathsf{r}^2] \circ e_t(\tilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))\, d\alpha(\mathfrak{y}_0, \mathfrak{y}_1) \\
&= \int [\phi(\mathsf{x}) \cdot \mathsf{r}^2] \circ e_t(\tilde{\Sigma}(\mathfrak{y}_0, \mathfrak{y}_1))\, d\tilde{\alpha}(\mathfrak{y}_0, \mathfrak{y}_1) \\
&= \int \phi(x)\, d(\mathfrak{h}_t^2 \tilde{\Sigma}_{\#}\tilde{\alpha}) = \int \phi(x)\, d\tilde{\mu}_t.
\end{aligned}$$

the third equality holds due to the optimal dynamic plan η concentrated on a set defined in (5.110), by Corollary 87.

□

Corollary 99. Continuing with Proposition 98, if (μ, v, w) is the optimal solution of $\mathbf{mHK}^2(\mu_0, \mu_1)$, then μ is unique, i.e. given two optimal solutions $(\tilde{\mu}, \tilde{v}, \tilde{w})$ and $(\hat{\mu}, \hat{v}, \hat{w})$ of $\mathbf{mHK}^2(\mu_0, \mu_1)$, we have $\tilde{\mu} = \hat{\mu}$.

Proof. According to Proposition 98, there exist $\tilde{\alpha}, \hat{\alpha} \in Opt_{\mathbf{mHK}_{\mathfrak{C}}}(\mu_0, \mu_1)$ such that

1. $\tilde{\alpha} = \mathrm{dil}_{\vartheta,2}\tilde{\alpha}$ and $\tilde{\mu}_t = \mathfrak{h}_t^2 \tilde{\Sigma}_{\#}\tilde{\alpha}$.

2. $\hat{\alpha} = \mathrm{dil}_{\vartheta,2}\hat{\alpha}$ and $\hat{\mu}_t = \mathfrak{h}_t^2 \tilde{\Sigma}_\# \hat{\alpha}$.

Due to Proposition 85, $\tilde{\alpha} = \hat{\alpha}$ so that $\tilde{\mu}_t = \hat{\mu}_t$ for all $t \in I$ almost everywhere.

$\square$

6. Numerical Schemes

In this section, we expand upon concepts introduced in [Chi+18a, Section 5] and [PPO14]. We aim to leverage these generalized ideas to devise an effective numerical approximation for the $\mathbf{mHK}^2$ problem in (4.19).

6.1 Discretization

Given the previous works in [Chi+18a] and [CP11], we also implemented algorithms on the centered and staggered grids.

6.1.1 Centered and Staggered Grids

Given the time domain $[0,1]$ and the space domain $[0,L]$, we discretize the time domain into T pieces and the space domain into N pieces, then $\Delta x = \frac{L}{N}$ and $\Delta t = \frac{1}{T}$. The centered grid is obtained from discretizing the domain $[0,1] \times [0,L]$ equally which does not include boundaries, that is

$$\mathcal{G}_c = \left\{ \big(t_i = (i - \frac{1}{2})\Delta t, x_j = (j - \frac{1}{2})\Delta x\big) : 1 \leqslant i \leqslant T, 1 \leqslant j \leqslant N \right\}.$$

The staggered grids on the space and time domains are given by

$$\mathcal{G}_s^x = \left\{ \big(t_i = (i - \frac{1}{2})\Delta t, x_j = (j - 1)\Delta x\big) : 1 \leqslant i \leqslant T, 1 \leqslant j \leqslant N + 1 \right\},$$

$$\mathcal{G}_s^t = \left\{ \big(t_i = (i - 1)\Delta t, x_j = (j - \frac{1}{2})\Delta x\big) : 1 \leqslant i \leqslant T + 1, 1 \leqslant j \leqslant N \right\}.$$

Figure 6.1 is an example of the centered grid $\mathcal{G}_c$ (labeled by the star), the staggered grid on the space domain $\mathcal{G}_s^x$ (labeled by the circle), and time domain $\mathcal{G}_s^t$ (labeled by the rectangle) when $N = 4, T = 3$.

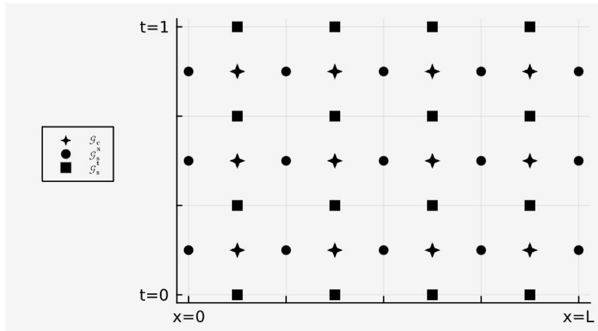


FIGURE 6.1: Centered and Staggered Grids

The variables discretized on the centered grid and staggered grid are

$$V = (\mu, \omega, \zeta) \in \mathcal{E}_c = (\mathbb{R}^{\mathcal{G}_c})^3 \quad \text{and} \quad U = (\bar{\mu}, \bar{\omega}, \bar{\zeta}) \in \mathcal{E}_s = \mathbb{R}^{\mathcal{G}_s^t} \times \mathbb{R}^{\mathcal{G}_s^x} \times \mathbb{R}^{\mathcal{G}_c}.$$

Since satisfying the continuity equation requires computing the partial derivative of $\bar{\mu}, \bar{\omega}$, then we act them on the staggered grid.

6.1.2 Discrete Continuity Constraint

The continuity equation with boundary conditions under the constraint in (4.20) is discretized on the staggered grid. Given $U = (\bar{\mu}, \bar{\omega}, \bar{\zeta}) \in \mathcal{E}_s$, the continuity constraint $\partial_t \bar{\mu} + \nabla \cdot \bar{\omega} = 2\bar{\zeta}$ is linear. Define a divergence operator $\mathbf{div}\colon \mathcal{E}_s \to \mathbb{R}^{\mathcal{G}_c}$ as

$$\mathbf{div}(U)_{i,j} := \frac{1}{\Delta t}(\bar{\mu}_{i+1,j} - \bar{\mu}_{i,j}) + \frac{1}{\Delta x}(\bar{\omega}_{i,j+1} - \bar{\omega}_{i,j}) \tag{6.1}$$

and $s_z(U) := 2\bar{\zeta}$, then the discrete scheme for the continuity constraint is

$$\mathbf{div}(U) - s_z(U) = 0. \tag{6.2}$$

Moreover, to discretize the boundary constraints, where $\bar{\mu}(0, \cdot)$ and $\bar{\mu}(1, \cdot)$ should be equivalent to initial and target distributions respectively, i.e.

$$\bar{\mu}(0, x) = \mu_0(x) \quad \text{and} \quad \bar{\mu}(1, x) = \mu_1(x) \quad \text{for all } x,$$

and $\bar{\omega}$ satisfies the Neumann boundary condition,

$$\bar{\omega}(t, 0) = 0 = \bar{\omega}(t, L) \quad \text{for all } t,$$

we define a function $s_b : \mathcal{E}_s \to \mathbb{R}^{2(N+T)}$ by $s_b(U) := b_0$ where b_0 is a vector of the boundary values, i.e.

$$b_0 := \begin{bmatrix}\mu_0(x_1) & \dots & \mu_0(x_N) & \mu_1(x_1) & \dots & \mu_1(x_N) & 0 & \dots & 0 & 0 & \dots & 0\end{bmatrix}^T$$

and the function s_b selects the values of U on the boundaries of the staggered grids (red points shown in Figure 6.2).

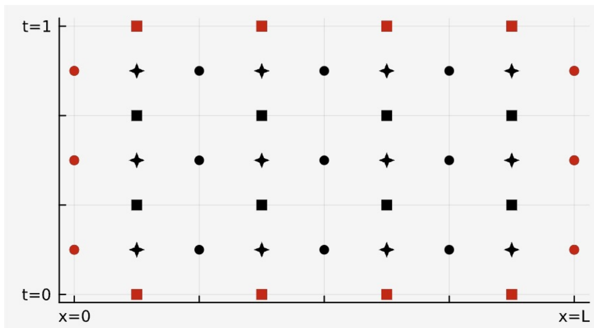


FIGURE 6.2: Boundaries on the Staggered Grids

Thus, the continuity equation with boundary conditions can be defined as a set

$$\mathcal{CE} := \left\{ U \in \mathcal{E}_s : A(U) = f_0,\ A = \begin{bmatrix}\mathbf{div} - s_z \\ s_b\end{bmatrix},\ f_0 = \begin{bmatrix}0 \\ b_0\end{bmatrix} \right\}. \tag{6.3}$$

6.1.3 Discrete Energy Functional and Interpolation Operator

The energy functional in (4.19) is discretized on the centered grid, i.e. $D(V) := \sum_{I \in \mathcal{G}_c} f(\mu_I, \omega_I, \zeta_I)$ where $f := f_1$ is given in (4.16). Since the energy functional and the continuity equation with boundary constraints are not defined on the same grid, we add a linear constraint $\mathcal{I}(U) = V$, where $\mathcal{I} : \mathcal{E}_s \to \mathcal{E}_c$ is a midpoint interpolation operator such that for any $U \in \mathcal{E}_s$,

$$\mathcal{I}(U)_{i,j} := \Big(\mathcal{I}_1(\bar{\mu})_{i,j}, \mathcal{I}_2(\bar{\omega})_{i,j}, \mathcal{I}_3(\bar{\zeta})_{i,j}\Big) := \left(\frac{\bar{\mu}_{i+1,j} + \bar{\mu}_{i,j}}{2}, \frac{\bar{\omega}_{i,j+1} + \bar{\omega}_{i,j}}{2}, \bar{\zeta}_{i,j}\right). \tag{6.4}$$

6.1.4 Discrete Optimization Problem

The discrete scheme of the convex minimization problem in (4.19) is

$$\text{minimize} \quad \Big\{ D(V) + \iota_{\mathcal{CE}}\big(U\big) + \iota_{\{V=\mathcal{I}(U)\}}(U,V) + \iota_{\mathcal{C}^+}(U,V) \Big\}, \tag{6.5}$$

where $\mathcal{C}^+ := \big\{(U,V) \in \mathcal{E}_s \times \mathcal{E}_c : \bar{\zeta}, \zeta \in \mathbb{R}_+^{\mathcal{G}_c}\big\}$ with $\mathbb{R}_+^{\mathcal{G}_c} := \big\{\zeta \in \mathbb{R}^{\mathcal{G}_c} : \zeta \geqslant 0\big\}$.

6.2 Minimization Algorithms

In this section, we will provide two methods to approximate the optimal solutions of the discrete minimization problem in (6.5). Before introducing the algorithms, let us provide some definitions. Denote by $\mathcal{H}$ be a Hilbert space.

Definition 6.6 ([CP11])**.** The projection $\Pi_C(x)$ of $x \in \mathbb{R}^d$ onto the nonempty closed convex set $C \subset \mathbb{R}^d$ is the solution to

$$\text{minimizie}_{y\in\mathbb{R}^d} \quad \iota_C(y) + \frac{1}{2}\|x-y\|^2. \tag{6.7}$$

Definition 6.8 ([CP11])**.** Let $f : \mathbb{R}^d \to [-\infty, \infty]$ be a lower semicontinuous convex function. For every $x \in \mathbb{R}^d$, the minimization problem

$$\text{minimizie}_{y\in\mathbb{R}^d} \quad f(y) + \frac{1}{2}\|x-y\|^2 \tag{6.9}$$

admits a unique solution, denoted by $\text{prox}_f(x)$. The operator $\text{prox}_f : \mathbb{R}^d \to \mathbb{R}^d$ is the proximal operator of f.

6.2.1 Douglas-Rachford Splitting Algorithm

Problem 6.10. *Define*

$$G_1(U,V) := D(V) + \iota_{\mathcal{CE}}(U), \tag{6.11}$$

$$G_2(U,V) := \iota_{\{V=\mathcal{I}(U)\}}\big(U,V\big) + \iota_{\mathcal{C}^+}\big(U,V\big). \tag{6.12}$$

Then the discrete minimization problem in (6.5) is equivalent to

$$\textit{minimize} \quad G_1(U,V) + G_2(U,V). \tag{6.13}$$

The minimal solutions of problem 6.10 can be approximated using the **Douglas-Rachford Splitting Algorithm**, which was originally proposed in [DR56] for solving linear problems and later generalized to nonlinear problems by [LM79]. According to [Chi+18a] and [CP11], the Douglas-Rachford algorithm facilitates the identification of minimal solutions for the sum of two convex, proper, and lower semicontinuous functions, G_1 and G_2, through iterative schemes: choose $(z^{(0)}, w^{(0)}) \in \mathcal{H}^2, \alpha \in (0,2)$, and $\gamma > 0$,

$$\begin{aligned} w^{(l+1)} &= w^{(l)} + \alpha\big(\text{prox}_{\gamma G_1}(2z^{(l)} - w^{(l)}) - z^{(l)}\big), \\ z^{(l+1)} &= \text{prox}_{\gamma G_2}\big(w^{(l+1)}\big), \end{aligned} \tag{6.14}$$

then $z^{(l)} \to z^*$. (see [[CP11]] for the details of convergence).

6.2.2 Alternating Direction Method of Multipliers

Problem 6.15. *Define*

$$G_1(U,V) := D(V) + \iota_{\mathcal{CE}}(U), \tag{6.16}$$

$$G_2(U,V) := \iota_{\{V=\mathcal{I}(U)\}}(U,V), \tag{6.17}$$

$$H(U,V) := \iota_{\mathcal{C}^+}((U,V)). \tag{6.18}$$

Then the discrete minimization problem in (6.5) is equivalent to

$$\begin{aligned} &\textit{minimize} \quad G_1(U_1,V_1) + G_2(U_2,V_2) + H(X,Y), \\ &\textit{subject to} \quad (U_i,V_i) = (X,Y), \quad i = 1,2. \end{aligned} \tag{6.19}$$

The minimal solutions of problem 6.15 can be approximated using the **Alternating Direction Method of Multiplier (ADMM) algorithm**. According to [Boy+11, Chapter 7], the problem 6.15 is a global variable consensus problem with regularization and it can be solved through iterative schemes: choose $(x_1^{(0)}, x_2^{(0)}, z^{(0)}, y_1^{(0)}, y_2^{(0)})$ and $\gamma > 0$,

$$\begin{aligned} x_i^{(\ell+1)} &= \text{prox}_{\gamma G_i}\big(z^{(\ell)} - \gamma y_i^{(\ell)}\big), \quad i = 1,2, \\ z^{(\ell+1)} &= \Pi_{\mathcal{C}^+}\left(\sum_{i=1}^{2} x_i^{(\ell+1)} + \gamma y_i^{(\ell)}\right), \\ y_i^{(\ell+1)} &= y_i^{(\ell)} + \frac{1}{\gamma}\big(x_i^{(\ell+1)} - z^{(\ell+1)}\big), \quad i = 1,2. \end{aligned}$$

6.2.3 Proximal Operators

Proposition 100. The proximal operator of γG_1 is equivalent to

$$\operatorname{prox}_{\gamma G_1}(U_0, V_0) = \big(\Pi_{\mathcal{CE}}(U_0),\, \operatorname{prox}_{\gamma D}(V_0)\big). \tag{6.20}$$

Proof of Proposition 100.

$$\begin{aligned}
&\operatorname{prox}_{\gamma G_1}(U_0, V_0)\\
&\quad= \arg\min_{U,V} \left\{\gamma\, D(V) + \gamma\, \iota_{\mathcal{CE}}(U) + \frac{1}{2}\|(U,V) - (U_0,V_0)\|^2\right\}\\
&\quad= \left(\arg\min_{U} \left\{\gamma\, \iota_{\mathcal{CE}}(U) + \frac{1}{2}\|U - U_0\|^2\right\},\, \arg\min_{V} \left\{\gamma\, D(V) + \frac{1}{2}\|V - V_0\|^2\right\}\right)\\
&\quad= \left(\Pi_{\mathcal{CE}}(U_0),\, \operatorname{prox}_{\gamma D}(V_0)\right).
\end{aligned}$$

□

Proposition 101. The projection operator of U onto $\mathcal{CE}$ is equivalent to

$$\begin{aligned}
\Pi_{\mathcal{CE}}(U) &= U + A^*(AA^*)^{-1}(f_0 - A(U))\\
&= \Pi_B(U) - (s_z^* - \mathbf{div}^*)S^{-1}(s_z - \mathbf{div})(\Pi_B(U)).
\end{aligned} \tag{6.21}$$

where $S = (\mathbf{div})(\mathbf{div}^*) + 4id$ is the Schur complement of the block $s_b s_b^*$ of AA^*, and $\Pi_B(U)$ is the protection onto the boundaries $B := \big\{U \in \mathcal{E}_c : s_b(U) = b_0\big\}$ such that

$$\begin{aligned}
\Pi_B(U) &= U + s_b^*(s_b s_b^*)^{-1}(b_0 - s_b(U))\\
&= U + s_b^*(b_0 - s_b(U)).
\end{aligned} \tag{6.22}$$

Moreover, given $p = (s_z - \mathbf{div})\big(\Pi_B(U)\big)$, we can approximate $u = S^{-1}p$ by solving $-\Delta u + 4u = p$ where u satisfies Neuwmann boundary conditions on the staggered grids. Then u can be numerically solved by taking the Fourier inverse of $\hat{u}$ where $\hat{u}$ solves

$$\begin{aligned}
\hat{u}(m,n) \cdot \Bigg(4 + \frac{1}{(\Delta x)^2}\left[2 - 2\cos\left(\frac{\pi}{N}(n + \frac{1}{2})\right)\right]\\
+ \frac{1}{(\Delta t)^2}\left[2 - 2\cos\left(\frac{\pi}{T}(m + \frac{1}{2})\right)\right]\Bigg) = \hat{p}(m,n).
\end{aligned} \tag{6.23}$$

Proof of Proposition 101. Compute

$$AA^* = \begin{bmatrix} \mathbf{div} - s_z \\ s_b \end{bmatrix} \begin{bmatrix} \mathbf{div}^* - s_z^* & s_b^* \end{bmatrix}$$

$$= \begin{bmatrix} (\mathbf{div} - s_z)(\mathbf{div}^* - s_z^*) & (\mathbf{div} - s_z)s_b^* \\ s_b(\mathbf{div}^* - s_z^*) & s_b s_b^* \end{bmatrix}$$

and $s_b s_b^* = id$, we obtain the Schur complement of the block $s_b s_b^*$ of AA^* as

$$\begin{aligned} S &= AA^*/s_b s_b^* \\ &= (\mathbf{div} - s_z)(\mathbf{div}^* - s_z^*) - (\mathbf{div} - s_z)s_b^* s_b(\mathbf{div}^* - s_z^*) \\ &= (\mathbf{div})(\mathbf{div}^*) + 4id \end{aligned}$$

and the last equality holds since

$$s_z\mathbf{div}^* = 0 = \mathbf{div}s_z^*, \quad s_b\mathbf{div}^* = 0 = \mathbf{div}s_b^*, \quad s_b s_z^* = 0 = s_z s_b^*, \quad s_z s_z^* = 4id.$$

Furthermore, we have

$$(AA^*)^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}(\mathbf{div} - s_z)s_b^* \\ -s_b(\mathbf{div}^* - s_z^*)S^{-1} & Id + s_b(\mathbf{div}^* - s_z^*)S^{-1}(\mathbf{div} - s_z)s_b^* \end{bmatrix}.$$

Substitute that into

$$\Pi_{\mathcal{CE}}(U) = U + A^*(AA^*)^{-1}(f_0 - A(U)),$$

we will have the desired result for $\Pi_{\mathcal{CE}}(U)$. Since $(\mathbf{div})(\mathbf{div}^*) \approx -\Delta$, then $Su = p$ is approximated by $-\Delta u + 4u = p$. To solve this equation, we can represent u by normalized discrete cosine transform, i.e.

$$u(i,j) = \sum_{m=1}^{T}\sum_{n=1}^{N} \hat{u}(m,n)\cos(\frac{\pi}{N}(n+\frac{1}{2})j)\cos(\frac{\pi}{T}(m+\frac{1}{2})i),$$

where $i = -\frac{1}{2}, \ldots, T + \frac{1}{2}, j = -\frac{1}{2}, \ldots, N + \frac{1}{2}$, and substitute that into $-\Delta u + 4u = p$. We obtain

$$\left(4 + \frac{1}{(\Delta x)^2}\big[2 - 2\cos(\frac{\pi}{N}(n+\frac{1}{2}))\big] + \frac{1}{(\Delta t)^2}\big[2 - 2\cos(\frac{\pi}{T}(m+\frac{1}{2}))\big]\right)\hat{u} = \hat{p}.$$

Taking the inverse Fourier transform, we can recover the solution of u. □

Proposition 102. The proximal operator of γD is equivalent to

$$\mathrm{prox}_{\gamma D}(V) = \big(\mathrm{prox}_{\gamma f}(V_I)\big)_{I\in\mathcal{G}_c}, \tag{6.24}$$

where for any $I \in \mathcal{G}_c$,

$$\mathrm{prox}_{\gamma f}(\mu_I, \omega_I, \zeta_I) = \begin{cases} (\tilde{\mu}_I, \frac{\tilde{\mu}_I\omega_I}{\tilde{\mu}_I+2\gamma}, \frac{\tilde{\mu}_I\zeta_I}{\tilde{\mu}_I+2\gamma}) & \text{if } \tilde{\mu}_I > 0,\\ (0,0,0) & \text{otherwise,}\end{cases} \tag{6.25}$$

and $\tilde{\mu}_I$ is the largest real root of $(X-\mu_I)(X+2\gamma)^2 - \gamma(|\omega_I|^2 + {\zeta_I}^2) = 0$.

Proof of Proposition 102. Since

$$\begin{aligned}\mathrm{Prox}_{\gamma D}(V) &= \arg\min_{\tilde{V}\in\mathcal{E}_c}\left\{\gamma D(\tilde{V}) + \frac{1}{2}\|\tilde{V}-V\|^2\right\}\\ &= \arg\min_{\tilde{V}\in\mathcal{E}_c}\left\{\sum_{I\in\mathcal{G}_c}\gamma f(V_I) + \frac{1}{2}\|\tilde{V}_I - V_I\|^2\right\}, \quad V_I = (\mu_I,\omega_I,\zeta_I)\\ &= \left(\arg\min_{\tilde{V}_I}\left\{\gamma f(V_I) + \frac{1}{2}\|\tilde{V}_I - V_I\|^2\right\}\right)_{I\in\mathcal{G}_c},\end{aligned}$$

then for any $I \in \mathcal{G}_c$, $\mathrm{prox}_{\gamma f}(V_I) = \tilde{V}_I = (\tilde{\mu}_I, \tilde{\omega}_I, \tilde{\zeta}_I)$ satisfies $\tilde{V}_I - V_I + \gamma\nabla f(\tilde{V}_I) = 0$. If $\tilde{\mu}_I > 0$, we have

$$\nabla f(\mu_I,\omega_I,\zeta_I) = (-\frac{|\omega_I|^2+\zeta_I^2}{\mu_I^2}, \frac{2\omega_I}{\mu_I}, \frac{2\zeta_I}{\mu_I}).$$

Simplify the linear systems,

$$\begin{cases}\tilde{\mu}_I - \mu_I - \gamma(|\tilde{\omega}_I|^2 + \tilde{\zeta}_I^2)/(\tilde{\mu}_I^2) & = 0\\ \tilde{\omega}_I - \omega_I + \gamma(2\tilde{\omega}_I)/(\tilde{\mu}_I) & = 0\\ \tilde{\zeta}_I - \zeta_I + \gamma(2\tilde{\zeta}_I)/(\tilde{\mu}_I) & = 0,\end{cases}$$

we obtain

$$\begin{cases}(\tilde{\mu}_I - \mu_I)(\tilde{\mu}_I + 2\gamma)^2 - \gamma({\omega_I}^2 + \zeta_I^2) = 0\\ \tilde{\omega}_I = \frac{\tilde{\mu}_I\omega_I}{2\gamma+\tilde{\mu}_I}\\ \tilde{\zeta}_I = \frac{\tilde{\mu}_I\zeta_I}{2\gamma+\tilde{\mu}_I}.\end{cases}$$

□

Proposition 103. The proximal operator of $\gamma\iota_{\{V=\mathcal{I}(U)\}}$ is equivalent to

$$\operatorname{prox}_{\gamma\iota_{\{V=\mathcal{I}(U)\}}}(U_0, V_0) = (\tilde{U}, \mathcal{I}(\tilde{U})), \tag{6.26}$$

where $\tilde{U} = (\tilde{\mu}, \tilde{\omega}, \tilde{\zeta})$ satisfies

$$\tilde{\mu} = Q_1^{-1}\big(\bar{\mu}^0 + \mathcal{I}_1^*(\mu^0)\big), \quad \tilde{\omega} = Q_2^{-1}\big(\bar{\omega}^0 + \mathcal{I}_2^*(\omega^0)\big), \quad \tilde{\zeta} = \frac{1}{2}(\bar{\zeta}^0 + \zeta^0), \tag{6.27}$$

with $Q_i = \mathrm{Id} + \mathcal{I}_i^*\mathcal{I}_i$ for $i = 1, 2$ and $\mathcal{I}_i^*$ is the adjoint operator of $\mathcal{I}_i$.

Proof of Proposition 103.

$$\begin{aligned}\operatorname{Prox}_{\gamma\iota_{\{V=\mathcal{I}(U)\}}}(U_0, V_0) &= \arg\min_{(U,V)} \left\{\gamma\iota_{\{V=\mathcal{I}(U)\}}\big((U,V)\big) + \frac{1}{2}\|(U_0, V_0) - (U, V)\|_2^2\right\} \\ &= \arg\min_{(U,\mathcal{I}(U))} \left\{\frac{1}{2}\|(U_0, V_0) - (U, \mathcal{I}(U))\|_2^2\right\}.\end{aligned}$$

Then $\tilde{U} = (\tilde{\mu}, \tilde{\omega}, \tilde{\zeta})$ s.t.

$$\begin{aligned}\tilde{\mu} &:= \arg\min_{\bar{\mu}} \left\{\frac{1}{2}\|(\bar{\mu}^0, \mu^0) - (\bar{\mu}, \mathcal{I}_1(\bar{\mu}))\|_2^2\right\}, \\ \tilde{\omega} &:= \arg\min_{\bar{\omega}} \left\{\frac{1}{2}\|(\bar{\omega}^0, \omega^0) - (\bar{\omega}, \mathcal{I}_2(\bar{\omega}))\|_2^2\right\}, \\ \tilde{\zeta} &:= \arg\min_{\bar{\zeta}} \left\{\frac{1}{2}\|(\bar{\zeta}^0, \zeta^0) - (\bar{\zeta}, \mathcal{I}_3(\bar{\zeta}))\|_2^2\right\}.\end{aligned}$$

W.L.O.G, take the derivative w.r.t $\bar{\mu}$, we have

$$2\tilde{\mu} - 2\bar{\mu}^0 + 2\mathcal{I}_1^*\mathcal{I}_1(\tilde{\mu}) - 2\mathcal{I}_1^*(\mu^0) = 0 \implies \tilde{\mu} = (\mathrm{Id} + \mathcal{I}_1^*\mathcal{I}_1)^{-1}(\bar{\mu}^0 + \mathcal{I}_1^*(\mu^0)).$$

□

Proposition 104. The proximal operator of $\gamma(\iota_{\{V=\mathcal{I}(U)\}} + \iota_{\mathcal{C}^+})$ is equivalent to

$$\operatorname{prox}_{\gamma(\iota_{\{V=\mathcal{I}(U)\}}+\iota_{\mathcal{C}^+})}(U_0, V_0) = (\tilde{U}, \mathcal{I}(\tilde{U})), \tag{6.28}$$

where $\tilde{U} = (\tilde{\mu}, \tilde{\omega}, \tilde{\zeta})$ such that $\tilde{\mu}$, $\tilde{\omega}$ are given in (6.27) and

$$\tilde{\zeta}_I = \max\left\{\left(\frac{\bar{\zeta}^0 + \zeta^0}{2}\right)_I, 0\right\} \quad \text{for } I \in \mathcal{G}_c.$$

Proof of Proposition 104.

$$\begin{aligned}
&\operatorname{Prox}_{\gamma(\iota_{\{V=\mathcal{I}(U)\}}+\iota_{\mathcal{C}^+})}(U_0,V_0)\\
&=\arg\min_{(U,V)}\left\{\gamma\,\iota_{\{V=\mathcal{I}(U)\}}(U,V)+\gamma\,\iota_{\mathcal{C}^+}(U,V)+\frac{1}{2}\|(U_0,V_0)-(U,V)\|_2^2\right\}\\
&=\arg\min_{(U,\mathcal{I}(U))}\left\{\gamma\,\iota_{\mathcal{C}^+}(U,\mathcal{I}(U))+\frac{1}{2}\|(U_0,V_0)-(U,\mathcal{I}(U))\|_2^2\right\}.
\end{aligned}$$

Then $\tilde{U}=(\tilde{\mu},\tilde{\omega},\tilde{\zeta})$ s.t.

$$\begin{aligned}
\tilde{\mu}&:=\arg\min_{\bar{\mu}}\left\{\frac{1}{2}\|(\bar{\mu}^0,\mu^0)-(\bar{\mu},\mathcal{I}_1(\bar{\mu}))\|_2^2\right\}\\
\tilde{\omega}&:=\arg\min_{\bar{\omega}}\left\{\frac{1}{2}\|(\bar{\omega}^0,\omega^0)-(\bar{\omega},\mathcal{I}_2(\bar{\omega}))\|_2^2\right\}\\
\tilde{\zeta}&:=\arg\min_{\bar{\zeta}}\left\{\gamma\iota_{\mathbb{R}_+^{\mathcal{G}_c}\times\mathbb{R}_+^{\mathcal{G}_c}}(\bar{\zeta},\mathcal{I}(\bar{\zeta}))+\frac{1}{2}\|(\bar{\zeta}^0,\zeta^0)-(\bar{\zeta},\mathcal{I}_3(\bar{\zeta}))\|_2^2\right\}.
\end{aligned}$$

□

Proposition 105. The projection of (U_0,V_0) onto $\mathcal{C}^+$ is $\Pi_{\mathcal{C}^+}(U_0,V_0)=(\tilde{U},\tilde{V})$ where for $I\in\mathcal{G}_c$,

$$\tilde{U}_I=\big(\bar{\mu}_I^0,\,\bar{\omega}_I^0,\,\max(\bar{\zeta}_I^0,0)\big)\qquad\text{and}\qquad\tilde{V}_I=\big(\mu_I^0,\,\omega_I^0,\,\max(\zeta_I^0,0)\big).$$

6.3 Numerical experiments

To reproduce the results, the source code in Julia and animations of the process are available in the GitHub link: `https://github.com/yd124/Unbalanced-OT`.

6.3.1 Examples: 1D Gaussian Bumps

Consider the initial density ρ_0 as a composite of two Gaussian densities with masses of 1 and 4, respectively, within the interval $\Omega=[0,3]$. These densities have peaks at 0.2 and 2.2. Conversely, the target density ρ_1 is formed from Gaussian densities of masses 4 and 1, supported on the same segment $\Omega=[0,3]$, but with modes at 0.4 and 2.0. The problem is discretized using $T=16$ in time steps and $N=256$ in spatial divisions.

FIGURE 6.3: Transport Between the Gassuain Bumps

Figure 6.4 illustrates the variance in transported densities between the unconstrained (HK) and constrained (mHK) scenarios. In the unconstrained case, the transported density $\rho_t^{\mathbf{HK}}$ exhibits local increases and decreases. Conversely, the constrained case, incorporating a non-negativity condition, prevents any reduction in the transported density $\rho_t^{\mathbf{mHK}}$. This results in a noticeable mass redistribution from the right peak to the left, demonstrating the significant impact of the non-negative constraint on the mass transport dynamics.

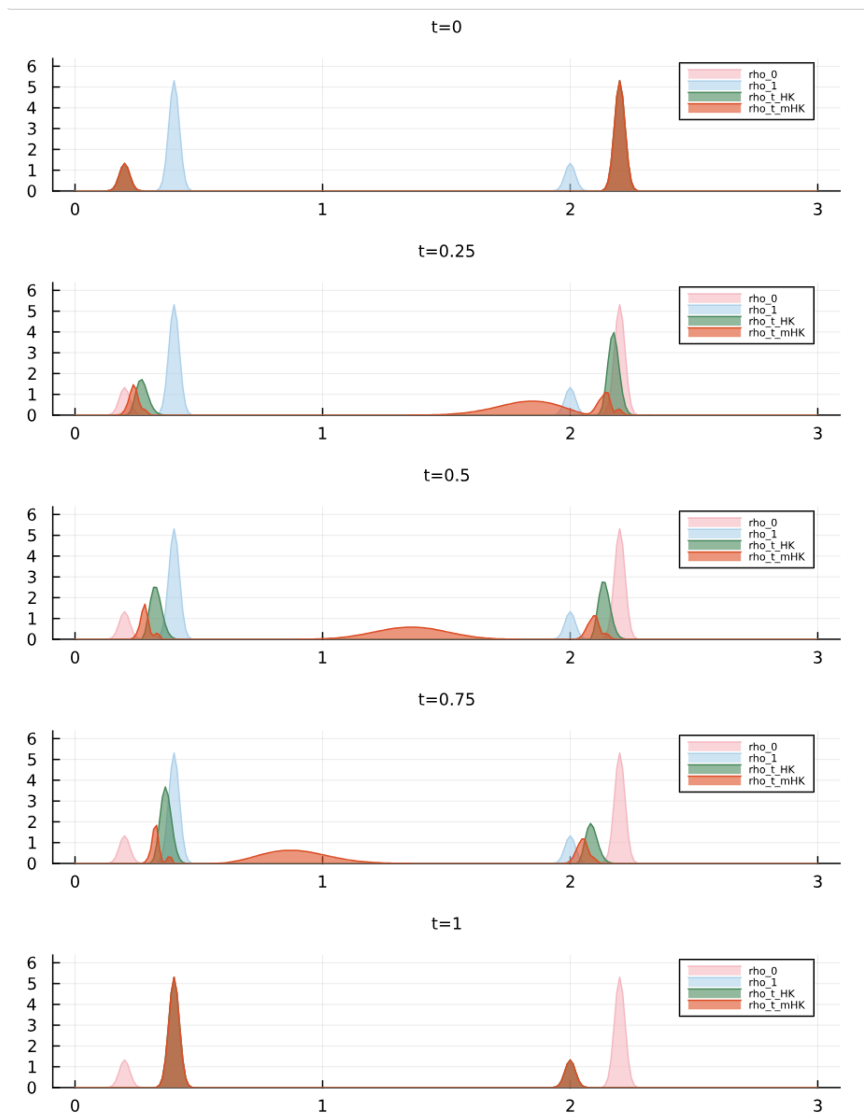


FIGURE 6.4: The Transport Between Gaussian Bumps

6.3.2 Examples: 2D Images Transport

Let the domain $\Omega = [-2, 2]$. Consider the initial density ρ_0 as composed of two distinct rings: the first ring is centered at $(-1, -1)$, featuring an inner diameter of 0.05 and an outer diameter of 0.15, while the second ring, centered at $(1, 1)$, has inner and outer radii of 0.25 and 0.45, respectively. On the other hand, the target density ρ_1 consists of two rings as well: the first, centered at $(-1, -1)$, has inner and outer radii of 0.25 and 0.45, mirroring the second ring of ρ_0; the second ring, centered at $(1, 1)$, is defined by an inner diameter of 0.05

and an outer diameter of 0.15, matching the first ring of ρ_0. This configuration inversely mirrors the ring arrangements between ρ_0 and ρ_1, establishing a symmetrical relationship in their compositions.

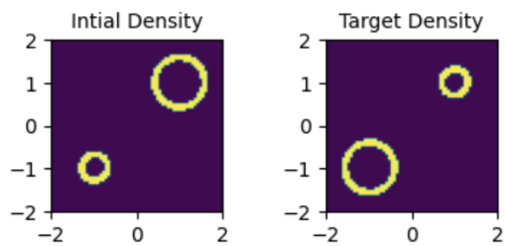


FIGURE 6.5: Intial and Target Densities

Figure 6.6 demonstrates the dynamic changes in density over a specified interval. The initial set of images (in the first row) vividly captures the evolution of the optimal density for the HK problem, highlighting the local contraction and expansion of two distinct rings. This visual progression succinctly illustrates the density adjustments necessary for optimizing the HK framework. In contrast, the subsequent series of images (in the second row) distinctly reveals the mass transfer process: a significant portion of the mass is relocated from the larger ring to the smaller one. Concurrently, there is evident growth of the smaller ring in its original location, indicating not just a transfer but also an accumulation process. This detailed portrayal underscores the complex interactions and adjustments within the system, providing a clear visual understanding of the density transformation phenomena.

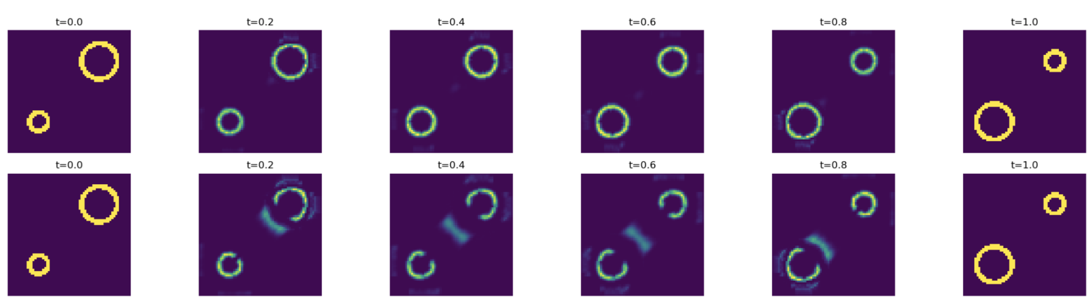


FIGURE 6.6: The Transport Between 2D Images

7. Conclusion

In this paper, we build several unbalanced optimal transport problems with growth constraints and connect their relationships. We succeed in proving the two major theorems: the existence and uniqueness of the optimal solutions for our problems. Moreover, we developed a numerical scheme based on previous work. In the future, we hope to apply our problem in a real scenario, for example, we want to uncover the change in the density of a tumor within a period, which might provide more treatment insights.

Appendix A. Proof of Results in Continuity Equations

Proof of Proposition 4. It suffices to check when $\varphi(t,x) = \eta(t)\phi(x)$ for any $\eta \in C_c^\infty((0,T))$ and $\phi \in C_c^\infty(\mathbb{R}^d)$, since D is dense in $C_c^\infty((0,T)\times\mathbb{R}^d)$ where

$$D := \left\{ \sum_{i=1}^{N} \eta_i(t)\phi_i(x) : \eta_i \in C_c^\infty((0,T)),\ \phi_i \in C_c^\infty(\mathbb{R}^d), i = 1,\dots,N \right\}.$$

Observe that

$$\begin{aligned} 0 &= \int_0^T \int_{\mathbb{R}^d} \Big(\partial_t \varphi + \nabla_x \varphi \cdot v_t + \varphi w_t \Big)\, d\mu_t\, dt \\ &= \int_0^T \int_{\mathbb{R}^d} \Big(\dot{\eta}(t)\phi(x) + \eta(t)\nabla_x \phi(x)\cdot v_t(x) + \eta(t)\phi(x) w_t(x) \Big)\, d\mu_t\, dt. \end{aligned}$$

Using the integration by parts, we can conclude that

$$\frac{d}{dt}\int_{\mathbb{R}^d} \phi\, d\mu_t = \int_{\mathbb{R}^d} \Big(\nabla_x \phi \cdot v_t + \phi w_t \Big)\, d\mu_t.$$

□

Proof of Proposition 5. Let us prove (i) at first. According to [SC16] and [Tes12], there exists a unique X_t and a unique λ_t of the following ODEs:

Cauchy problem	Leibniz formula
$\begin{cases} \frac{d}{dt}X_t = v_t(X_t) \\ X_0(x) = x \end{cases}$	$\begin{cases} \frac{d}{dt}\lambda_t = w_t(X_t)\lambda_t \\ \lambda_0(x) = 1 \end{cases}$

Then $\mu_t := (X_t)_\#(\lambda_t \mu_0)$ is unique. Since X_t and λ_t are continuous in t, then so is μ_t. Moreover, we want to show that (μ, v, w) solves (CE) in (1.14). Given any $\phi \in C_c^\infty(\mathbb{R}^d)$ and any $h > 0$, we obtain

$$\int_{\mathbb{R}^d} \phi(x)\, d\mu_{t+h} - \int_{\mathbb{R}^d} \phi(x)\, d\mu_t = \int_{\mathbb{R}^d} \Big(\phi(X_{t+h})\lambda_{t+h} - \phi(X_t)\lambda_t \Big) d\mu_0. \tag{A.1}$$

Since

$$\begin{aligned} \phi(X_{t+h})\lambda_{t+h} - \phi(X_t)\lambda_t &= \big[\phi(X_{t+h})\lambda_{t+h} - \phi(X_t)\lambda_{t+h}\big] + \big[\phi(X_t)\lambda_{t+h} - \phi(X_t)\lambda_t\big] \\ &= \int_t^{t+h} \Big(\lambda_{t+h}\frac{d}{ds}\phi(X_s) + \phi(X_t)\frac{d}{ds}\lambda_s \Big) ds. \end{aligned}$$

then (A.1) is equivalent to

$$\int_{\mathbb{R}^d} \phi(x)\, d\mu_{t+h} - \int_{\mathbb{R}^d} \phi(x)\, d\mu_t = \int_{\mathbb{R}^d} \int_t^{t+h} \Big(\lambda_{t+h} \nabla\phi(X_s) \cdot v_s + \phi(X_t)\lambda_s w_s(X_s)\Big)\, ds\, d\mu_0.$$

Take the derivative w.r.t time t from both sides, we can conclude that

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}^d} \phi\, d\mu_t &= \int_{\mathbb{R}^d} \lambda_t \Big(\nabla\phi(X_t) \cdot v_t + \phi(X_t) w_t(X_t)\Big) d\mu_0 \\ &= \int_{\mathbb{R}^d} \Big(\nabla\phi \cdot v_t + \phi w_t\Big) d\mu_t. \end{aligned}$$

This implies that (μ, v, w) solves (1.14) in the distributional sense.

Now, let us prove (ii). The triplet $(\mu^\varepsilon, v^\varepsilon, w^\varepsilon)$ solves (1.14) since

$$\partial_t \mu^\varepsilon + \nabla \cdot (v^\varepsilon \mu^\varepsilon) - w^\varepsilon \mu^\varepsilon = \Big(\partial_t \mu + \nabla \cdot (v\mu) + (w\mu)\Big) * \rho_\varepsilon = 0.$$

Let $\nu_t^x(dy) := \frac{\rho_\varepsilon(x-y)\, d\mu_t(y)}{\int \rho_\varepsilon(x-y)\, d\mu_t(y)} \in \mathcal{P}(\mathbb{R}^d)$, we can compute that

$$\begin{aligned} &\int_{\mathbb{R}^d} |v_t^\varepsilon(x)|^p + \frac{1}{4}|w_t^\varepsilon(x)|^p\, d\mu_t^\varepsilon(x) \\ &\qquad = \int_{\mathbb{R}^d} \left(\left| \int_{\mathbb{R}^d} v_t(y)\, d\nu_t^x(y) \right|^p + \frac{1}{4} \left| \int_{\mathbb{R}^d} w_t(y)\, d\nu_t^x(y) \right|^p \right) d\mu_t^\varepsilon(x) \\ &\qquad \leqslant \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(|v_t(y)|^p + \frac{1}{4}|w_t(y)|^p \right) d\nu_t^x(y)\, d\mu_t^\varepsilon(x) \\ &\qquad = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(|v_t(y)|^p + \frac{1}{4}|w_t(y)|^p \right) \rho_\varepsilon(x-y)\, d\mu_t(y)\, dx \\ &\qquad = \int_{\mathbb{R}^d} |v_t(y)|^p + \frac{1}{4}|w_t(y)|^p\, d\mu_t(y). \end{aligned}$$

Let $\omega = v\mu$ and $\zeta = w\mu$. Since $(\mu_t, \omega_t, \zeta_t) \mapsto \int_{\mathbb{R}^d} \left|\frac{\omega_t(dx)}{\mu_t(dx)}\right|^p + \left|\frac{\zeta_t(dx)}{\mu_t(dx)}\right|^p d\mu_t(x)$ is lower semi-continuous by Proposition 40, then the convergence in (1.22) holds.

□

Appendix B. Supplementary Calculations

Claim B.1. Let $f(\rho) := (\theta - \beta)\gamma - \rho$. When $\theta > \frac{\pi}{2}$, the equation $f(\rho) = 0$ has a unique solution ρ_θ in $(0, 1)$. When $\theta = \frac{\pi}{2}$, 0 is the only solution of $f(\rho) = 0$. In summary, $(\theta - \beta)\gamma = \rho$ has a unique solution in $[0, 1)$ when $\theta \geqslant \frac{\pi}{2}$.

Proof. Take the derivative of f, we obtain

$$f'(\rho) = -(\theta - \beta) \cdot \frac{\rho}{\gamma} < 0 \text{ for } \rho \in (0, 1).$$

Hence, f is a monotone decreasing function in $(0, 1)$.

1. If $\theta > \frac{\pi}{2}$, then $f(0) = (\theta - \frac{\pi}{2}) > 0$ and $f(1) = -1 < 0$. By the intermediate value theorem, there exists a unique $\rho_\theta \in (0, 1)$ s.t. $f(\rho_\theta) = 0$.
2. If $\theta = \frac{\pi}{2}$, then $f(0) = 0$. Thus, $\rho_\theta = 0$ is a unique solution of $(\theta - \beta)\gamma = \rho$.

□

Claim B.2. Fix $x, y \in \mathbb{R}^d$ and $\theta = d(x, y)$. Let

$$g_\theta(\rho) := (\rho(\theta - \beta) + \gamma) \cdot \gamma, \quad \forall \rho \in [\cos(\theta_{\frac{\pi}{2}}), 1].$$

Then

1. When $\theta \leqslant \frac{\pi}{2}$, $g_\theta(\rho) \leqslant 1$ for all $\rho \in [\cos(\theta), 1]$.
2. When $\theta > \frac{\pi}{2}$, then $g_\theta(\rho) > 1$ if $\rho \in (0, \rho_\theta)$ and $g_\theta(\rho) < 1$ if $\rho \in (\rho_\theta, 1)$ where ρ_θ is the unique solution of $(\theta - \beta)\gamma = \rho$.

Proof. Observe that

$$g_\theta(\rho) = (\rho(\theta - \beta) + \gamma) \cdot \gamma = \rho[(\theta - \beta)\gamma - \rho] + 1 := \rho f(\rho) + 1.$$

1. When $\theta \leqslant \frac{\pi}{2}$, since $f'(\rho) = -(\theta - \beta) \cdot \frac{\rho}{\gamma} < 0$, then $f(\rho) \leqslant f(\cos(\theta)) = \sin^2(\theta) \leqslant 1$.
2. When $\theta > \frac{\pi}{2}$, $g_\theta(0) = 1 = g_\theta(\rho_\theta)$. From the proof of Claim B.1, $f(\rho) > 0$ if $\rho \in (0, \rho_\theta)$ and $f(\rho) < 0$ if $\rho \in (\rho_\theta, 1)$. Therefore, we obtain $g_\theta(\rho) > 1$ if $\rho \in (0, \rho_\theta)$ and $g_\theta(\rho) < 1$ if $\rho \in (\rho_\theta, 1)$.

□

Claim B.3. Fix $x, y \in \mathbb{R}^d$ and $\theta = d(x, y)$. Let

$$f(u) := c(x, \rho, y, u) + c(x, 0, y, \sqrt{1-u^2}) - c(x, \rho, y, 1) \quad \forall u \in [\rho, 1], \tag{B.4}$$

where the cost is defined in 5.50. Then $f(u) < 0$ if and only if $\theta > \frac{\pi}{2}$ and $\rho\rho_\theta^{-1} < u$ where ρ_θ is the unique solution of $(\theta - \beta)\gamma = \rho$.

Proof. Take the derivative of f,

$$f'(u) = \left[g_\theta(\rho u^{-1}) - 1\right] \cdot 2u, \tag{B.5}$$

where g_θ is defined in Claim B.2 and by that claim, $f(u) < f(1) = 0$ if and only if $f'(u) > 0$ if and only if $g_\theta(\rho u^{-1}) > 1$ if and only if $\rho u^{-1} \in (0, \rho_\theta)$. □

Claim B.6. Fix $x, y \in \mathbb{R}^d$ with $\theta = d(x, y)$. Fix $\kappa \in (0, 1]$. Let $f(\rho) := c(x, \rho, y, 1) - F(\theta, \kappa) \cdot \rho^2$ for $\rho \in (0, 1]$ where the cost c and F are defined in (5.52) and (5.55). Then $f(\rho) \geqslant f(\kappa)$ in $(0, 1]$.

Proof. Take the partial derivative of f, we have

$$f'(\rho) = 2\rho \cdot \left[F(\theta, \rho) - F(\theta, \kappa)\right] \quad \forall \rho \in (0, 1].$$

Since $F(\theta, \cdot)$ is increasing and $f'(\kappa) = 0$, then $f(\rho) \geqslant f(\kappa)$ in $(0, 1]$. Moreover, $f(\kappa) = G(\theta, \kappa)$ where G is defined in (5.56), so $f(\kappa) \leqslant 1 = f(0)$. Thus, $f(\rho) \geqslant f(\kappa)$ in $(0, 1]$. □

Bibliography

[ABS21] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on optimal transport.* Springer, 2021.

[AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* 2nd. Birkhäuser Verlag, 2005.

[BB00] Jean-David Benamou and Yann Brenier. "A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem". *Numer. Math.* 84.375-393 (2000), pp. 2563–2609.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry.* American Mathematical Society, 2001.

[Boy+11] Stephen Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122.

[Bre91] Yann Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". *Communications on Pure and Applied Mathematics* XLIV (1991), pp. 375–417.

[Chi+18a] Lénaïc Chizat et al. "An interpolating distance between optimal transport and Fisher–Rao metrics". *Found Comput Math* 18 (2018), pp. 1–44.

[Chi+18b] Lénaïc Chizat et al. "Scaling algorithms for unbalanced optimal transport problems". *Mathematics of computation* 87.314 (2018), pp. 2563–2609.

[Com07] Patrick L. Combettes. "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery". *IEEE J. Selected Topics Signal Process* 1 (2007), pp. 564–574.

[CP11] Patrick L. Combettes and Jean-Christophe Pesquet. "Proximal Splitting Methods in Signal Processing". *Springer Optimization and Its Applications* 49 (2011).

[DNS09] Jean Dolbeault, Bruno Nazaret, and Giuseppe Savaré. "A new class of transport distances between measures". *Calc. Var.* 34 (2009), pp. 193–231.

[DR56] Jim Douglas and H. H. Rachford. "On the numerical solution of heat conduction problems in two or three space variables". *Trans. Amer. Math. Soc.* 82.2 (1956), pp. 421–439.

[FG10] Alessio Figalli and Nicola Gigli. "A new transportation distance between non-negative measures, with applications to gradients, flows with Dirichlet boundary conditions". *Journal de Mathématiques Pures et Appliquées* 94 (2 2010), pp. 107–130.

[Kan39] Lenoid V. Kantorovich. *Mathematical Methods of Organizing and Planning Production.* Vol. 28. 4. Management Science, 1939, pp. 366–422.

[KMV16] Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. "A new optimal transport distance on the space of finite radon measures" (2016).

[KS84] Martin Knott and Cyril S. Smith. "On the optimal mapping of distributions". *Journal of optimization theory and applications* 43.1 (1984).

[Lis06] Stefano Lisini. "Characterization of absolutely continuous curves in Wasserstein spaces". *Calc. Var.* 28 (2006), pp. 85–120.

[LM79] Pierre-Louis Lions and Brigitte Mercier. "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis* 16.6 (1979), pp. 964–979.

[LMS16] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves". *SIAM* 48.4 (2016), pp. 2869–2911.

[LMS18] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures". *Inventiones mathematics* 211 (2018), pp. 969–1117.

[Man17] Stefania Maniglia. "Probabilistic representation and uniqueness results for measure-valued solutions of transport equations". *J. Math Pures Appl.* 81 (2017), pp. 601–626.

[MC20] Simone D. Marino and Lénaïc Chizat. "A tumor growth model of hele-shaw type as a gradient flow". *ESAIM: COCV* 26.103 (2020 2020).

[Oxt70] John C. Oxtoby. "Homeomorphic measures in metric space". *American Mathematical Society* 24 (1970), pp. 419–423.

[PPO14] Nicolas Papadakis, Gabriel Peyré, and Edouard Oudet. "Optimal Transport with Proximal Splitting". *SIAM J. IMAGING SCIENCES* 7.1 (2014), pp. 212–238.

[PR12] Benedetto Piccoli and Francesco Rossi. "Generalized Wasserstein distance and its application to transport equations with source" (2012).

[PR13] Benedetto Piccoli and Francesco Rossi. "On properties of the Generalized Wasserstein distance" (2013).

[Pra07] Aldo Pratelli. "On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation". *Annales de l'Institut Henri Poincare (B) Probability and Statistics* 43 (1 2007), pp. 1–13.

[Rao45] Cramér Rao. "Information and accuracy attainable in the estimation of statistical parameters". *Bullentin of the Calcutta Mathematical Society* 37 (3 1945), pp. 81–91.

[San15] Filippo Santambrogio. *Optimal transport for applied mathematics*. Springer, 2015.

[SC16] David G. Schaeffer and John W. Cain. *Ordinary Differential Equations: Basics and Beyond*. 1st ed. Vol. 65. Springer New York, NY, 2016.

[Tes12] Gerald Teschl. *Ordinary differential equations and dynamical systems*. American Mathematical Society, 2012, pp. 969–1117.

[Vil03] Céderic Villani. *Topics in optimal transportation*. 2nd. American Mathematical Society, 2003.

[Vil09] Céderic Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009.

Biography

Yuqing Dai comes from Tai Zhou City in Zhe Jiang Province, China. In 2015, she attended the University of Wisconsin-Madison in the U.S. where she earned a Bachelor's Degree with honors in Mathematics. During her undergraduate, Yuqing engaged in volunteer work and explored a broad topic of mathematics courses. Building on her strong academic foundation, Yuqing was honored with the Prof. Linnaeus Wayland Dowling Award and the Mary Ellen Rudin Foundation Scholarship.

Motivated by her passion for mathematics, Yuqing advanced her studies at Duke University, enrolling in a Ph.D. program in Mathematics. Initially drawn to probability theory, she soon discovered a profound interest in optimal transport theory. Yuqing's long-standing fascination with solving the shortest path problem in everyday contexts naturally aligned with her research in optimal transport, making it an ideal focus for her doctoral studies.