VARIANTS OF THE OPTIMAL TRANSPORT PROBLEM AND THEIR DUALITY

By

Chamila Malagoda Gamage

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics—Doctor of Philosophy

2023

ABSTRACT

The classical Optimal Transport (OT) problem studies how to transport one distribution to another in the most efficient way. In the past few decades it has emerged as a very powerful tool in various fields, such as optimization theory, probability theory, partial differential equations, machine learning and data analysis. In this thesis, we will discuss some existing variants of the classical optimal transport problem, such as the capacity constrained OT problem, multi-marginal OT problem, entropy-regularized OT problem and barycenters, and we will introduce a couple of new variants by combining the existing versions. We will also discuss their duality results and some characterizations.

To my loving daughter Mishini Olivia...

ACKNOWLEDGEMENTS

The journey of my doctoral studies has been an arduous yet immensely rewarding experience. I have faced numerous challenges, encountered countless obstacles, and navigated uncharted territories. Today, I stand at the culmination of my efforts, filled with a deep sense of satisfaction and accomplishment. None of this would have been possible without the amazing people I have had around me.

First and foremost, I am extremely grateful to my advisor, Dr. Jun Kitagawa, for his unwavering support, guidance, and encouragement throughout this challenging process. His knowledge, insights, and belief in my abilities have been invaluable, and I feel incredibly fortunate to have had the opportunity to learn from him. I would also like to express the deepest of gratitude towards the members of my dissertation committee: Dr. Baisheng Yan, Dr. Zhengfang Zhou, and Dr. Russell Schwab, for their guidance through different stages of my Ph.D. career. I also take this opportunity to thank the National Science Foundation (NSF) as I was partially supported through my advisor's NSF grants DMS-1700094 and DMS-2000128 many times in the form of Research Assistantships, which were very helpful for me.

I am deeply indebted to Dr. Dejan Slepčev, at Carnegie Mellon University, whose guidance, expertise, and invaluable insights have shaped my research and made this thesis possible. Your time, patience, and fruitful discussions have not only expanded my knowledge but also helped me think deeper and approach problems from new perspectives. Your mentorship has been invaluable, and I am truly grateful for the profound impact you have had on my academic growth.

Furthermore, I would like to acknowledge the contributions of the professors who have taught and mentored me throughout my academic journey. Special thanks to Dr. Archil Gulisashvili and Dr. Sergiu Aizicovici, at Ohio University: Your dedication to teaching and commitment to excellence have provided me with a solid foundation in Analysis and PDEs. I am grateful for the knowledge and skills you imparted, which have been crucial to my research.

I would also like to extend a special thanks to my teaching mentors, Andrew Krause, and Tsveta Sendova. Your guidance and support in developing my teaching skills have been instrumental in my growth as an educator. The knowledge and experiences you shared have not only enriched my teaching abilities but have also enhanced my research journey.

To my dear friends, thank you for your unwavering support during the ups and downs of this Ph.D. journey. While I cannot mention everyone individually, I would like to express my special thanks to Dimitris Vardakis, Estefania Garcia, Mihalis Paparizos, Ana-Maria Raicu, Arman Travakoli, Joe Melby, Craig Gross, Keshav Sutrave, Rachel Domagalski, Franciska Domokos, Kai Huang, Rui Wang, Danika Van Niel, Samara Chamoun, Reshma Menon, Hitesh Gakhar, Leonardo Abbrescia and Kasun Fernando. Your friendship and support have been a constant source of motivation and encouragement. Your willingness to listen, provide feedback, and engage in intellectual discussions has enriched my research and personal growth. I am grateful for the countless hours spent together, both academically and socially, and for the memories we have created.

I could not have undertaken this journey without the immense support of my family, who have been my rock throughout this journey. First of all, I would like to thank my loving husband, Lubashan Pathirana, for his continuous support throughout this journey. His love, encouragement, and understanding have been my anchor during the most difficult moments. He has believed in me even when I doubted myself, and his constant presence has given me the strength to persevere. I would like to say a special thank you to my daughter, Mishini Olivia. She came into my life during this doctoral journey, and her presence has given me a reason to try harder, to push through the most difficult moments, and to keep going even when things seemed overwhelming. I would also like to thank my parents, my parents-in-law, and my sister, for their love, encouragement, and constant belief in my abilities. Your support has been the cornerstone of my achievements, and I am forever grateful for your sacrifices and understanding.

Finally, I would like to express my sincere appreciation to all the individuals who have contributed to this thesis, directly or indirectly. Your support, advice, and encouragement have played a significant role in shaping my research and shaping me as a researcher. To all those who have believed in me and supported me throughout this endeavor, I offer my deepest gratitude. Your contributions have been invaluable, and I am humbled by your presence in my life.

TABLE OF CONTENTS

INTRODUCTION

The origin of the Optimal Transport (OT) Problem goes back to the year of 1781, where the French Mathematician Gaspard Monge introduced a problem in his classic paper *Memoire sur la theorie des deblais et des remblais* [41], which is about finding the most efficient way of moving dirt from one place to another which was inspired due to military and economic purposes. However, this problem remained unsolved for almost two centuries, until the Russian Mathematician and economist Leonid Kantorovich's involvement in this problem ([30]) who made some progress with the invention of Linear Programming.

To get an insight of this problem, we will consider an example in the discrete setting. Suppose there is a large number of iron mines and the iron has to be transported to the refining factories. The problem is to find where each unit of iron should be transported so that the total transportation cost is minimized. Such an assignment from an initial position to its final position, is known as an "optimal transport plan".

Over the past few decades, the theory of OT has gained a lot of attention and it has been applied in various fields such as optimization theory, probability theory partial differential equations, machine learning, etc. In 1987, in [10] Yann Brenier showed that under certain conditions, there exists a unique transport plan that minimizes the cost associated to the Euclidian distance squared. In 1995, Wilfrid Gangbo and Robert McCann generalized this result for cost functions which are strictly convex or concave ([22, 23]). In [6], Benamou and Brenier presents a dynamical formulation of the OT problems which connects the OT theory to many other fields such as fluid mechanics ([11]), image processing ([42]), data analysis ([36]), etc. The field of Computational OT is another rapidly growing area as it serves as a powerful tool to compare probability distributions. Object recognition ([25]), label classification ([55]), and generative modelling ([50]) are few among many sub-fields in machine learning that widely apply OT tools.

Viewing the OT problem as a linear programming problem enables us to construct duality theory for the OT problem. It plays a significant role in understanding and solving the

OT problem. Furthermore, it helps to characterize optimal solutions which is often challenging to do without duality theory. For instance, in computation OT theory, the duality theory enables to use efficient computational algorithms, such as Sinkhorn algorithm ([45]) to approximate the solutions to the OT problem.

In this thesis, we will discuss a few variants of the classical OT problem, namely, the capacity constrained OT problem, multi-marginal OT problem, entropy-regularized OT problem and barycenters, and their duality theory. We will also discuss a couple of new variants by combining already existing versions, such as the capacity constrained multi-marginal OT (CCMMOT) problem and capacity constrained barycenters. By combining the two notions of the capacity constrained OT problem and the barycenter problem, we will introduce capacity constrained barycenters in Wasserstein space. Under certain assumptions, we will prove that the problem attains a minimizer and present some duality results. The notion of the CCMMOT problem already exists in the literature ([18]); however, a dual formulation for this problem does not exist. We will present a dual formulation for this problem and prove the strong duality result and the existence of dual maximizers. The entropy-regularized version of Wasserstein barycenters and their dual formulation also exist in the literature ([38]). The authors have proven that the strong duality holds and the existence of the primal problem via duality result. In this thesis, we will provide a direct proof for the existence of a minimizer for the primal problem and the existence of dual maximizers.

CHAPTER 1

PRELIMINARIES

Some standard symbols, definitions, and theorems used in this thesis are given below.

Notation

Let X be a Polish space (see Definition 1.0.1).

- $\mathcal{B}(X)$: The sigma algebra of Borel sets of X.
- $\mathcal{P}(X)$: The space of Borel probability measures on X.
- $\mathcal{P}_2(X)$: The space of Borel probability measures with finite second moment.
- $\mathcal{M}(X)$: The space of finite Borel measures.
- $\mathcal{M}_+(X)$: The space of positive, finite Borel measures.
- $C(X)$: Continuous functions on X.
- $C_b(X)$: Bounded, continuous functions on X.
- $L^1(\mathbb{R}^d)$: Functions integrable w.r.t. Lebesgue measure on $\mathbb{R}^d$.
- $L^1(X, d\mu)$: Functions integrable w.r.t. measure μ on X.
- $L^0(X, d\mu)$: Measurable functions on the space (X, μ).
- $[f]_+$: Positive part of the function f (see Definition 1.0.16).
- $[f]_-$: Negative part of the function f (see Definition 1.0.17).

Definitions

Definition 1.0.1. *(Polish Spaces) A Polish space is a separable completely metrizable topological space.*

Definition 1.0.2. *(Push forward of a measure) Given two Polish spaces X, Y, a Borel map $T : X \mapsto Y$, and a probability measure $\mu \in \mathcal{P}(X)$, we define the push forward of μ through T, denoted by $T_{\#}\mu \in \mathcal{P}(Y)$, as*

$$T_{\#}\mu(E) = \mu(T^{-1}(E)), \qquad \forall E \subset Y, \text{ Borel.}$$

Definition 1.0.3. *(Convex set) A subset C of a vector space V is convex if $(1-\lambda)x+\lambda y \in C$ whenever $x, y \in C$, and $0 \leq \lambda \leq 1$.*

Definition 1.0.4. *(Weak Convergence) A sequence $\{\mu_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}(X)$ converges weakly to $\mu \in \mathcal{P}(X)$, if for all $f \in C_b(X)$,*

$$\lim_{n\to\infty} \int_X f \mathrm{d}\mu_n = \int_X f \mathrm{d}\mu.$$

Definition 1.0.5. *(Tightness) A set $\mathcal{A} \subseteq \mathcal{P}(X)$ is tight, if $\forall \varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that*

$$\mu(X \setminus K_\varepsilon) < \varepsilon, \qquad \forall \mu \in \mathcal{A}.$$

Definition 1.0.6. *(Lower semi-continuity) Let (X, d) be a metric space. A function $f : X \mapsto \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous, if for every sequence x_n such that $x_n \to x$, we have*

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

Definition 1.0.7. *(Support of a measure) Let X be a separable metric space. We define the support of a measure γ, denoted by $\mathrm{spt}(\gamma)$, as the smallest closed set on which γ is concentrated.*

$$\mathrm{spt}(\gamma) := \bigcap_{\{E : E \text{ is closed and } \gamma(X\setminus E)=0\}} E.$$

Definition 1.0.8. *(Finite p^{th} moment) A measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ has finite p^{th} moment, if*

$$\int_{\mathbb{R}^d} |x|^p \; d\mu(x) < +\infty.$$

Definition 1.0.9. *(Vanishing measures on small sets) A probability measure* $\mu \in \mathcal{P}(\mathbb{R}^d)$ *is said to vanish on small sets if and only if* $\mu(E) = 0$, $\forall E \subset \mathcal{B}(\mathbb{R}^d)$ *having Hausdorff dimension* $d-1$ *or less.*

Definition 1.0.10. *(ω-continuity) A function* $f : X \mapsto \mathbb{R}$ *is said to be ω-continuous, if there exists a function* $\omega : [0,\infty] \mapsto [0,\infty]$ *such that* $\lim_{t\to 0} \omega(t) = \omega(0) = 0$ *and*

$$|f(x) - f(y)| \leq \omega(|x-y|), \qquad \forall x, y \in X.$$

Definition 1.0.11. *(K-Convexity along curves) Given a metric space* (X, d)*, a functional* $\phi : X \mapsto (-\infty, \infty]$ *is called* K*-convex on a curve* $\gamma : t \in [0,1] \mapsto \gamma_t \in X$*, for some* $K \in \mathbb{R}$*, if*

$$\phi(\gamma_t) \leq (1-t)\phi(\gamma_0) + t\phi(\gamma_1) - \frac{1}{2}Kt(1-t)d^2(\gamma_0, \gamma_1), \qquad \forall t \in [0,1].$$

Definition 1.0.12. *(Proper Convex function) A convex function* $f : X \mapsto [-\infty, \infty]$ *is called proper, if* $f(x) < \infty$ *for at least one* $x \in X$ *and* $f(x) > -\infty$ *for all* $x \in X$*.*

Definition 1.0.13. *(Infimal Convolution) Given two proper convex functions* f, g *on* $\mathbb{R}^d$*, we define their infimal convolution, denoted by* $f \Box g$*, as*

$$(f \Box g)(x) = \inf_y \{f(x-y) + g(y)\}, \qquad \forall x \in \mathbb{R}^d.$$

Definition 1.0.14. *(L-Lipschitzness) Given two metric spaces* (X, d_X) *and* (Y, d_Y)*, a function* $f : X \mapsto Y$ *is said to be L-Lipschitz, if there is a real constant* $L \geq 0$ *such that, for all* $x_1, x_2 \in X$*,*

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

Definition 1.0.15. *(Legendre-Fenchel Transform) Let* E *be a normed vector space, and* φ *a convex function on* E *with values in* $\mathbb{R} \cup \{\infty\}$*. Then the Legendre-Fenchel transform of* φ *is the function* φ^**, defined on the dual space* E^* *by the formula*

$$\varphi^*(z^*) = \sup_{z \in E}\{z^* \cdot z - \varphi(z)\}.$$

Definition 1.0.16. *(Positive part) Given a function $f : \mathbb{R} \mapsto \mathbb{R}$, we define its positive part, denoted by $[f]_+$, as*

$$[f(x)]_+ = \max\{f(x), 0\}.$$

Definition 1.0.17. *(Negative part) Given a function $f : \mathbb{R} \mapsto \mathbb{R}$, we define its negative part, denoted by $[f]_-$, as*

$$[f(x)]_- = -\min\{f(x), 0\}.$$

Theorems

Theorem 1.0.18. *(Prokhorov) Let (X, d) be a Polish space. Then a family $\mathcal{A} \subset \mathcal{P}(X)$ is relatively compact w.r.t. the weak topology if and only if it is tight.*

Theorem 1.0.19. *(Fatou's Lemma) Let $f_n : X \mapsto [0, \infty]$ be measurable, for each $n \in \mathbb{N}$. Then*

$$\int_X \liminf_{n\to\infty} f_n \; d\mu \leq \liminf_{n\to\infty} \int_X f_n \; d\mu.$$

Theorem 1.0.20. *(Monotone Convergence theorem) Let $\{f_n\}$ be a sequence of measurable functions on X such that*

(i) $0 \leq f_k(x) \leq f_{k+1}(x) \leq \infty$, *for all $k \in \mathbb{N}$ and all $x \in X$,*

(ii) $\lim_{n\to\infty} f_n(x) = f(x)$, *for all $x \in X$.*

Then,

$$\lim_{n\to\infty} \int_X f_n \; d\mu = \int_X f \; d\mu.$$

CHAPTER 2

THE CLASSICAL OPTIMAL TRANSPORT PROBLEM

2.1 The Primal Problem

Let X and Y be two Polish Spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be two Borel probability measures, and $c : X \times Y \mapsto \mathbb{R} \cup \{\infty\}$ be a Borel measurable cost function. The Monge Problem is the following:

Problem 2.1.1. *Find a Borel map $T : X \mapsto Y$, that minimizes the cost*

$$M(S) := \int_X c(x, S(x))\ d\mu(x) \tag{2.1.1}$$

among all Borel maps $S : X \mapsto Y$ such that $S_{\#}\mu = \nu$.

Such maps are called *transport maps* from μ to ν. Maps that minimize the cost $M(S)$ are called *optimal transport maps.*

The push forward condition $S_{\#}\mu = \nu$ can be characterized by

$$\int_Y f(y)\mathrm{d}\nu(y) = \int_Y f(y)\ dS_{\#}\mu(y) = \int_X f \circ S(x)\ d\mu(x), \qquad \forall f \in L^1(Y, d\nu). \tag{2.1.2}$$

There are few major drawbacks in the Monge formulation. For example:

- The constraint set could be empty.

  Eg: For $\mu = \delta_0$ and $\nu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, the condition (2.1.2) cannot hold for any $S : X \mapsto Y$ that is μ-a.e. single-valued.

- The cost $M(S)$ could be non-linear in S (depending on c), hence could be difficult to solve.

- The constraint $S_{\#}\mu = \nu$ may not be closed under weak convergence in general.

  Eg: ([2], Chapter 1) Let $\mu = \mathcal{L}|_{[0,1]}$ and $\nu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$. Consider the sequence of functions given by $S_n(x) := S(nx)$ where $S : \mathbb{R} \mapsto \mathbb{R}$ is a 1-periodic function defined by

$$S(x) = \begin{cases} 1 & \text{on } [0,1/2) \\ -1 & \text{on } [1/2,1) \end{cases} \tag{2.1.3}$$

Then, $(S_n)_{\#}\mu = \nu, \forall n \in \mathbb{N}$, but (S_n) weakly converges to $S = 0$ in $L^p, \forall 1 \leq p < \infty$, so that $S_{\#}\mu = \delta_0 \neq \nu$.

Due to these issues, we consider a relaxation of the Monge Problem, which is known as the Kantorovich Problem.

Definition 2.1.2. *For given two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, we define the set of all transport plans from μ to ν by*

$$\Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(X \times Y) : \text{Proj}_x(x,y)_{\#}\gamma = \mu, \text{Proj}_y(x,y)_{\#}\gamma = \nu\}. \tag{2.1.4}$$

The conditions on γ above, are known as the marginal conditions and they can also be defined as

$$\gamma(A \times Y) = \mu(A), \quad \forall A \in \mathcal{B}(X), \qquad \textit{and} \qquad \gamma(X \times B) = \nu(B), \quad \forall B \in \mathcal{B}(Y).$$

Then, the Kantorovich Problem is defined as below:

Problem 2.1.3. *Find a $\gamma_0 \in \Pi(\mu,\nu)$ that minimizes the cost*

$$K(\gamma) = \int_{X \times Y} c(x,y)\ d\gamma(x,y) \tag{2.1.5}$$

among all transport plans $\gamma \in \Pi(\mu,\nu)$.

Transport plans that minimize the cost $K(\gamma)$ are called *optimal transport plans.*

When compared to the Monge formulation, there are many advantages in the Kantorovich formulation, such as:

- The set $\Pi(\mu,\nu)$ is always non-empty as it contains $\mu \otimes \nu$.
- The cost $K(\gamma)$ is linear in γ (regardless of c), hence much easier to solve.

- The set $\Pi(\mu, \nu)$ is a convex set.

  We can easily see that for any $\gamma_1, \gamma_2 \in \Pi(\mu, \nu)$ and $0 \leq \lambda \leq 1$, $\lambda\gamma_1 + (1-\lambda)\gamma_2 \in \Pi(\mu, \nu)$.

- If $T_{\#}\mu = \nu$, then $\gamma := (\mathrm{Id} \times T)_{\#}\mu \in \Pi(\mu, \nu)$, hence the set of transport plans contains all transport maps.

Now, we will discuss the existence of a minimizer for the Kantorovich problem.

Theorem 2.1.4. *([51], Theorem 1.7) Let X, Y be two Polish spaces and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If $c : X \times Y \mapsto [0, \infty]$ is lower semi-continuous, then the Kantorovich problem* (2.1.3) *has a minimizer.*

The proof is based on the tightness of the set $\Pi(\mu, \nu)$ and the Prokhorov theorem.

From here onwards, we will call the Kantorovich problem, the classical Optimal Transport (OT) problem and we will denote it by

$$\mathrm{OT}_c := \inf_{\gamma \in \Pi(\mu,\nu)} \left\{ \int_{X \times Y} c(x, y)\ d\gamma(x, y) \right\}. \tag{2.1.6}$$

2.2 Duality

Let X, Y be two compact Polish spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \mapsto [0, \infty)$ be continuous. We will define the dual formulation of the OT problem as the following maximization problem:

$$\hat{\mathrm{OT}}_c^* := \sup_{(\phi,\psi) \in \Phi_c} \left\{ \int_X \phi(x)\ d\mu(x) + \int_Y \psi(y)\ d\nu(y) \right\} \tag{2.2.1}$$

where

$$\Phi_c := \{(\phi, \psi) \in C_b(X) \times C_b(Y) : \phi(x) + \psi(y) \leq c(x, y)\}. \tag{2.2.2}$$

Due to the lack of compactness of the above class of admissible functions, we will consider an alternative dual formulation.

Definition 2.2.1. *Given a function* $f : X \mapsto \mathbb{R} \cup \{\pm\infty\}$*, we define its c-transform,* $f^c : Y \mapsto \mathbb{R} \cup \{\pm\infty\}$ *by*

$$f^c(y) := \inf_{x \in X} \{c(x,y) - f(x)\}. \tag{2.2.3}$$

Similarly, given a function $g : Y \mapsto \mathbb{R} \cup \{\pm\infty\}$*, we define its* c^**-transform,* $g^{c^*} : X \mapsto \mathbb{R} \cup \{\pm\infty\}$ *by*

$$g^{c^*}(x) := \inf_{y \in Y} \{c(x,y) - g(y)\}. \tag{2.2.4}$$

Definition 2.2.2. *A function* $f : X \mapsto \mathbb{R} \cup \{\pm\infty\}$ *is called c-concave, if there exists a function* $g : Y \mapsto \mathbb{R} \cup \{\pm\infty\}$ *such that* $f = g^{c^*}$*. A function* $g : Y \mapsto \mathbb{R} \cup \{\pm\infty\}$ *is called* c^**-concave, if there exists a function* $f : X \mapsto \mathbb{R} \cup \{\pm\infty\}$ *such that* $g = f^c$*.*

We will denote the set of c-concave functions on X *by c-conc*(X) *and the set of* c^**-concave functions on* Y *by* c^**-conc*(Y)*.*

Observe that, given an admissible pair (ϕ, ψ) in $\hat{\mathrm{OT}}^*$, if we replace (ϕ, ψ) by (ϕ, ϕ^c) and then again by (ϕ^{cc^*}, ϕ^c), the value will be increased while satisfying the constraints ([51], Definition 1.10). Hence, we consider the following dual formulation.

$$\mathrm{OT}_c^* := \sup_{\phi \in c\text{-}conc(X)} \left\{ \int_X \phi(x)\ d\mu(x) + \int_Y \phi^c(y)\ d\nu(y) \right\}. \tag{2.2.5}$$

Functions that maximize OT_c^* are called *Kantorovich potentials.*

Now, we will present the existence of dual maximizers and strong duality results.

Theorem 2.2.3. *([51], Proposition 1.11) Let* X, Y *be compact subsets of* $\mathbb{R}^d$ *and c be a continuous function. Then* OT_c^* *has a solution* (ϕ, ψ) *such that* $\phi \in$ *c-conc*$(X), \psi \in c^*$*-conc*(Y) *and* $\psi = \phi^c$*.*

In the proof, one starts with a maximizing sequence (ϕ_n, ψ_n) and take the c-transforms so that it improves the dual formulation. This transformation will make (ϕ_n, ψ_n) equi-continuous and equi-bounded, so that one can apply the Arzelà-Ascoli theorem to get the existence result.

Theorem 2.2.4. *([51], Theorem 1.39) Let X, Y be two Polish spaces and suppose that $c : X \times Y \mapsto \mathbb{R}$ is uniformly continuous and bounded. Then,*

$$\mathrm{OT}_c = \mathrm{OT}_c^*.$$

The proof uses the concept of c-cyclical monotonicity, which we will discuss next.

2.3 Properties of the optimizers

Definition 2.3.1. *Given a function $c : X \times Y \mapsto \mathbb{R} \cup \{\infty\}$, we say a set $\Gamma \subset X \times Y$ is c-cyclically monotone, if for any positive integer p, any permutation σ of $\{1, \dots, p\}$, and any finite family of points $(x_1, y_1), \dots, (x_p, y_p) \in \Gamma$, we have*

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

Theorem 2.3.2. *([51], Theorem 1.38) Let γ be an optimal transport plan for OT_c and let c be a continuous function. Then, $\mathrm{spt}(\gamma)$ is a c-cyclically monotone set.*

Now, we will present a more general theorem.

Theorem 2.3.3. *([53], Theorem 5.9) Let X, Y be two Polish spaces and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Suppose that $c : X \times Y \mapsto \mathbb{R} \cup \{\infty\}$ is a lower semi-continuous function, such that there exist some real-valued, upper semi-continuous functions $u \in L^1(X, d\mu)$, $v \in L^1(Y, d\nu)$ satisfying*

$$c(x, y) \ge u(x) + v(y), \qquad \forall (x, y) \in X \times Y.$$

Then,

1. *Duality holds:*

$$\begin{aligned}
\mathrm{OT}_c &= \sup_{\substack{(\phi,\psi)\in C_b(X)\times C_b(Y)\\ \phi+\psi\le c}} \left\{ \int_X \phi(x)\ d\mu(x) + \int_Y \psi(y)\ d\nu(y) \right\} \\
&= \sup_{\substack{(\phi,\psi)\in L^1(X,d\mu)\times L^1(Y,d\nu)\\ \phi+\psi\le c}} \left\{ \int_X \phi(x)\ d\mu(x) + \int_Y \psi(y)\ d\nu(y) \right\} \\
&= \sup_{\phi\in L^1(X,d\mu)} \left\{ \int_X \phi(x)\ d\mu(x) + \int_Y \phi^c(y)\ d\nu(y) \right\}
\end{aligned}$$

$$= \sup_{\psi \in L^1(Y,d\nu)} \left\{ \int_X \psi^{c^*}(x)\ d\mu(x) + \int_Y \psi(y)\ d\nu(y) \right\}.$$

Note that, in the above suprema, we may take $\phi \in c$-conc(X) and $\psi \in c^*$-conc(Y).

2. *Suppose c is real-valued and the cost OT_c is finite. Then there is a measurable c-cyclically monotone set $\Gamma \subset X \times Y$, such that for any $\gamma \in \Pi(\mu, \nu)$, the following statements are equivalent.*

   a) *γ is optimal;*

   b) *γ is c-cyclically monotone;*

   c) *There exists a c-concave function $\phi : X \mapsto \mathbb{R} \cup \{-\infty\}$ such that $\phi(x) + \phi^c(y) = c(x, y)$, γ-a.e.;*

   d) *There exist functions $\phi : X \mapsto \mathbb{R} \cup \{\infty\}$ and $\psi : Y \mapsto \mathbb{R} \cup \{\infty\}$, such that $\phi(x) + \phi^c(y) \le c(x, y)$, $\forall (x, y)$, with equality γ-a.e.;*

   e) *γ is concentrated on Γ.*

Now, we will present a uniqueness result for the optimal transport plans. This is known as the *Brenier-McCann's Theorem.*

Theorem 2.3.4. *([54], Theorem 2.12) Let $c(x, y) = |x - y|^2$ and $\mu \in \mathcal{P}_2(X), \nu \in \mathcal{P}_2(Y)$. Suppose that μ vanishes on small sets. Then,*

1. *There exists a unique optimal transport plan, given by*

$$\gamma = (\mathrm{Id} \times \nabla u)_{\#}\mu,$$

   where ∇u is uniquely determined μ-a.e. such that u is convex and $\nabla u_{\#}\mu = \nu$.

   Furthermore,

$$\mathrm{spt}(\nu) = \overline{\nabla u(\mathrm{spt}(\mu))}.$$

2. *∇u is the unique solution to the Monge problem given by* (2.1.1).

3. *If ν also vanishes on smalls sets, then ∇u^* is a ν-a.e. unique solution to the Monge problem from ν to μ, such that u^* is convex and*

$$\nabla u(\nabla u^*(y)) = y \qquad \nu\text{-a.e. } y \qquad \text{and}$$
$$\nabla u^*(\nabla u(x)) = x \qquad \mu\text{-a.e. } x.$$

Finally, we will briefly discuss the notion of the Wasserstein distance.

Definition 2.3.5. *Let (X, d) be a Polish metric space and $\mu, \nu \in \mathcal{P}(X)$. For a given $p \in [1, \infty)$, we define the Wasserstein distance of order p between μ and ν by*

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)^p \; d\gamma(x, y) \right)^{1/p}.$$

Proposition 2.3.6. *([3], Chapter 7.1) W_p defines a distance on $\mathcal{P}_p(X)$.*

Proposition 2.3.7. *([53], Corollary 6.9) W_p is lower semi-continuous w.r.t. weak convergence of measures.*

CHAPTER 3

CAPACITY CONSTRAINED OT PROBLEM

3.1 Introduction

In the capacity constrained OT problem, we impose capacity constraints which limit the amount transported between any given source and corresponding target. As an example in the discrete case, we can consider a large number of coal mines from which coal has to be transported to refining factories and the problem is to find where each unit of coal should go with a minimum transportation cost. In the unconstrained OT problem, we assume that any amount of coal can be transported, whereas in the capacity constrained case, there is a limit to the amount of coal that can be transported from one mine to the corresponding factory.

Formally, given two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, that represent the distributions in source and target, respectively, and a finite Borel measure $\tilde{\gamma}$ on $\mathbb{R}^d \times \mathbb{R}^d$, that represents the capacity constraint for the transport plans, we minimize the cost:

$$\inf_{\gamma \in \Pi^{\tilde{\gamma}}(\mu,\nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\ d\gamma(x,y) \right\} \tag{3.1.1}$$

Here, the set $\Pi^{\tilde{\gamma}}(\mu,\nu)$ represents the set of transport plans from μ to ν bounded by $\tilde{\gamma}$.

In [46], Rachev and Rüschendorf introduced this problem on compact spaces where they study bounded below, Borel measurable and lower semicontinuous cost functions and obtained a dual formulation of the minimization problem. Recently, in a series of papers by Korman, McCann and Seis, [33, 35, 34], the authors have considered this problem for continuous, bounded cost functions on $\mathbb{R}^d \times \mathbb{R}^d$ and for finite, bounded capacity constraints. There, they have obtained the equivalence between the primal problem and a dual problem with the existence of minimizers of the capacity constrained OT problem and existence of dual maximizers. However, unlike in the classical case, any further information about dual maximizers such as regularity or inheriting properties from the cost function is still unknown.

In this chapter, we will present the existing results regarding this Capacity Constrained

OT problem and provide some characterization of the optimizers of the primal and dual problems.

3.2 The Primal Problem

Similar to the work in [33], we will use functions to represent mass densities.

Let f and g be two non-negative, compactly supported density functions in $L^1(\mathbb{R}^d)$ with equal total masses, i.e. $\int_{\mathbb{R}^d} f(x)\ dx = \int_{\mathbb{R}^d} g(y)\ dy$, representing the source and the target densities. Let c be a Borel measurable function on $\mathbb{R}^d \times \mathbb{R}^d$ representing the cost function. Let $\tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ be a compactly supported function representing the capacity constraint. Denote by $\Pi^{\tilde{h}}(f,g)$, the set of all joint densities $h \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals f and g, and bounded by $\tilde{h}$, i.e.

$$f(x) = \int_{\mathbb{R}^d} h(x,y)\ dy,\ \ g(y) = \int_{\mathbb{R}^d} h(x,y)\ dx, \quad \text{and} \quad 0 \leq h \leq \tilde{h}.$$

We define the two-marginal Capacity Constrained Optimal Transport (CCOT) problem between f and g under the capacity $\tilde{h}$ as the following minimization problem:

$$\mathrm{OT}_{\mathrm{CC}} := \inf_{h \in \Pi^{\tilde{h}}(f,g)} I_c(h) \tag{3.2.1}$$

where

$$I_c(h) := \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) h(x,y)\ dxdy.$$

Unlike in the unconstrained problem, it is not always guaranteed that the set $\Pi^{\tilde{h}}(f,g)$ is non-empty. The necessary and sufficient conditions for $\Pi^{\tilde{h}}(f,g)$ to be non-empty are as follows:

Proposition 3.2.1. *([46], Corollary 4.6.15) The set* $\Pi^{\tilde{h}}(f,g) \neq \emptyset$ *if and only if*

$$\int_A f(x)\ dx + \int_B g(y)\ dy \leq 1 + \int_{A \times B} \tilde{h}(x,y)\ dxdy,$$

for any Borel measurable sets $A, B \subset \mathbb{R}^d$.

The following theorem states that $\mathrm{OT}_{\mathrm{CC}}$ has a minimizer.

Theorem 3.2.2. *([33], Theorem 3.1) Let c be a bounded, continuous function on $\mathbb{R}^d \times \mathbb{R}^d$ and $0 \le \tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ be compactly supported. Take $0 \le f, g \in L^1(\mathbb{R}^d)$ compactly supported functions such that the set $\Pi^{\tilde{h}}(f,g)$ is non-empty. Then, OT_{CC} has a minimizer in $\Pi^{\tilde{h}}(f,g)$.*

In order to get uniqueness of the minimizers, we require the cost c to satisfy three conditions.

(C1) $c(x,y)$ is bounded,

(C2) there is a Lebesgue negligible closed set $Z \subset \mathbb{R}^d \times \mathbb{R}^d$ such that $c(x,y) \in C^2(\mathbb{R}^d \times \mathbb{R}^d \backslash Z)$ and,

(C3) $c(x,y)$ is non-degenerate: i.e. $\det \nabla^2_{xy} c(x,y) \neq 0$ for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \setminus Z$.

Theorem 3.2.3. *([33], Theorem 8.1) Suppose that the cost $c(x,y)$ satisfies the conditions (C1), (C2), and (C3). Let $0 \le \tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ be compactly supported. Take $0 \le f, g \in L^1(\mathbb{R}^d)$ compactly supported functions such that the set $\Pi^{\tilde{h}}(f,g)$ is non-empty. Then, OT_{CC} has a unique minimizer.*

Now, we will give a characterization of the minimizers of $\mathrm{OT}_{\mathrm{CC}}$.

Definition 3.2.4. *Let $\tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$. A density $h \in \Pi^{\tilde{h}}(f,g)$ is called geometrically extreme, if $h(x,y) = \mathbb{1}_W(x,y)\tilde{h}(x,y)$ for almost all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, for some Lebesgue measurable set $W \subset \mathbb{R}^d \times \mathbb{R}^d$.*

Theorem 3.2.5. *([33], Theorem 7.2) Suppose that the cost $c(x,y)$ satisfies the conditions (C1), (C2), and (C3). Let $0 \le \tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ be compactly supported. Take $0 \le f, g \in L^1(\mathbb{R}^d)$ compactly supported functions such that the set $\Pi^{\tilde{h}}(f,g)$ is non-empty. If $h \in \Pi^{\tilde{h}}(f,g)$ minimizes OT_{CC}, then h is geometrically extreme.*

3.3 Duality - Version I

For given $f, g \in L^1(\mathbb{R}^d)$ and $0 \leq \tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$, we consider the following maximization problem:

$$\mathrm{OT}^*_{\mathrm{CC}} := \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{h}}} J(u,v,w) \tag{3.3.1}$$

where

$$J(u,v,w) := \int_{\mathbb{R}^d} u(x)f(x)\ dx + \int_{\mathbb{R}^d} v(y)g(y)\ dy + \int_{\mathbb{R}^d\times\mathbb{R}^d} w(x,y)\tilde{h}(x,y)\ dxdy, \tag{3.3.2}$$

and

$$\begin{aligned}\mathrm{Lip}_{c,\tilde{h}} := \Big\{&(u,v,w) : u \in L^1(\mathbb{R}^d, fdx), v \in L^1(\mathbb{R}^d, gdy), w \in L^1(\mathbb{R}^d \times \mathbb{R}^d, \tilde{h}dxdy), \\ &u(x) + v(y) + w(x,y) \leq c(x,y), \text{and } w(x,y) \leq 0\Big\}.\end{aligned} \tag{3.3.3}$$

Now, we will present the strong duality result for the CCOT problem.

Theorem 3.3.1. *([35], Theorem 1) Let $f, g \in L^1(\mathbb{R}^d)$ be two probability densities such that $\Pi^{\tilde{h}}(f,g)$ is non-empty and $0 \leq \tilde{h} \in L^\infty(\mathbb{R}^d \times \mathbb{R}^d)$ be compactly supported. Let $c \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$. Then,*

$$\mathrm{OT}_{CC} = \mathrm{OT}^*_{CC}.$$

Remark 3.3.2. *In [35], the authors use an infinite dimensional linear programming duality with a quadratic penalization to get this result.*

In [34], the same authors prove the existence of dual maximizers for $\mathrm{OT}^*_{\mathrm{CC}}$.

Let X and Y be two compact subsets of $\mathbb{R}^d$ with unit volumes, f and g be probability densities on X and Y, respectively, and $0 \leq \tilde{h} \in L^\infty(X \times Y)$.

Instead of the dual functional (3.3.2), we consider the functional

$$J'(u,v) := \int_X uf\ dx + \int_Y vg\ dy - \int_{X\times Y} [-c+u+v]_+\, \tilde{h}\ dxdy. \tag{3.3.4}$$

and define

$$\mathrm{OT}^{*'}_{\mathrm{CC}} := \sup_{u\in L^1(X,fdx), v\in L^1(Y,gdy)} J'(u,v). \tag{3.3.5}$$

Note that $\mathrm{OT}^*_{\mathrm{CC}} = \mathrm{OT}^{*'}_{\mathrm{CC}}$ (see Appendix 1).

The existence result of dual maximizers is given below:

Theorem 3.3.3. *([34], Theorem 4.2) Let f, g and $\tilde{h}$ be continuous and strictly positive on their compact supports X, Y, and $X \times Y$, respectively. Let $c \in L^1(X \times Y)$. Fix an $\eta > 1$ and assume that $\Pi^{\tilde{h}/\eta}(f, g)$ is non-empty. Then, there exist functions $(u, v) \in L^1(X, f dx) \times L^1(Y, g dy)$, such that*

$$\mathrm{OT}^{*'}_{CC} = J'(u, v).$$

The authors also provide a characterization of the optimizers of the primal and the dual problems as follows:

Corollary 3.3.4. *([34], Corollary 1.1) Under the assumptions of Theorem 3.3.3, any $h \in \Pi^{\tilde{h}}(f, g)$ is optimal if and only if there exist functions $(u, v) \in L^1(X, f dx) \times L^1(Y, g dy)$, such that*

$$c - u - v \begin{cases} \geq 0 & \text{where } h = 0, \\ = 0 & \text{where } 0 < h < \tilde{h}, \\ \leq 0 & \text{where } h = \tilde{h}. \end{cases} \tag{3.3.6}$$

3.4 Duality - Version II

In [46], the authors consider the CCOT problem in a different setting.

Let X, Y be two compact subsets of $\mathbb{R}^d$ and let $c : X \times Y \mapsto \mathbb{R} \cup \{\infty\}$ be Borel measurable and bounded below. Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $\tilde{\gamma}$ be a finite Borel measure on $X \times Y$.

Then, the CCOT is defined as the following minimization problem:

$$\overline{\mathrm{OT}_{\mathrm{CC}}} := \inf_{\gamma \in \Pi^{\tilde{\gamma}}(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) \, d\gamma(x, y) \right\}, \tag{3.4.1}$$

where

$$\Pi^{\tilde{\gamma}}(\mu, \nu) := \{\gamma \in \Pi(\mu, \nu) : \gamma(A \times B) \leq \tilde{\gamma}(A \times B), \forall (A, B) \in \mathcal{B}(X) \times \mathcal{B}(Y)\}. \tag{3.4.2}$$

Its dual formulation is given by the following maximization problem:

$$\overline{\mathrm{OT}^*_{\mathrm{CC}}} := \sup_{\mathcal{B}} \left\{ \int_X u(x)\ d\mu(x) + \int_Y v(y)\ d\nu(y) + \int_{X\times Y} w(x,y)\ d\tilde{\gamma}(x,y) \right\}, \tag{3.4.3}$$

where the supremum is taken over the set $\mathcal{B}$ of real-valued functions u, v, w satisfying $u \in C_b(X), v \in C_b(Y)$, $w \in C_b(X\times Y)$ and $w \leq 0$ with $u(x)+v(y)+w(x,y) \leq c(x,y)$ everywhere.

Now, we will present the duality theorem for this version.

Theorem 3.4.1. *([46], Theorem 4.6.14) Let $c : X \times Y \mapsto \mathbb{R}\cup\{\infty\}$ be Borel measurable and bounded below. Then, the following statements are equivalent:*

(a) c is lower semi-continuous on $X \times Y$.

(b) The duality holds for all $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and $\tilde{\gamma} \in \mathcal{M}_+(X \times Y)$.

$$i.e. \qquad \overline{\mathrm{OT}_{CC}} = \overline{\mathrm{OT}^*_{CC}}.$$

The proof is based on the *abstract duality theorem* (see [46], Theorem 4.6.1).

3.5 A further characterization on the optimizers

Even though the idea of the CCOT problem is quite as natural as the classical OT problem, only a little is known about the optimizers when compared to other variants of OT problem. In this section, we will present some characterization on the optimizers of the primal and the dual problems for CCOT problem.

Let X, Y be compact subsets of $\mathbb{R}^d$ and c be a non-negative, bounded, continuous function on $X\times Y$. Let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ be probability measures which are absolutely continuous w.r.t. Lebesgue measure with densities $f \in L^1(X)$ and $g \in L^1(Y)$ and $\tilde{\gamma}$ be a compactly supported finite measure on $X \times Y$ that is absolutely continuous w.r.t. Lebesgue measure with a bounded density $\tilde{h} \in L^\infty(X \times Y)$.

We will redefine the primal and the dual problems as follows:

$$\mathrm{I}_{cap} := \inf_{\gamma\in\Pi^{\tilde{\gamma}}(\mu,\nu)} \left\{ \int_{X\times Y} c(x,y)\ d\gamma(x,y) \right\}. \tag{3.5.1}$$

$$\mathrm{I}^*_{cap} := \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{\gamma}}} J(u,v,w), \tag{3.5.2}$$

where

$$J(u,v,w) = \int_X u(x)\ d\mu(x) + \int_Y v(y)\ d\nu(y) + \int_{X\times Y} w(x,y)\ d\tilde{\gamma}(x,y)$$

and

$$\begin{aligned} \mathrm{Lip}_{c,\tilde{\gamma}} := \Big\{ &(u,v,w) : u \in L^1(X,d\mu), v\in L^1(Y,d\nu), w\in L^1(X\times Y, d\tilde{\gamma}), \\ &u(x)+v(y)+w(x,y) \le c(x,y), \text{ and } w(x,y)\le 0 \Big\}. \end{aligned} \tag{3.5.3}$$

Let $\gamma \in \Pi^{\tilde{\gamma}}(\mu,\nu)$ be a minimizer for I_{cap} and $(u,v,w)\in \mathrm{Lip}_{c,\tilde{\gamma}}$ be a maximizer for I^*_{cap}. Let $E \subseteq \mathrm{spt}(\tilde{\gamma})$ be the compact support of γ.

Then, by Theorem 3.2.5, we have that

$$\gamma = \begin{cases} \tilde{\gamma} & \text{on } E \subset X\times Y, \\ 0 & \text{elsewhere.} \end{cases} \tag{3.5.4}$$

By duality (Theorem 3.3.1), we have that

$$\begin{aligned} \int_{X\times Y} c\ d\gamma &= \int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\tilde{\gamma} \\ &= \int_X u\ d\mu + \int_Y v\ d\nu + \int_{(X\times Y)\cap E} w\ d\tilde{\gamma} + \int_{(X\times Y)\cap E^c} w\ d\tilde{\gamma}. \end{aligned}$$

By (3.5.4) and $w \le 0$ on $(X\times Y)\cap E^c$, we have that

$$\int_{X\times Y} c\ d\gamma \le \int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\gamma. \tag{3.5.5}$$

On the other hand, since $(u,v,w)\in \mathrm{Lip}_{c,\tilde{\gamma}}$, we have that

$$u(x)+v(y)+w(x,y) \le c(x,y), \quad \forall (x,y)\in X\times Y.$$

By integrating both sides with respect to γ, we get

$$\int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\gamma \le \int_{X\times Y} c\ d\gamma. \tag{3.5.6}$$

By combining (3.5.5) and (3.5.6), we get

$$\int_{X\times Y} c\ d\gamma = \int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\gamma. \tag{3.5.7}$$

Thus, by using the marginal conditions, we can write

$$\int_{X\times Y} (c-u-v-w)\ d\gamma = 0. \tag{3.5.8}$$

Since we have the inequality $c-u-v-w \geq 0$, we can conclude that

$$c(x,y) - u(x) - v(y) - w(x,y) = 0 \qquad \gamma\text{-a.e.} \tag{3.5.9}$$

Now, let $n \in \mathbb{N}$, σ be any permutation of $\{1,\dots,n\}$ and pick an arbitrary collection of points $(x_1,y_1),\dots,(x_n,y_n) \in \operatorname{spt}(\gamma)$. Then,

$$\begin{aligned}\sum_{i=1}^n c(x_i,y_i) - \sum_{i=1}^n w(x_i,y_i) &= \sum_{i=1}^n u(x_i) + v(y_i)\\ &= \sum_{i=1}^n u(x_i) + v(y_{\sigma(i)})\\ &\leq \sum_{i=1}^n c(x_i,y_{\sigma(i)}) - w(x_i,y_{\sigma(i)}).\end{aligned}$$

This shows that $\operatorname{spt}(\gamma)$ is $(c-w)$-cyclical monotone (see Definition 2.3.1).

From here onwards, we will assume that the capacity $\tilde{\gamma} = \kappa\mu\otimes\nu$ for some $\kappa \geq 1$.

Let $\gamma^* \in \Pi^{\tilde{\gamma}}(\mu,\nu)$ be a minimizer for I_{cap} and $(u,v,w) \in \mathrm{Lip}_{c,\tilde{\gamma}}$ be a maximizer for I^*_{cap}. Thus, by (3.5.7), we have

$$\int_{X\times Y} c\ d\gamma^* = \int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\gamma^*. \tag{3.5.10}$$

Again, consider the inequality

$$u(x) + v(y) + w(x,y) \leq c(x,y).$$

By integrating both sides with respect to an arbitrary $\gamma \in \Pi(\mu,\nu)$, we get

$$\int_X u\ d\mu + \int_Y v\ d\nu + \int_{X\times Y} w\ d\gamma \leq \int_{X\times Y} c\ d\gamma.$$

i.e.

$$\int_X u\ d\mu + \int_Y v\ d\nu \le \int_{X\times Y} (c-w)\ d\gamma.$$

Now, taking the infimum over $\gamma \in \Pi(\mu,\nu)$ from both sides, we get

$$\int_X u\ d\mu + \int_Y v\ d\nu \le \inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} (c-w)\ d\gamma.$$

Observe that, since

$$\Pi^{\tilde{\gamma}}(\mu,\nu) \subseteq \Pi(\mu,\nu),$$

we have

$$\begin{aligned}\inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} (c-w)\ d\gamma &\le \inf_{\gamma\in\Pi^{\tilde{\gamma}}(\mu,\nu)} \int_{X\times Y} (c-w)\ d\gamma \\ &\le \int_{X\times Y} (c-w)\ d\gamma^* \qquad (3.5.11) \\ &= \int_X u\ d\mu + \int_Y v\ d\nu, \qquad \text{(by (3.5.10))}.\end{aligned}$$

Recall that the strong duality result holds for the classical OT problem with Borel measurable and $\mu\otimes\nu$-a.e. finite costs (see [5], Theorem 2). Since $c-w\ge 0$ is Borel measurable and $\mu\otimes\nu$-a.e. finite, we have that

$$\inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} (c-w)\ d\gamma = \sup_{\substack{u\in L^1(X), v\in L^1(Y)\\ u+v\le c}} \int_X u\ d\mu + \int_Y v\ d\nu. \qquad (3.5.12)$$

Thus, by (3.5.11) and (3.5.12), we can conclude that (u,v) maximizes the dual of the classical OT problem with cost $(c-w)$.

Finally, we will list the above characterization in the following proposition.

Proposition 3.5.1. *Let $\gamma \in \Pi^{\tilde{\gamma}}(\mu,\nu)$ be a minimizer for I_{cap} and $(u,v,w)\in Lip_{c,\tilde{\gamma}}$ be a maximizer for I^*_{cap}. Then,*

1. $c(x,y)-u(x)-v(y)-w(x,y)=0 \qquad \gamma$*-a.e..*

2. spt(γ) *is $(c-w)$-cyclical monotone.*

3. *If $\tilde{\gamma}=\kappa\mu\otimes\nu$ for some $\kappa\ge 1$, then (u,v) maximizes the dual of the classical OT problem with cost $(c-w)$.*

CHAPTER 4

MULTI-MARGINAL OT PROBLEM

4.1 The Multi-Marginal OT (MMOT) Problem

4.1.1 Introduction

As opposed to the classical OTP, which is a two-marginal problem, multi-marginal OTP (also known as the multi-dimensional Monge-Kantorovich Problem) studies transporting mass from a single source to multiple targets. More generally, given a p-tuple of probability measures $(\nu_1, \ldots, \nu_p)$ in $\mathbb{R}^d$, the problem consists of finding an optimal way of successively rearranging ν_1 onto ν_i against a certain cost on $\mathbb{R}^{pd}$ where ν_1 represents the mass distribution at the source and $(\nu_j)_{j\neq 1}$ represent the mass distributions at the targets.

MMOT is a versatile framework that can be applied in various fields, including image processing, computer vision, economics, machine learning, natural language processing, and medical imaging, among others. For example, MMOT is used in economics to model and solve problems involving the distribution of resources, such as the allocation of goods among multiple buyers and sellers [13]. Also, MMOT is used in machine learning tasks such as clustering, where it is used to group similar data points together based on their feature similarities.

This was first discussed by Gangbo and Święch in [24], for a specific cost function. More characterization of the optimal solutions and some applications have been discussed in [43, 44] and a few variants such as multi-marginal partial OTP [31] have been introduced later on.

4.1.2 The Primal Problem (MMOT Problem)

Let p be a positive integer. For a given p-tuple of probability measures $(\nu_1, \ldots, \nu_p)$ in $\mathcal{P}(\mathbb{R}^d)$, and a lower semi-continuous cost function $c(x_1, \ldots, x_p) : \mathbb{R}^{pd} \mapsto \mathbb{R}$, we consider the minimization problem:

$$\mathrm{OT}_{\mathrm{MM}}(\nu_1, \ldots, \nu_p) := \inf_{\gamma \in \Pi(\nu_1, \ldots, \nu_p)} \int_{\mathbb{R}^{pd}} c(x_1, \ldots, x_p) d\gamma(x_1, \ldots, x_p), \tag{4.1.1}$$

where

$$\Pi(\nu_1, \ldots, \nu_p) := \{\gamma \in \mathcal{P}(\mathbb{R}^{pd}) : \mathrm{Proj}_{x_i}(x_1, \ldots, x_p)_{\#}\gamma = \nu_i, \forall 1 \leq i \leq p\}. \tag{4.1.2}$$

When $p = 2$, (4.1.1) becomes the classical OT problem.

Remark 4.1.1. *In [24], the authors consider the cost*

$$c(x_1, \ldots, x_p) = \sum_{j \neq k}^{p} |x_j - x_k|^2 \tag{4.1.3}$$

and the minimization problem

$$\mathrm{OT}'_{MM}(T) := \inf_{\mathcal{K}} \sum_{j \neq k}^{p} \int_{\mathbb{R}^d} \frac{|T_j(x) - T_k(x)|^2}{2} \, d\nu_1(x). \tag{4.1.4}$$

where $\mathcal{K}$ is the set of all p-tuples of maps $T = (T_1, \ldots, T_p)$ such that $T_i : \mathbb{R}^d \mapsto \mathbb{R}^d$ ($i = 1, \ldots, p$) are Borel measurable and satisfy $\nu_i = T_{i_\#}\nu_1$.

Theorem 4.1.2. *Given $\nu_i \in \mathcal{P}(\mathbb{R}^d)$ for each $i = \{1, \ldots, p\}$, and a lower semi-continuous cost $c : \mathbb{R}^{pd} \mapsto \mathbb{R}$, OT_{MM} has a solution.*

The proof of the existence of a minimizer for (4.1.1) is quite standard. Similar to the proof of existence of solutions for the classical OT Problem, since the set $\Pi(\nu_1, \ldots, \nu_p)$ is non-empty, convex and compact with respect to the weak topology and the functional $\gamma \mapsto \int c \, d\gamma$ is linear with respect to γ, we can guarantee the existence of a minimizer for (4.1.1).

Remark 4.1.3. *In [24], the authors have proven that if the measures $\nu_1, \ldots, \nu_p$ are vanishing on $(d-1)$-rectifiable sets and have finite second moments, then the MMOT Problem with cost $\sum_{j \neq k}^{p} |x_j - x_k|^2$ has a unique solution (See [24], Corollary 2.2).*

4.1.3 Duality

Similar to the dual formulation of the two-marginal OT problem, we define the dual problem of the MMOT Problem as follows:

Given probability measures $(\nu_1, \ldots, \nu_p)$ in $\mathcal{P}(\mathbb{R}^d)$, we consider the maximizing problem

$$\mathrm{OT}^*_{\mathrm{MM}}(u_1, \ldots, u_p) := \sup_{\mathcal{A}} \left\{ \sum_{i=1}^{p} \int_{\mathbb{R}^d} u_i(x_i)\ d\nu_i(x_i) \right\}. \tag{4.1.5}$$

Here, $\mathcal{A}$ is the set of all p-tuples of functions $(u_1, \ldots, u_p)$ such that $u_i \in L^1(\mathbb{R}^d, d\nu_i)$ and upper semi-continuous, and $\sum_{i=1}^{p} u_i(x_i) \le c(x_1, \ldots, x_p),\ \forall (x_1, \ldots, x_p) \in \mathbb{R}^{pd}$.

The duality results between (4.1.1) and (4.1.5) have been discussed in the literature. However, in [24], the authors provide duality results with extensive characterization. Therefore, we will present the duality results given in [24] for the cost given by (4.1.3).

Theorem 4.1.4. *([24], Theorem 2.1) Assume that $\nu_1, \ldots, \nu_p$ are non-negative Borel probability measures vanishing on $(d-1)$-rectifiable sets and having finite second moments. Set $X_i := \mathrm{spt}(\nu_i)$ for $i = 1, \ldots, p$. Then:*

(i) OT^*_{MM} *admits a maximizer $u = (u_1, \ldots, u_p) \in \mathcal{A}$.*

(ii) There is a minimizer $S = (S_1, \ldots, S_p)$ for (4.1.4) *satisfying $S_1(x) = x$ $(x \in \mathbb{R}^d)$. The S_i are one-to-one ν_i-a.e., are uniquely determined, and have the form $S_i(x) = \nabla f_i^*(\nabla f_1(x))$ $(x \in \mathbb{R}^d)$, where*

$$f_i(x) = \frac{|x|^2}{2} + \phi_i(x),$$

the ϕ_i are convex functions, and $f_i^* \in C^1(\mathbb{R}^d)$ where f_i^* denotes the Legendre transform of f_i (see [47], Section 26).

(iii) Duality holds: the optimal values in (4.1.4) *and* (4.1.5) *coincide.*

(iv) If $\bar{u} = (\bar{u}_1, \ldots, \bar{u}_p) \in \mathcal{A}$ is another maximizer for (4.1.5)*, we can modify the $\bar{u}_i$'s on sets of zero ν_i measure to obtain a maximizer, still denoted $\bar{u}$, such that $\bar{u}_i$ is differentiable ν_i-a.e. Furthermore,*

$$\nabla u_i = \nabla \bar{u}_i \quad \nu_i\text{-a.e.}$$

4.2 Capacity Constrained Multi-Marginal OT Problem

The notion of the capacity constrained multi-marginal OT Problem already has been introduced in [18]. The Capacity Constrained Multi-Marginal Optimal Transport (CCMMOT) problem introduces capacities that limit the amount transported between the source and the targets.

4.2.1 The Primal Problem

Suppose we are given a positive integer p, and p probability measures $\nu_1, \dots, \nu_p \in \mathcal{P}(\mathbb{R}^d)$. Let $\tilde{\gamma}$ be a compactly supported finite measure on $\mathbb{R}^{pd}$ such that $\tilde{\gamma} \ll \nu_1 \otimes \dots \otimes \nu_p$.

Define

$$\Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p) := \{\gamma \in \Pi(\nu_1, \dots, \nu_p) : \gamma(A_1 \times \dots \times A_p) \leq \tilde{\gamma}(A_1 \times \dots \times A_p), \forall A_i \in \mathcal{B}(\mathbb{R}^d), \forall i\}. \tag{4.2.1}$$

We assume that the set $\Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$ is non-empty.

Then, the CCMMOT problem is to minimize the cost:

$$\mathrm{OT}_{\mathrm{CCMM}} := \inf_{\gamma \in \Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)} \int_{\mathbb{R}^{pd}} c(x_1, \dots, x_p)\, d\gamma(x_1, \dots, x_p). \tag{4.2.2}$$

In [18], it has been shown that the CCMMOT problem has a solution and some characterization of the optimal solution is given. We will present some of the results from [18] below.

Theorem 4.2.1. *([18], Theorem 3.1) If $c \in L^\infty(\mathbb{R}^{pd})$, then OT_{CCMM} has a solution.*

Next, we will provide a characterization of the optimal solutions of (4.2.2).

Definition 4.2.2. *A measure $\gamma \in \Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$ is called an extreme point of the convex set $\Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$ if γ is not the midpoint of a non-trivial line segment in $\Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$.*

Theorem 4.2.3. *([18], Theorem 4.1) Suppose $\nu_1, \dots, \nu_p$ are non-atomic Borel probability measures on $\mathbb{R}^d$. Then, $\gamma \in \Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$ is an extreme point of the set $\Pi^{\tilde{\gamma}}(\nu_1, \dots, \nu_p)$ if and only if $\gamma = \mathbb{1}_W \tilde{\gamma}$ for a γ-measurable set $W \subset \mathbb{R}^{pd}$.*

To get the uniqueness of optimal solutions, we make the following two assumptions on the cost function.

1. Assume that the function $c \in L^\infty(\mathbb{R}^{pd})$ is such that the mixed partial derivative of order p
$$\partial_{i_1,\ldots,i_p} c = \frac{\partial^p c}{\partial x_1^{i_1}, \ldots, x_p^{i_p}}$$
exists for each set of variables $(x_1^{i_1}, \ldots, x_p^{i_p})$, where $1 \leq i_k \leq d$ for any $k \leq p$.

2. Assume that for some fixed number $N \in \mathbb{N} \cup \{\infty\}$ there exist at most countably many disjoint open sets $\{G_k\}_{k=1}^N,\ G_k \subseteq \mathbb{R}^d$, such that the following conditions are satisfied:

(C1) each set G_k has positive Lebesgue measure;

(C2) the union of all sets in $\{G_k\}_{k=1}^N$ has full Lebesgue measure;

(C3) for every $k \leq N$ there exist a set of variables $(x_1^{k_1}, \ldots, x_p^{k_p})$ such that the functions $\partial_{k_1,\ldots,k_p} c$ is either strictly positive or strictly negative on G_k.

Theorem 4.2.4. *([18], Theorem 6.1) Suppose we are given a cost function c on $\mathbb{R}^d$ satisfying the conditions (C1), (C2), and (C3). Then, any $\gamma \in \mathcal{P}(\mathbb{R}^{pd})$ that is an optimal plan is an extreme point of the set $\Pi^{\tilde{\gamma}}(\nu_1, \ldots, \nu_p)$.*

Corollary 4.2.5. *([18], Corollary 6.1.1) If c on $\mathbb{R}^d$ satisfies the conditions (C1), (C2), and (C3), then the optimal plan is unique.*

4.2.2 Duality

To the best of our knowledge, a dual formulation for the CCMMOT problem does not exist in the literature. Therefore, in this section we present a dual formulation for the CCMMOT problem and prove the strong duality result and the existence of dual maximizers.

We will generalize the techniques used in [34, 35] for the two-marginal capacity constrained OT problem to get our results.

Let ν_i be a probability measure in $\mathcal{P}(\mathbb{R}^d)$ which is absolutely continuous w.r.t. Lebesgue measure with density $f_i \in L^1(\mathbb{R}^d)$ for each $1 \le i \le p$ and $\tilde{\gamma}$ be a compactly supported finite measure that is absolutely continuous w.r.t. Lebesgue measure with a bounded density $\tilde{h}$. We denote by $\Pi^{\tilde{h}}(f_1, \dots, f_p)$ the set of all non-negative measurable joint densities on $\mathbb{R}^{pd}$ that are bounded by $\tilde{h}$. i.e. for each $1 \le i \le p$, $f_i(x_i) = \int h(x_1, \dots, x_p)\, dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_p$, and $0 \le h \le \tilde{h}$. As the dual formulation of the CCMMOT problem, we consider the following maximization problem:

$$\mathrm{OT}^*_{\mathrm{CCMM}} := \sup_{(u_1,\dots,u_p,w)\in \mathrm{Lip}_c^{\tilde{h}}} J(u_1, \dots, u_p, w) \tag{4.2.3}$$

where

$$J(u_1, \dots, u_p, w) := \sum_{i=1}^{p} \int_{\mathbb{R}^d} u_i(x_i) f_i(x_i)\, dx_i + \int_{\mathbb{R}^{pd}} w(x_1, \dots, x_p) \tilde{h}(x_1, \dots, x_p)\, dx_1 \dots dx_p \tag{4.2.4}$$

and

$$\begin{aligned} \mathrm{Lip}_c^{\tilde{h}} := \Bigg\{ & (u_1, \dots, u_p, w) : u_i \in L^1(\mathbb{R}^d), w \in L^1(\mathbb{R}^{pd}), \\ & \sum_{i=1}^{p} u_i(x_i) + w(x_1, \dots, x_p) \le c(x_1, \dots, x_p), \text{and } w(x_1, \dots, x_p) \le 0 \Bigg\}. \end{aligned} \tag{4.2.5}$$

First, we will prove that the strong duality holds for the CCMMOT problem.

Theorem 4.2.6. *Let ν_i be a probability measure in $\mathcal{P}(\mathbb{R}^d)$ which is absolutely continuous w.r.t. Lebesgue measure with density $f_i \in L^1(\mathbb{R}^d)$ for each $1 \le i \le p$, $\tilde{\gamma}$ be a compactly supported finite measure that is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^{pd}$ with a bounded density $\tilde{h}$, $c \in L^1(\mathbb{R}^{pd})$ and assume $\Pi^{\tilde{h}}(f_1, \dots, f_p) \neq \emptyset$. Then,*

$$\mathrm{OT}_{CCMM} = \mathrm{OT}^*_{CCMM}.$$

Here, we will redefine $\mathrm{OT}_{\mathrm{CCMM}}$ as

$$\mathrm{OT}_{\mathrm{CCMM}} = \inf_{h\in\Pi^{\tilde{h}}(f_1,\dots,f_p)} I_c(h) := \int_{\mathbb{R}^{pd}} ch\, dx_1 \dots dx_p. \tag{4.2.6}$$

Proof. The proof of Theorem 4.2.6 is a consequence of the two propositions that we will be proving below.

Proposition 4.2.7. *Under the hypotheses of Theorem 4.2.6, we have*

$$\mathrm{OT}_{CCMM} \geq \mathrm{OT}^*_{CCMM} \tag{4.2.7}$$

Proposition 4.2.8. *Under the hypotheses of Theorem 4.2.6, there exists a sequence* $\{(u_1^\varepsilon, \dots, u_p^\varepsilon, w^\varepsilon)\}_{\varepsilon\downarrow 0}$ *in* $Lip_c^{\tilde{h}}$ *such that*

$$I_c(h_0) = \lim_{\varepsilon\downarrow 0} J(u_1^\varepsilon, \dots, u_p^\varepsilon, w^\varepsilon), \tag{4.2.8}$$

where h_0 *is a minimizer of* (4.2.6) *of the form* $h_0 = \mathbb{1}_W \tilde{h}$*, for a Lebesgue measurable set* $W \subset \mathbb{R}^{pd}$*.*

Proof of Proposition 4.2.7: Let $h \in \Pi^{\tilde{h}}(f_1, \dots, f_p)$ with $I_c(h)$ finite and let $(u_1, \dots, u_p, w) \in \mathrm{Lip}_c^{\tilde{h}}$. Then we have

$$\begin{aligned}
I_c(h) &= \int_{\mathbb{R}^{pd}} ch\ dx_1 \dots dx_p \\
&= \sum_{i=1}^p \int_{\mathbb{R}^d} u_i f_i\ dx_i + \int_{\mathbb{R}^{pd}} w\tilde{h}\ dx_1 \dots dx_p + \int_{\mathbb{R}^{pd}} (c - \sum_{i=1}^p u_i - w)h\ dx_1 \dots dx_p \\
&\quad + \int_{\mathbb{R}^{pd}} w(h - \tilde{h})\ dx_1 \dots dx_p \\
&\geq J(u_1, \dots, u_p, w).
\end{aligned}$$

Note that, in the second line, we use the marginal conditions on h and in the last line, we used the properties of the set $\mathrm{Lip}_c^{\tilde{h}}$ along with the fact that $h \leq \tilde{h}$. By taking the infimum over $h \in \Pi^{\tilde{h}}(f_1, \dots, f_p)$ and supremum over $(u_1, \dots, u_p, w) \in \mathrm{Lip}_c^{\tilde{h}}$ from both sides, we get the inequality (4.2.7).

Proof of Proposition 4.2.8: Similar to [35], we introduce a *relaxed* version of the MMOT problem with capacity constraints. Let $\varepsilon > 0$ be a small number. Define

$$I_c^\varepsilon(h) := \int_{\mathbb{R}^{pd}} ch\ dx_1 \dots dx_p + \frac{1}{2\varepsilon} \sum_{i=1}^p || \langle h \rangle_{x_i} - f_i ||_2^2 \tag{4.2.9}$$

where $\langle g \rangle_{x_i}$ denotes the i^{th} marginal of g, for a given function $g = g(x_1, \ldots, x_p)$ defined on $\mathbb{R}^{pd}$, i.e. for each $1 \leq i \leq p$, $\langle g \rangle_{x_i} = \int g(x_1, \ldots, x_p)\ dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p$.

For $(u_1, \ldots, u_p, w) \in \text{Lip}_c^{\tilde{h}}$ such that each $u_i \in L^2(\mathbb{R}^d)$, we define

$$J^{\varepsilon}(u_1, \ldots, u_p, w) := \sum_{i=1}^{p} \int_{\mathbb{R}^d} u_i f_i\ dx_i + \int_{\mathbb{R}^{pd}} w\tilde{h}\ dx_1 \ldots dx_p - \frac{\varepsilon}{2} \sum_{i=1}^{p} ||u_i||_2^2. \tag{4.2.10}$$

Note that, if $u_i \notin L^2(\mathbb{R}^d)$ for some $i \in \{1, \ldots, p\}$, we can extend J^{ε} to a functional on the entire set $\text{Lip}_c^{\tilde{h}}$ by setting $J^{\varepsilon}(u_1, \ldots, u_p, w) := -\infty$.

First, we will prove that the statement of Proposition 4.2.7 holds for the relaxed version.

Lemma 4.2.9. *Let $\varepsilon > 0$ be given. Under the hypotheses of Theorem 4.2.6, we have*

$$\inf_{0 \leq h \leq \tilde{h}} I_c^{\varepsilon}(h) \geq \sup_{(u_1, \ldots, u_p, w) \in Lip_c^{\tilde{h}}} J^{\varepsilon}(u_1, \ldots, u_p, w). \tag{4.2.11}$$

Proof. Let $0 \leq h \leq \tilde{h}$ and $(u_1, \ldots, u_p, w) \in \text{Lip}_c^{\tilde{h}}$ be such that $I_c^{\varepsilon}(h)$ and $J^{\varepsilon}(u_1, \ldots, u_p, w)$ are both finite. Then, observe that,

$$\begin{aligned}
I_c^{\varepsilon}(h) &= \int_{\mathbb{R}^{pd}} ch\ dx_1 \ldots dx_p + \frac{1}{2\varepsilon} \sum_{i=1}^{p} || \langle h \rangle_{x_i} - f_i ||_2^2 \\
&= \sum_{i=1}^{p} \int_{\mathbb{R}^d} u_i f_i\ dx_i + \int_{\mathbb{R}^{pd}} w\tilde{h}\ dx_1 \ldots dx_p - \frac{\varepsilon}{2} \sum_{i=1}^{p} ||u_i||_2^2 \\
&\quad + \int_{\mathbb{R}^{pd}} (c - \sum_{i=1}^{p} u_i - w) h\ dx_1 \ldots dx_p + \int_{\mathbb{R}^{pd}} w(h - \tilde{h})\ dx_1 \ldots dx_p \\
&\quad + \frac{1}{2\varepsilon} \sum_{i=1}^{p} || \langle h \rangle_{x_i} - f_i + \varepsilon u_i ||_2^2 \\
&\geq J^{\varepsilon}(u_1, \ldots, u_p, w).
\end{aligned}$$

The last line holds true by the definition of the set $\text{Lip}_c^{\tilde{h}}$, the fact that $0 \leq h \leq \tilde{h}$ and the non-negativity of the L^2 norms.

Finally, by taking the infimum over $h \in \Pi^{\tilde{h}}(f_1, \ldots, f_p)$ and supremum over $(u_1, \ldots, u_p, w) \in \text{Lip}_c^{\tilde{h}}$ from both sides, we get the inequality (4.2.11). ■

Next, we will prove that the relaxed problem (4.2.9) has a minimizer.

Lemma 4.2.10. *Let $\varepsilon > 0$ be given. Under the hypotheses of Theorem 4.2.6, there exists a minimizer h_ε of I_c^ε, and h_ε can be chosen of the form $h_\varepsilon = \mathbb{1}_{W_\varepsilon}\tilde{h}$ for some Lebesgue measurable set $W_\varepsilon \subseteq \mathbb{R}^{pd}$. Furthermore, if $\hat{h}_\varepsilon$ is another minimizer of I_c^ε, then for each $1 \le i \le p$, $\langle h_\varepsilon \rangle_{x_i} = \left\langle \hat{h}_\varepsilon \right\rangle_{x_i}$.*

Proof. Let h_0 be a minimizer of (4.2.6). Since it is admissible for I_c^ε, and h_0 has marginals $f_1, \dots, f_p$, we have that $I_c^\varepsilon(h_0) = I_c(h_0)$. Thus, we have that, $-||\tilde{h}||_\infty ||c||_{L^1(\tilde{W})} \le \inf_h I_c^\varepsilon(h) \le I_c^\varepsilon(h_0) = I_c(h_0) < \infty$, where $\widetilde{W}$ is the support of $\tilde{h}$. So, we pick a minimizing sequence $\{h_n\}_{n\in\mathbb{N}}$ of I_c^ε. Since we have $0 \le h_n \le \tilde{h}$, for each n, we can find a subsequence (without relabeling) $\{h_n\}_{n\in\mathbb{N}}$ that converges weak-* in $(L^1(\mathbb{R}^{pd}))^* = L^\infty(\mathbb{R}^{pd})$ (see [48], Chapter 19) to some L^∞ function h_ε that satisfies $0 \le h_\varepsilon \le \tilde{h}$.

Also note that for each $1 \le i \le p$,

$$\langle h_n \rangle_{x_i} - f_i = \int_{\mathbb{R}^{(p-1)d}} (h_n - h_0)\; dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_p.$$

Hence,

$$\int_{\mathbb{R}^d} |\langle h_n \rangle_{x_i} - f_i|^2 \; dx_i \le \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^{(p-1)d}} |h_n - h_0| \; dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_p \right)^2 \; dx_i.$$

Since $h_n, h_0 \le \tilde{h} \in L^\infty(\mathbb{R}^{pd})$ with compact support, the sequence $\{\langle h_n \rangle_{x_i} - f_i\}_{n\in\mathbb{N}}$ is bounded in L^2. Hence, we can find a further subsequence (without relabeling) $\{h_n\}_{n\in\mathbb{N}}$ such that the sequences $\{\langle h_n \rangle_{x_i} - f_i\}_{n\in\mathbb{N}}$ weakly converge to $\left(\langle h_\varepsilon \rangle_{x_i} - f_i\right)$ in L^2 for each i.

Now, fix an $i \in \{1, \dots, p\}$ and let $\xi = \xi(x_i)$ be an arbitrary smooth and compactly supported test function. Then,

$$\begin{aligned}\int_{\mathbb{R}^d} [\left(\langle h_n \rangle_{x_i} - f_i\right) - \left(\langle h_\varepsilon \rangle_{x_i} - f_i\right)]\xi \; dx_i &= \int_{\mathbb{R}^d} \left(\langle h_n \rangle_{x_i} - \langle h_\varepsilon \rangle_{x_i}\right) \xi \; dx_i \\ &= \int_{\mathbb{R}^{pd}} (h_n - h_\varepsilon)\xi \; dx_1 \dots dx_p.\end{aligned}$$

Since h_n converges to h_ε weak-*, we get that $\langle h_n \rangle_{x_i} - f_i$ converges to $\langle h_\varepsilon \rangle_{x_i} - f_i$.

Now, by the lower semi-continuity of the L^2 norm with respect to weak L^2 convergence, for each i, we have that

$$|| \langle h_\varepsilon \rangle_{x_i} - f_i ||_{L^2(\mathbb{R}^d)} \le \liminf_{n\to\infty} || \langle h_n \rangle_{x_i} - f_i ||_{L^2(\mathbb{R}^d)}.$$

Also, since $c \in L^1(\mathbb{R}^{pd})$, and h_ε, h_n are supported in $\widetilde{W}$, by the weak-* convergence on $L^\infty(\widetilde{W})$, we get

$$\int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1 \dots dx_p = \lim_{n\to\infty} \int_{\mathbb{R}^{pd}} ch_n \; dx_1 \dots dx_p.$$

Thus, we have that

$$\begin{aligned}
\inf_h I_c^\varepsilon(h) &= \liminf_{n\to\infty} I_c^\varepsilon(h_n) \\
&= \liminf_{n\to\infty} \left\{ \int_{\mathbb{R}^{pd}} ch_n \; dx_1 \dots dx_p + \frac{1}{2\varepsilon} \sum_{i=1}^p || \langle h_n \rangle_{x_i} - f_i ||_2^2 \right\} \\
&\geq \liminf_{n\to\infty} \int_{\mathbb{R}^{pd}} ch_n \; dx_1 \dots dx_p + \frac{1}{2\varepsilon} \sum_{i=1}^p \liminf_{n\to\infty} || \langle h_n \rangle_{x_i} - f_i ||_2^2 \\
&\geq \int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1 \dots dx_p + \frac{1}{2\varepsilon} \sum_{i=1}^p || \langle h_\varepsilon \rangle_{x_i} - f_i ||_2^2 \\
&= I_c^\varepsilon(h_\varepsilon).
\end{aligned}$$

This concludes that h_ε is a minimizer for I_c^ε. Since the relaxed problem is strictly convex (see Appendix 2), it is clear that h_ε has unique marginals. Also, since h_ε is a minimizer for I_c in the class $\Pi^{\tilde{h}}(\langle h_\varepsilon \rangle_{x_1}, \dots, \langle h_\varepsilon \rangle_{x_p})$, we can choose h_ε such that $h_\varepsilon = \mathbb{1}_{W_\varepsilon} \tilde{h}$ for some Lebesgue measurable set $W_\varepsilon \subset \widetilde{W}$ (see Theorem 4.2.3 and Theorem 4.2.4). ■

For the rest of the proof, we define the following functions.

$$u_i^\varepsilon := -\frac{1}{\varepsilon}(\langle h_\varepsilon \rangle_{x_i} - f_i) \qquad \forall 1 \leq i \leq p, \tag{4.2.12}$$

and

$$w^\varepsilon := \min \left\{ c - \sum_{i=1}^p u_i^\varepsilon, 0 \right\}. \tag{4.2.13}$$

Note that, by the definition of w^ε, we get that $w^\varepsilon \leq 0$ and $\sum_{i=1}^p u_i^\varepsilon + w^\varepsilon \leq c$. Thus, $(u_1^\varepsilon, \dots, u_p^\varepsilon, w^\varepsilon) \in \mathrm{Lip}_c^{\tilde{h}}$ defined in (4.2.5). Also note that for each i, u_i^ε are determined independently of the choice of h_ε in $\Pi^{\tilde{h}}(f_1, \dots, f_p)$. We will claim that $(u_1^\varepsilon, \dots, u_p^\varepsilon, w^\varepsilon)$ maximizes J^ε in $\mathrm{Lip}_c^{\tilde{h}}$.

Lemma 4.2.11. *Let $h_\varepsilon = \mathbb{1}_{W_\varepsilon}\tilde{h}$ be a minimizer for I_c^ε. Then*

$$c - \sum_{i=1}^{p} u_i^\varepsilon \begin{cases} \leq 0 & a.e.\ in\ W_\varepsilon, \\ \geq 0 & a.e.\ in\ \widetilde{W} \setminus W_\varepsilon. \end{cases} \tag{4.2.14}$$

Proof. Let $\xi \geq 0$ be an arbitrary smooth test function. Define, for any $\sigma \in \mathbb{R}$

$$h_\varepsilon^\sigma := h_\varepsilon + \sigma\xi(\tilde{h} - h_\varepsilon).$$

Since $h_\varepsilon = \mathbb{1}_{W_\varepsilon}\tilde{h}$, we have

$$h_\varepsilon^\sigma = \begin{cases} h_\varepsilon & \text{a.e. in } W_\varepsilon, \\ \sigma\xi\tilde{h} & \text{a.e. in } \widetilde{W} \setminus W_\varepsilon. \end{cases} \tag{4.2.15}$$

Observe that, $h_\varepsilon^0 = h_\varepsilon$ and for $0 \leq \sigma \leq ||\xi||_\infty^{-1},\ 0 \leq h_\varepsilon^\sigma \leq \tilde{h}$. Since h_ε minimizes I_c^ε, we have $I_c^\varepsilon(h_\varepsilon) = I_c^\varepsilon(h_\varepsilon^0) \leq I_c^\varepsilon(h_\varepsilon^\sigma)$.

Now observe that,

$$\begin{aligned} I_c^\varepsilon(h_\varepsilon^\sigma) &= \int_{\mathbb{R}^{pd}} c h_\varepsilon^\sigma \ dx_1 \dots dx_p + \frac{1}{2\varepsilon}\sum_{i=1}^{p} || \left\langle h_\varepsilon^\sigma \right\rangle_{x_i} - f_i||_2^2 \\ &= \int_{\mathbb{R}^{pd}} c(h_\varepsilon + \sigma\xi(\tilde{h} - h_\varepsilon)) \ dx_1 \dots dx_p + \frac{1}{2\varepsilon}\sum_{i=1}^{p} || \Big\langle (h_\varepsilon + \sigma\xi(\tilde{h} - h_\varepsilon)) \Big\rangle_{x_i} - f_i||_2^2 \\ &= \int_{\mathbb{R}^{pd}} c h_\varepsilon \ dx_1 \dots dx_p + \int_{\mathbb{R}^{pd}} c\sigma\xi(\tilde{h} - h_\varepsilon) \ dx_1 \dots dx_p + \frac{1}{2\varepsilon}\sum_{i=1}^{p} || \left\langle h_\varepsilon \right\rangle_{x_i} - f_i||_2^2 \\ &\quad + \frac{1}{2\varepsilon}\sum_{i=1}^{p} \int_{\mathbb{R}^d} \left| \sigma\xi \Big\langle (\tilde{h} - h_\varepsilon) \Big\rangle_{x_i} \right|^2 \ dx_i \\ &\quad + \frac{1}{\varepsilon}\sum_{i=1}^{p} \int_{\mathbb{R}^d} \left(\sigma\xi \Big\langle (\tilde{h} - h_\varepsilon) \Big\rangle_{x_i} \right) \cdot \left(\left\langle h_\varepsilon \right\rangle_{x_i} - f_i \right) \ dx_i. \end{aligned}$$

Then, the right derivative of I_c^ε at $\sigma = 0$ is given by

$$\begin{aligned} \partial_+ I_c^\varepsilon(0) &= \lim_{\sigma \to 0^+} \frac{I_c^\varepsilon(\sigma) - I_c^\varepsilon(0)}{\sigma} \\ &= \lim_{\sigma \to 0^+} \frac{1}{\sigma} \Bigg\{ \int_{\mathbb{R}^{pd}} c h_\varepsilon \ dx_1 \dots dx_p + \int_{\mathbb{R}^{pd}} c\sigma\xi(\tilde{h} - h_\varepsilon) \ dx_1 \dots dx_p \\ &\quad + \frac{1}{2\varepsilon}\sum_{i=1}^{p} || \left\langle h_\varepsilon \right\rangle_{x_i} - f_i||_2^2 + \frac{1}{2\varepsilon}\sum_{i=1}^{p} \int_{\mathbb{R}^d} \left| \sigma\xi \Big\langle (\tilde{h} - h_\varepsilon) \Big\rangle_{x_i} \right|^2 \ dx_i \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\varepsilon}\sum_{i=1}^{p}\int_{\mathbb{R}^d}\left(\sigma\xi\left\langle(\tilde{h}-h_\varepsilon)\right\rangle_{x_i}\right)\cdot\left(\langle h_\varepsilon\rangle_{x_i}-f_i\right)\,dx_i\\
&-\int_{\mathbb{R}^{pd}}ch_\varepsilon\,dx_1\dots dx_p-\frac{1}{2\varepsilon}\sum_{i=1}^{p}||\langle h_\varepsilon\rangle_{x_i}-f_i||_2^2\Bigg\}\\
=&\lim_{\sigma\to 0^+}\left\{\int_{\mathbb{R}^{pd}}c\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p+\frac{\sigma}{2\varepsilon}\sum_{i=1}^{p}\int_{\mathbb{R}^d}\left|\xi\left\langle(\tilde{h}-h_\varepsilon)\right\rangle_{x_i}\right|^2\,dx_i\right.\\
&\left.+\frac{1}{\varepsilon}\sum_{i=1}^{p}\int_{\mathbb{R}^d}\left(\xi\left\langle(\tilde{h}-h_\varepsilon)\right\rangle_{x_i}\right)\cdot\left(\langle h_\varepsilon\rangle_{x_i}-f_i\right)\,dx_i\right\}\\
=&\int_{\mathbb{R}^{pd}}c\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p+\frac{1}{\varepsilon}\sum_{i=1}^{p}\int_{\mathbb{R}^d}\left(\xi\left\langle(\tilde{h}-h_\varepsilon)\right\rangle_{x_i}\right)\cdot\left(\langle h_\varepsilon\rangle_{x_i}-f_i\right)\,dx_i\\
=&\int_{\mathbb{R}^{pd}}c\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p+\frac{1}{\varepsilon}\sum_{i=1}^{p}\int_{\mathbb{R}^{pd}}\left(\langle h_\varepsilon\rangle_{x_i}-f_i\right)\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p\\
=&\int_{\mathbb{R}^{pd}}\left(c-\sum_{i=1}^{p}u_i^\varepsilon\right)\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p.
\end{aligned}
$$

Since $h_\varepsilon^0=h_\varepsilon$ minimizes I_c^ε, we have

$$0\leq\partial_+I_c^\varepsilon(0)=\int_{\mathbb{R}^{pd}}\left(c-\sum_{i=1}^{p}u_i^\varepsilon\right)\xi(\tilde{h}-h_\varepsilon)\,dx_1\dots dx_p.$$

As this holds for any arbitrary test function $\xi\geq 0$, we have that $\left(c-\sum_{i=1}^{p}u_i^\varepsilon\right)(\tilde{h}-h_\varepsilon)\geq 0$ a.e..

Since $\tilde{h}-h_\varepsilon\geq 0$ a.e. and $\tilde{h}-h_\varepsilon>0$ a.e in $\widetilde{W}\setminus W_\varepsilon$, we get the second part of the inequality in (4.2.14).

With a similar argument for $h_\varepsilon^\sigma:=h_\varepsilon-\sigma\xi h_\varepsilon$, we can prove the first inequality in (4.2.14).

■

Next, we will prove a duality result for the relaxed version.

Lemma 4.2.12. *Let $h_\varepsilon=\mathbb{1}_{W_\varepsilon}\tilde{h}$ be a minimizer for I_c^ε and $u_i^\varepsilon,w^\varepsilon$ be defined by (4.2.12) and (4.2.13). Then*

$$I_c^\varepsilon(h_\varepsilon)=J^\varepsilon(u_1^\varepsilon,\dots,u_p^\varepsilon,w^\varepsilon).\tag{4.2.16}$$

In particular, $(u_1^\varepsilon,\dots,u_p^\varepsilon,w^\varepsilon)$ is a maximizer for $J^\varepsilon(u_1,\dots,u_p,w)$ in $Lip_c^{\tilde{h}}$.

Proof. Observe that,

$$
\begin{aligned}
J^\varepsilon(u_1^\varepsilon,\dots,u_p^\varepsilon,w^\varepsilon) &= \sum_{i=1}^p \int_{\mathbb{R}^d} u_i^\varepsilon f_i \; dx_i + \int_{\mathbb{R}^{pd}} w^\varepsilon \tilde{h} \; dx_1\dots dx_p - \frac{\varepsilon}{2}\sum_{i=1}^p ||u_i^\varepsilon||_2^2 \\
&= \int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1\dots dx_p - \int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1\dots dx_p + \int_{\mathbb{R}^{pd}} w^\varepsilon h_\varepsilon \; dx_1\dots dx_p \\
&\quad - \int_{\mathbb{R}^{pd}} w^\varepsilon h_\varepsilon \; dx_1\dots dx_p + \sum_{i=1}^p \int_{\mathbb{R}^d} u_i^\varepsilon \left\langle h_\varepsilon\right\rangle_{x_i} \; dx_i - \sum_{i=1}^p \int_{\mathbb{R}^d} u_i^\varepsilon \left\langle h_\varepsilon\right\rangle_{x_i} \; dx_i \\
&\quad + \sum_{i=1}^p \int_{\mathbb{R}^d} u_i^\varepsilon f_i \; dx_i + \int_{\mathbb{R}^{pd}} w^\varepsilon \tilde{h} \; dx_1\dots dx_p - \frac{\varepsilon}{2}\sum_{i=1}^p ||u_i^\varepsilon||_2^2 \\
&= \int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1\dots dx_p - \int_{\mathbb{R}^{pd}} \left(c - \sum_{i=1}^p u_i^\varepsilon - w^\varepsilon\right) h_\varepsilon \; dx_1\dots dx_p \\
&\quad + \int_{\mathbb{R}^{pd}} w^\varepsilon(\tilde{h} - h_\varepsilon) \; dx_1\dots dx_p + \sum_{i=1}^p \int_{\mathbb{R}^d} u_i^\varepsilon (f_i - \left\langle h_\varepsilon\right\rangle_{x_i}) \; dx_i \\
&\quad - \frac{\varepsilon}{2}\sum_{i=1}^p ||u_i^\varepsilon||_2^2 \\
&= \int_{\mathbb{R}^{pd}} ch_\varepsilon \; dx_1\dots dx_p + \frac{1}{2\varepsilon}\sum_{i=1}^p || \left\langle h_\varepsilon\right\rangle_{x_i} - f_i||_2^2 \\
&\quad + \int_{\mathbb{R}^{pd}} w^\varepsilon(\tilde{h} - h_\varepsilon) \; dx_1\dots dx_p \\
&\quad - \int_{\mathbb{R}^{pd}} \left(c - \sum_{i=1}^p u_i^\varepsilon - w^\varepsilon\right) h_\varepsilon \; dx_1\dots dx_p \\
&= I_c^\varepsilon(h_\varepsilon) + \int_{\mathbb{R}^{pd}} w^\varepsilon(\tilde{h} - h_\varepsilon) \; dx_1\dots dx_p \\
&\quad - \int_{\mathbb{R}^{pd}} \left(c - \sum_{i=1}^p u_i^\varepsilon - w^\varepsilon\right) h_\varepsilon \; dx_1\dots dx_p.
\end{aligned} \tag{4.2.17}
$$

From (4.2.14) and (4.2.13), we observe that, on W_ε, $h_\varepsilon = \tilde{h}$, hence $w^\varepsilon(\tilde{h} - h_\varepsilon) = 0$ and since $c - \sum_{i=1}^p u_i^\varepsilon \leq 0$, $w^\varepsilon = c - \sum_{i=1}^p u_i^\varepsilon$, we obtain $c - \sum_{i=1}^p u_i^\varepsilon - w^\varepsilon = 0$. On the other hand, on $\widetilde{W} \setminus W_\varepsilon$, $h_\varepsilon = 0$, hence $(c - \sum_{i=1}^p u_i^\varepsilon - w^\varepsilon)h_\varepsilon = 0$, and since $c - \sum_{i=1}^p u_i^\varepsilon \geq 0$, $w^\varepsilon = 0$, we obtain $w^\varepsilon(\tilde{h} - h_\varepsilon) = 0$.

Thus, (4.2.17) becomes

$$J^\varepsilon(u_1^\varepsilon,\dots,u_p^\varepsilon,w^\varepsilon) \geq I_c^\varepsilon(h_\varepsilon).$$

By (4.2.11) and the fact that $(u_1^\varepsilon, \ldots, u_p^\varepsilon, w^\varepsilon) \in \mathrm{Lip}_c^{\tilde{h}}$, we can conclude that $(u_1^\varepsilon, \ldots, u_p^\varepsilon, w^\varepsilon)$ is a maximizer for $J^\varepsilon(u_1^\varepsilon, \ldots, u_p^\varepsilon, w^\varepsilon)$ in $\mathrm{Lip}_c^{\tilde{h}}$.

■

Now, we will prove that the minimizer of the relaxed problem (4.2.9) approximates the original problem (4.2.6).

Lemma 4.2.13. *Let $\{h_\varepsilon\}_{\varepsilon \downarrow 0}$ be a sequence of minimizers for I_c^ε. Then, it is precompact in the L^∞-weak-* topology and every limit point h_0 is a minimizer of I_c. Furthermore,*

$$\lim_{\varepsilon \downarrow 0} I_c(h_\varepsilon) = I_c(h_0), \tag{4.2.18}$$

$$\lim_{\varepsilon \downarrow 0} \varepsilon ||u_i^\varepsilon||_2^2 = 0, \qquad \forall 1 \leq i \leq p. \tag{4.2.19}$$

Proof. Since we have $0 \leq h_\varepsilon \leq \tilde{h}$, there exists a subsequence $\{h_{\varepsilon_n}\}_{n \in \mathbb{N}}$ that converges weakly-* in $L^\infty(\widetilde{W})$ to some function $0 \leq \bar{h} \leq \tilde{h}$.

Let $\bar{h}_0$ be a minimizer of I_c. Note that, $\bar{h}_0$ is admissible for the relaxed problem and since h_{ε_n} is a minimizer for the relaxed problem, we have

$$I_c^{\varepsilon_n}(h_{\varepsilon_n}) \leq I_c^{\varepsilon_n}(\bar{h}_0) = I_c(\bar{h}_0). \tag{4.2.20}$$

By the weak-* convergence of $\{h_{\varepsilon_n}\}$ in $L^\infty(\widetilde{W})$, we have that

$$\int_{\mathbb{R}^{pd}} c\bar{h} \; dx_1 \ldots dx_p = \lim_{\varepsilon_n \downarrow 0} \int_{\mathbb{R}^{pd}} c h_{\varepsilon_n} \; dx_1 \ldots dx_p.$$

i.e.

$$I_c(\bar{h}) = \lim_{\varepsilon_n \downarrow 0} I_c(h_{\varepsilon_n}). \tag{4.2.21}$$

Since $I_c(h_{\varepsilon_n}) \leq I_c^{\varepsilon_n}(h_{\varepsilon_n})$, we have

$$\liminf_{\varepsilon_n \downarrow 0} I_c(h_{\varepsilon_n}) \leq \liminf_{\varepsilon_n \downarrow 0} I_c^{\varepsilon_n}(h_{\varepsilon_n}). \tag{4.2.22}$$

By taking the $\liminf$ in (4.2.20), we have

$$\liminf_{\varepsilon_n \downarrow 0} I_c^{\varepsilon_n}(h_{\varepsilon_n}) \leq I_c(\bar{h}_0). \tag{4.2.23}$$

Thus, by combining (4.2.21), (4.2.22), and (4.2.23)

$$I_c(\bar{h}) \le I_c(\bar{h}_0). \tag{4.2.24}$$

Now, we will claim that $\bar{h}$ has the marginals $f_1, \ldots, f_p$.

From (4.2.20), we have

$$\int_{\mathbb{R}^{pd}} c h_{\varepsilon_n} \, dx_1 \ldots dx_p + \frac{1}{2\varepsilon_n} \sum_{i=1}^{p} || \langle h_{\varepsilon_n} \rangle_{x_i} - f_i ||_2^2 \le \int_{\mathbb{R}^{pd}} c\bar{h}_0 \, dx_1 \ldots dx_p.$$

So,

$$\sum_{i=1}^{p} || \langle h_{\varepsilon_n} \rangle_{x_i} - f_i ||_2^2 \le 2\varepsilon_n \int_{\mathbb{R}^{pd}} c(\bar{h}_0 - h_{\varepsilon_n}) \, dx_1 \ldots dx_p.$$

Note that, since $\bar{h}_0, h_{\varepsilon_n} \in L^\infty(R^{pd})$ with $0 \le h_{\varepsilon_n} \le \tilde{h}, \forall \varepsilon_n$, and $c \in L^1(\mathbb{R}^{pd})$, we have $\int_{\mathbb{R}^{pd}} c(\bar{h}_0 - h_{\varepsilon_n}) \, dx_1 \ldots dx_p \le 2||\tilde{h}||_{L^\infty(\mathbb{R}^{pd})}||c||_{L^1(\mathbb{R}^{pd})}$.

Hence, $\sup_n \int_{\mathbb{R}^{pd}} c(\bar{h}_0 - h_{\varepsilon_n}) \, dx_1 \ldots dx_p < \infty$.

Therefore, as $\varepsilon_n \downarrow 0$, for each i,

$$\langle h_{\varepsilon_n} \rangle_{x_i} \to f_i \quad \text{in } L^2. \tag{4.2.25}$$

Now, fix an $i \in \{1, \ldots, p\}$ and let $\xi = \xi(x_i)$ be an arbitrary smooth and compactly supported test function. Then,

$$\begin{aligned}
\int_{\mathbb{R}^d} \left(\langle \bar{h} \rangle_{x_i} - f_i \right) \xi \, dx_i &= \int_{\mathbb{R}^d} \left(\langle h_{\varepsilon_n} \rangle_{x_i} - f_i \right) \xi \, dx_i + \int_{\mathbb{R}^d} \left(\langle \bar{h} \rangle_{x_i} - \langle h_{\varepsilon_n} \rangle_{x_i} \right) \xi \, dx_i \\
&= \int_{\mathbb{R}^d} \left(\langle h_{\varepsilon_n} \rangle_{x_i} - f_i \right) \xi \, dx_i + \int_{\mathbb{R}^{pd}} (\bar{h} - h_{\varepsilon_n}) \xi \, dx_1 \ldots dx_p \\
&\le \int_{\mathbb{R}^d} | \left(\langle h_{\varepsilon_n} \rangle_{x_i} - f_i \right) \xi | \, dx_i + \int_{\mathbb{R}^{pd}} (\bar{h} - h_{\varepsilon_n}) \xi \, dx_1 \ldots dx_p \\
&\le || \langle h_{\varepsilon_n} \rangle_{x_i} - f_i ||_{L^2(\mathbb{R}^d)} ||\xi||_{L^2(\mathbb{R}^d)} + \int_{\mathbb{R}^{pd}} (\bar{h} - h_{\varepsilon_n}) \xi \, dx_1 \ldots dx_p.
\end{aligned}$$

In the last line, the first integral on the right converges to zero by (4.2.25) and the second integral converges to zero by the weak-* convergence of h_{ε_n}. Since ξ is arbitrary, we get that for each i, $\langle \bar{h} \rangle_{x_i} = f_i$. Thus, we have $\bar{h} \in \Pi^{\tilde{h}}(f_1, \ldots, f_p)$. Hence, by (4.2.24), we have that

$$I_c(\bar{h}) = I_c(\bar{h}_0) = \min_{h \in \Pi^{\tilde{h}}(f_1, \ldots, f_p)} I_c(h). \tag{4.2.26}$$

Thus, we have

$$\min_{h\in\Pi^{\tilde{h}}(f_1,\ldots,f_p)} I_c(h) = I_c(\bar{h}) = \liminf_{\varepsilon_n\downarrow 0} I_c(h_{\varepsilon_n}) \le \liminf_{\varepsilon_n\downarrow 0} I_c^{\varepsilon_n}(h_{\varepsilon_n}) \le I_c(\bar{h}_0) = \min_{h\in\Pi^{\tilde{h}}(f_1,\ldots,f_p)} I_c(h).$$

So, we can find a subsequence $\{h_{\varepsilon_n}\}_{n\in\mathbb{N}}$ (not relabeled), such that

$$\min_{h\in\Pi^{\tilde{h}}(f_1,\ldots,f_p)} I_c(h) = \lim_{\varepsilon_n\downarrow 0} I_c(h_{\varepsilon_n}) = \lim_{\varepsilon_n\downarrow 0} I_c^{\varepsilon_n}(h_{\varepsilon_n}). \tag{4.2.27}$$

Therefore,

$$\begin{aligned} I_c(\bar{h}) &= \lim_{\varepsilon_n\downarrow 0}\left\{\int_{\mathbb{R}^{pd}} ch_{\varepsilon_n}\ dx_1\ldots dx_p + \frac{1}{2\varepsilon_n}\sum_{i=1}^{p} ||\left\langle h_{\varepsilon_n}\right\rangle_{x_i} - f_i||_2^2\right\} \\ &= \lim_{\varepsilon_n\downarrow 0}\left\{\int_{\mathbb{R}^{pd}} ch_{\varepsilon_n}\ dx_1\ldots dx_p\right\} + \lim_{\varepsilon_n\downarrow 0}\left\{\frac{1}{2\varepsilon_n}\sum_{i=1}^{p} ||\left\langle h_{\varepsilon_n}\right\rangle_{x_i} - f_i||_2^2\right\} \\ &= I_c(\bar{h}) + \lim_{\varepsilon_n\downarrow 0}\left\{\frac{1}{2\varepsilon_n}\sum_{i=1}^{p} ||\left\langle h_{\varepsilon_n}\right\rangle_{x_i} - f_i||_2^2\right\}. \end{aligned}$$

This gives us that

$$\lim_{\varepsilon_n\downarrow 0}\left\{\frac{1}{\varepsilon_n}\sum_{i=1}^{p} ||\left\langle h_{\varepsilon_n}\right\rangle_{x_i} - f_i||_2^2\right\} = 0.$$

Using the definition of $u_i^{\varepsilon_n}$ in (4.2.12), we get

$$\lim_{\varepsilon_n\downarrow 0}\sum_{i=1}^{p}\varepsilon_n||u_i^{\varepsilon_n}||_2^2 = 0.$$

Hence, for each $1 \le i \le p$,

$$\lim_{\varepsilon_n\downarrow 0}\varepsilon_n||u_i^{\varepsilon_n}||_2^2 = 0.$$

This completes the proof of Lemma 4.2.13. ■

Finally, rewriting (4.2.16) using $J(u_1^{\varepsilon_n},\ldots,u_p^{\varepsilon_n},w^{\varepsilon_n})$ and $I_c(h_{\varepsilon_n})$, we get that

$$J(u_1^{\varepsilon_n},\ldots,u_p^{\varepsilon_n},w^{\varepsilon_n}) = I_c(h_{\varepsilon_n}) + \sum_{i=1}^{p}\varepsilon_n||u_i^{\varepsilon_n}||_2^2.$$

Letting $\varepsilon_n \downarrow 0$, we get

$$\lim_{\varepsilon_n\downarrow 0} J(u_1^{\varepsilon_n},\ldots,u_p^{\varepsilon_n},w^{\varepsilon_n}) = \lim_{\varepsilon_n\downarrow 0} I_c(h_{\varepsilon_n}) = I_c(h_0).$$

This completes the proof of Proposition 4.2.8, hence the proof of Theorem 4.2.6.

■

Remark 4.2.14. *Note that the assumption of $\tilde{h}$ being compactly supported, automatically gives us that the densities f_i's are compactly supported.*

Next, we will prove the existence of dual maximizers for the CCMMOT problem.

Recall that the CCMMOT problem is given by

$$\mathrm{OT}_{\mathrm{CCMM}} := \inf_{\gamma \in \Pi^{\tilde{\gamma}}(\nu_1,\ldots,\nu_p)} \int_{\mathbb{R}^{pd}} c(x_1,\ldots,x_p)\ d\gamma(x_1,\ldots,x_p)$$

and the dual problem is given by

$$\mathrm{OT}^*_{\mathrm{CCMM}} := \sup_{(u_1,\ldots,u_p,w)\in \mathrm{Lip}_c^{\tilde{h}}} J(u_1,\ldots,u_p,w),$$

where

$$J(u_1,\ldots,u_p,w) := \sum_{i=1}^{p} \int_{\mathbb{R}^d} u_i(x_i) f_i(x_i)\ dx_i + \int_{\mathbb{R}^{pd}} w(x_1,\ldots,x_p)\tilde{h}(x_1,\ldots,x_p)\ dx_1 \ldots dx_p.$$

Let $\tilde{h} \in L^\infty(X_1 \times \ldots \times X_p)$, and the probability densities $f_i \in L^1(X_i)$, for each $1 \le i \le p$ are compactly supported. We will assume that the sets $X_i = \mathrm{spt}(f_i)$ have unit Lebesgue measure.

In order to get the existence result, we will consider the following dual formulation:

$$\mathrm{OT}^{*'}_{\mathrm{CCMM}} := \sup_{u_i \in L^1(X_i), \forall 1 \le i \le p} J'(u_1,\ldots,u_p), \tag{4.2.28}$$

where

$$J'(u_1,\ldots,u_p) := \sum_{i=1}^{p} \int_{X_i} u_i f_i\ dx_i - \int_{X_1\times\ldots\times X_p} \left[-c + \sum_{i=1}^{p} u_i\right]_+ \tilde{h}\ dx_1 \ldots dx_p. \tag{4.2.29}$$

Note that, (see Appendix 1)

$$\mathrm{OT}^*_{\mathrm{CCMM}} = \mathrm{OT}^{*'}_{\mathrm{CCMM}}.$$

Now, we will show that (4.2.28) has a solution.

Theorem 4.2.15. *Let f_i and $\tilde{h}$ be continuous and strictly positive on their compact supports $X_i \subseteq \mathbb{R}^d$ and $X_1 \times \ldots \times X_p$ respectively. Let $c \in L^1(X_1 \times \ldots \times X_p)$. Fix an $\eta > 1$ and* *assume that* $\Pi^{\tilde{h}/\eta}(f_1,\dots,f_p)$ *is non-empty. Then, there exist functions* $(u_1,\dots,u_p)$ *where* $u_i \in L^1(X_i)\ \forall i$*, such that*

$$\mathrm{OT}^{*'}_{CCMM} = J'(u_1,\dots,u_p).$$

As in [34], our main goal will be to get a coercivity estimate of maximizing sequences which will guarantee L^1-boundedness.

For each $1 \le i \le p$, let

$$\overline{u_i f_i} := \int_{\mathbb{R}^d} u_i(x_i) f_i(x_i)\ dx_i.$$

Note that for a p-tuple of constants $(k_1,\dots,k_p)$ with $\sum_{i=1}^p k_i = 0$, we have

$$J'(u_1,\dots,u_p) = J'(u_1+k_1,\dots,u_p+k_p).$$

Hence, we can find a p-tuple of constants $(k_1,\dots,k_p)$ with $\sum_{i=1}^p k_i = 0$, so that

$$\overline{(u_i+k_i)f_i} = \overline{(u_j+k_j)f_j},\ \forall i \ne j.$$

First, we will find a bound on the means $\overline{u_i f_i}$.

Lemma 4.2.16. *Fix* $u_i \in L^1(X_i), c \in L^1(X_1\times\ldots\times X_p)$ *and a probability density* $h \in L^\infty(X_1\times\ldots\times X_p)$ *with marginals* $(f_1,\dots,f_p)$*. Suppose there is some* $\eta > 1$ *such that* $h \le \frac{\tilde{h}}{\eta}$*. Then,*

$$J'(u_1,\dots,u_p) \le \sum_{i=1}^p \overline{u_i f_i} \le \frac{||\eta h c||_{L^1(X_1\times\ldots\times X_p)} - J'(u_1,\dots,u_p)}{\eta - 1}. \tag{4.2.30}$$

Proof. From definition (4.2.29), we have

$$J'(u_1,\dots,u_p) = \sum_{i=1}^p \int_{X_i} u_i f_i\ dx_i - \int_{X_1\times\ldots\times X_p}\left[-c + \sum_{i=1}^p u_i\right]_+ \tilde{h}\ dx_1\ldots dx_p.$$

By the non-positivity of the second integral, we get one direction of the inequality:

$$J'(u_1,\dots,u_p) \le \sum_{i=1}^p \overline{u_i f_i}. \tag{4.2.31}$$

On the other hand, since $0 \le h \le \frac{\tilde{h}}{\eta}$ and the fact that $(-c+\sum_{i=1}^p u_i) \le [-c+\sum_{i=1}^p u_i]_+$, we have

$$J'(u_1,\dots,u_p) = \sum_{i=1}^p \overline{u_i f_i} - \int_{X_1\times\ldots\times X_p}\left[-c + \sum_{i=1}^p u_i\right]_+ \tilde{h}\ dx_1\ldots dx_p$$

$$\leq \sum_{i=1}^{p} \overline{u_i f_i} - \eta \int_{X_1\times\ldots\times X_p} \left(-c + \sum_{i=1}^{p} u_i\right) h \; dx_1 \ldots dx_p$$
$$= \sum_{i=1}^{p} \overline{u_i f_i} + \eta \int_{X_1\times\ldots\times X_p} ch \; dx_1 \ldots dx_p - \eta \sum_{i=1}^{p} \overline{u_i f_i}$$
$$\leq -(\eta - 1) \sum_{i=1}^{p} \overline{u_i f_i} + ||\eta ch||_{L^1(X_1\times\ldots\times X_p)}.$$

Thus, we have

$$\sum_{i=1}^{p} \overline{u_i f_i} \leq \frac{||\eta ch||_{L^1(X_1\times\ldots\times X_p)} - J'(u_1, \ldots, u_p)}{\eta - 1}. \tag{4.2.32}$$

By combining (4.2.31) and (4.2.32), we get the inequality (4.2.30) we desired. ■

Remark 4.2.17. *Note that, if we assume that* $\overline{(u_i + k_i) f_i} = \overline{(u_j + k_j) f_j}$, $\forall i \neq j$ *for some constants* $(k_1, \ldots, k_p)$, *the bounds in* (4.2.30) *imply that* $\overline{u_i f_i}$ *is also bounded* ,$\forall 1 \leq i \leq p$.

Next, we will obtain a bound on the oscillation of $u_i f_i$ around its mean for each i.

Lemma 4.2.18. *Let* $u_i, f_i \in L^1(X_i)$ *for each* $1 \leq i \leq p$, $c \in L^1(X_1 \times \ldots \times X_p)$ *and* $\tilde{h} \in L^\infty(X_1 \times \ldots \times X_p)$. *Suppose that there is some* $0 < \varepsilon \leq 1$ *such that* $\varepsilon f_1 \ldots f_p \leq \tilde{h}$ *for all* $(x_1, \ldots, x_p) \in X_1 \times \ldots \times X_p$. *Then, for each* $1 \leq i \leq p$ *we have*

$$\frac{\varepsilon}{6} ||u_i f_i - \overline{u_i f_i}||_{L^1(X_1\times\ldots\times X_p)} \leq -J'(u_1, \ldots, u_p) + ||c f_1 \ldots f_p||_{L^1(X_1\times\ldots\times X_p)} + \sum_{i=1}^{p} |\overline{u_i f_i}|. \tag{4.2.33}$$

Proof. Fix an $i \in \{1, \ldots, p\}$. Define the oscillation around the mean as

$$\sigma_i := \int_{X_i} |u_i f_i - \overline{u_i f_i}| \; dx_i.$$

Let

$$X_i^{\pm} = \left\{ x_i : \pm(u_i(x_i) f_i(x_i) - \overline{u_i f_i}) > \frac{\sigma_i}{3} \right\}$$

and

$$m_i^{\pm} = |X_i^{\pm}|$$

Also, let

$$A_i^{\pm} = \pm \int_{X_i^{\pm}} (u_i(x_i) f_i(x_i) - \overline{u_i f_i}) \; dx_i.$$

Note that by definition, $A_i^{\pm} \geq 0$.

Now, observe that

$$\begin{aligned}
\int_{\{|u_i f_i - \overline{u_i f_i}| \leq \frac{\sigma_i}{3}\}} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i &= \int_{X_i} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i \\
&\quad - \int_{\{u_i f_i - \overline{u_i f_i} < -\frac{\sigma_i}{3}\}} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i \\
&\quad - \int_{\{u_i f_i - \overline{u_i f_i} > \frac{\sigma_i}{3}\}} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i \\
&= \overline{u_i f_i}|X_i| - \overline{u_i f_i} - \int_{X_i^-} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i \\
&\quad - \int_{X_i^+} (\overline{u_i f_i} - u_i(x_i) f_i(x_i))\, dx_i \\
&= \int_{X_i^+} (u_i(x_i) f_i(x_i) - \overline{u_i f_i})\, dx_i \\
&\quad + \int_{X_i^-} (u_i(x_i) f_i(x_i) - \overline{u_i f_i})\, dx_i \\
&= A_i^+ - A_i^-.
\end{aligned} \tag{4.2.34}$$

In the second line, first term, we used the fact that the sets X_i have unit Lebesgue measure.

Then, by the definition of σ_i,

$$\begin{aligned}
\sigma_i &= \int_{X_i} |u_i f_i - \overline{u_i f_i}|\, dx_i \\
&= \int_{\{|u_i f_i - \overline{u_i f_i}| \leq \frac{\sigma_i}{3}\}} |u_i f_i - \overline{u_i f_i}|\, dx_i + \int_{\{u_i f_i - \overline{u_i f_i} < -\frac{\sigma_i}{3}\}} |u_i f_i - \overline{u_i f_i}|\, dx_i \\
&\quad + \int_{\{u_i f_i - \overline{u_i f_i} > \frac{\sigma_i}{3}\}} |u_i f_i - \overline{u_i f_i}|\, dx_i \\
&\leq \frac{\sigma_i}{3} \left|\left\{|u_i f_i - \overline{u_i f_i}| \leq \frac{\sigma_i}{3}\right\}\right| + \int_{X_i^-} |u_i f_i - \overline{u_i f_i}|\, dx_i + \int_{X_i^+} |u_i f_i - \overline{u_i f_i}|\, dx_i \\
&\leq \frac{\sigma_i}{3}\left(1 - m_i^+ - m_i^-\right) + A_i^+ + A_i^-.
\end{aligned}$$

Thus,

$$A_i^+ + A_i^- \geq \left(\frac{2}{3} + \frac{m_i^+}{3} + \frac{m_i^-}{3}\right)\sigma_i \geq \frac{2}{3}\sigma_i. \tag{4.2.35}$$

On the other hand, by (4.2.34) we have

$$\begin{aligned} A_i^+ - A_i^- &= \int_{\{|u_if_i - \overline{u_if_i}| \le \frac{\sigma_i}{3}\}} (\overline{u_if_i} - u_i(x_i)f_i(x_i))\ dx_i \\ &\ge -\frac{\sigma_i}{3} \left| \left\{ -\frac{\sigma_i}{3} \le u_if_i - \overline{u_if_i} \le \frac{\sigma_i}{3} \right\} \right| \\ &= -\frac{\sigma_i}{3}\left(1 - m_i^+ - m_i^-\right) \\ &\ge -\frac{\sigma_i}{3}. \end{aligned} \tag{4.2.36}$$

Thus by (4.2.35) and (4.2.36), we get

$$A_i^+ \ge \frac{\sigma_i}{6}. \tag{4.2.37}$$

Therefore, first we will find a bound on A_i^+, so that we get a bound on σ_i.

Now, observe that

$$\begin{aligned} A_i^+ &= \int_{X_i^+} (u_if_i - \overline{u_if_i})\ dx_i \\ &= -\int_{X_i^+} \overline{u_if_i}\ dx_i + \int_{X_i^+} u_if_i\ dx_i \\ &= -(\overline{u_if_i})m_i^+ + \int_{X_1\times\ldots\times X_{i-1}\times X_i^+\times X_{i+1}\times\ldots\times X_p} u_if_1\ldots f_p\ dx_1\ldots dx_p \\ &= -(\overline{u_if_i})m_i^+ + \int_{X_1\times\ldots\times X_{i-1}\times X_i^+\times X_{i+1}\times\ldots\times X_p} \left(-c + \sum_{k=1}^p u_k\right) f_1\ldots f_p\ dx_1\ldots dx_p \\ &\quad + \int_{X_1\times\ldots\times X_{i-1}\times X_i^+\times X_{i+1}\times\ldots\times X_p} cf_1\ldots f_p\ dx_1\ldots dx_p - \sum_{j\neq i}\overline{u_jf_j}\int_{X_i^+} f_i\ dx_i \\ &\le \frac{1}{\varepsilon}\int_{X_1\times\ldots\times X_p}\left[-c + \sum_{k=1}^p u_k\right]_+ \tilde{h}\ dx_1\ldots dx_p - \frac{1}{\varepsilon}\sum_{k=1}^p \overline{u_kf_k} + \frac{1}{\varepsilon}\sum_{k=1}^p\overline{u_kf_k} \\ &\quad + \int_{X_1\times\ldots\times X_p} |c|f_1\ldots f_p\ dx_1\ldots dx_p - \sum_{j\neq i}\overline{u_jf_j}\int_{X_i^+} f_i\ dx_i - (\overline{u_if_i})m_i^+ \\ &\le -\frac{1}{\varepsilon}J'(u_1,\ldots,u_p) + ||cf_1\ldots f_p||_{L^1(X_1\times\ldots\times X_p)} + \left(\frac{1}{\varepsilon} - m_i^+\right)\overline{u_if_i} \\ &\quad + \left(\frac{1}{\varepsilon} - \int_{X_i^+} f_i\ dx_i\right)\sum_{j\neq i}\overline{u_jf_j}. \end{aligned} \tag{4.2.38}$$

Note that in the penultimate line, we used the fact that $\varepsilon f_1\ldots f_p \le \tilde{h}$.

So, from (4.2.38), we have that,

$$\varepsilon A_i^+ \leq -J'(u_1,\ldots,u_p) + \varepsilon||cf_1\ldots f_p||_{L^1(X_1\times\ldots\times X_p)} + \left(1-\varepsilon m_i^+\right)\overline{u_i f_i}$$
$$+\left(1-\varepsilon\int_{X_i^+} f_i\ dx_i\right)\sum_{j\neq i}\overline{u_j f_j}.$$

Since each f_i is a probability density and $m_i^+,\varepsilon \leq 1$, we have $0 \leq 1-\varepsilon m_i^+ \leq 1$ and $0 \leq 1-\varepsilon\int_{X_i^+} f_i\ dx_i \leq 1$.

So, we get that

$$\varepsilon A_i^+ \leq -J'(u_1,\ldots,u_p) + ||cf_1\ldots f_p||_{L^1(X_1\times\ldots\times X_p)} + \sum_{k=1}^{p}\left[\overline{u_k f_k}\right]_+ \tag{4.2.39}$$

Since we have $A_i^+ \geq \frac{\sigma_i}{6}$ by (4.2.37), we get

$$\frac{\varepsilon}{6}\sigma_i \leq -J'(u_1,\ldots,u_p) + ||cf_1\ldots f_p||_{L^1(X_1\times\ldots\times X_p)} + \sum_{k=1}^{p}\left|\overline{u_k f_k}\right|.$$

This is the desired inequality (4.2.33). ■

Now, we will obtain L^1-bounds on the u_i's.

Proposition 4.2.19. *Let $c \in L^1(X_1\times\ldots\times X_p)$ and $\tilde{h} \in L^\infty(X_1\times\ldots\times X_p)$. Let h be a probability density with marginals $f_1,\ldots,f_p$ such that $h \leq \frac{\tilde{h}}{\eta}$ for some $\eta > 1$. Suppose that there is some $\varepsilon > 0$ such that $\varepsilon f_1(x_1)\ldots f_p(x_p) \leq \tilde{h}(x_1,\ldots,x_p)$ and $\varepsilon \leq \min\{f_1(x_1),\ldots,f_p(x_p)\}$ for almost all $(x_1,\ldots,x_p) \in X_1\times\ldots\times X_p$ and suppose that there exist some functions $u_i \in L^1(X_i), i \in \{1,\ldots,p\}$ such that $\overline{u_i f_i} = \overline{u_j f_j}, \forall i \neq j$. Then $\mathrm{OT}^*_{CCMM} - 1 \leq J'(u_1,\ldots,u_p)$ implies that $||u_i||_{L^1(X_i)}$ is bounded for each $1 \leq i \leq p$.*

Remark 4.2.20. *Note that the above bound depends only on $\mathrm{OT}^*_{CCMM}, \varepsilon, \eta, ||c||_{L^1}$ and $||\tilde{h}||_{L^\infty}$.*

Proof. Fix an $i \in \{1,\ldots,p\}$. Observe that

$$||u_i||_{L^1(X_i)} \leq ||\frac{1}{f_i}||_{L^\infty(X_i)}||u_i f_i||_{L^1(X_i)} \leq \frac{1}{\varepsilon}||u_i f_i||_{L^1(X_i)} \leq \frac{1}{\varepsilon}||\overline{u_i f_i}||_{L^1(X_i)} + \frac{1}{\varepsilon}||u_i f_i - \overline{u_i f_i}||_{L^1(X_i)}$$

Using Lemma 4.2.16, Lemma 4.2.18 and the fact that $\mathrm{OT}^*_{\mathrm{CCMM}} - 1 \leq J'(u_1,\ldots,u_p)$, we get a bound for $||u_i||_{L^1(X_i)}$ for each i. ■

Now, we will discuss the proof of the existence of maximizers.

By Proposition 4.2.19, we can pick a maximizing sequence $(u_1, \ldots, u_p)$ with uniform L^1-bounds for $\mathrm{OT}^*_{\mathrm{CCMM}}$. However, since the L^1-unit ball is not compact, an L^1-bound is not sufficient to get the convergence of a subsequence. Therefore, in order to get better compactness properties, we extend the definition of $J'(u_1, \ldots, u_p)$ from L^1 space to the space of signed measures with finite total variation.

Let $M(X)$ denote the space of signed Radon measures on X. Let $C \in M(X_1 \times \ldots \times X_p)$ be such that $dC(x_1, \ldots, x_p) = c(x_1, \ldots, x_p)\ dx_1 \ldots dx_p$ and $U_i \in M(X_i)$ for each $1 \leq i \leq p$.

Define

$$\tilde{J}'(U_1, \ldots, U_p) := \sum_{i=1}^{p} \int_{X_i} f_i\ dU_i - \int_{X_1 \times \ldots \times X_p} \tilde{h}\ d\left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i \otimes \ldots \otimes H_p^d\right]_+ . \tag{4.2.40}$$

Here, $dH_i^d(x) = dx_i$ denotes Lebesgue measure on $\mathbb{R}^d$.

First, we will prove that the functional $\tilde{J}'$ is upper semi-continuous with respect to weak-* convergence in $M(X_1) \times \ldots \times M(X_p)$.

Lemma 4.2.21. *Let $c \in L^1(X_1 \times \ldots \times X_p)$ be such that $dC = c\ dH^{pd}$. Let $f_1, \ldots, f_p$ and $\tilde{h}$ be continuous, non-negative functions on the compact sets $X_1, \ldots, X_p$ and $X_1 \times \ldots \times X_p$ respectively. Then, the functional $\tilde{J}'$ is upper semi-continuous with respect to weak-* convergence in $M(X_1) \times \ldots \times M(X_p)$.*

Proof. Let $(U_1^n, \ldots, U_p^n)_{n\in\mathbb{N}}^{\infty}$ be a bounded sequence in $M(X_1) \times \ldots \times M(X_p)$ such that $(U_1^n, \ldots, U_p^n)$ converges to $(U_1, \ldots, U_p)$ when tested against functions in $C(X_1) \times \ldots \times C(X_p)$. We need to show that

$$\limsup_{n\to\infty} \tilde{J}'(U_1^n, \ldots, U_p^n) \leq \tilde{J}'(U_1, \ldots, U_p).$$

Observe that,

$$\limsup_{n\to\infty} \left\{ \sum_{i=1}^{p} \int_{X_i} f_i\ dU_i^n - \int_{X_1 \times \ldots \times X_p} \tilde{h}\ d\left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^n \otimes \ldots \otimes H_p^d\right]_+ \right\}$$

$$\leq \limsup_{n\to\infty} \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^n + \limsup_{n\to\infty} - \int_{X_1\times\ldots\times X_p} \tilde{h} \; d \left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^n \otimes \ldots \otimes H_p^d \right]_+$$
$$= \limsup_{n\to\infty} \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^n - \liminf_{n\to\infty} \int_{X_1\times\ldots\times X_p} \tilde{h} \; d \left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^n \otimes \ldots \otimes H_p^d \right]_+ .$$

We already have that, for each $1 \leq i \leq p$,

$$\int_{X_i} f_i \; dU_i = \lim_{n\to\infty} \int_{X_i} f_i \; dU_i^n. \tag{4.2.41}$$

So, it remains to show that

$$\int_{X_1\times\ldots\times X_p} \tilde{h} \; d \left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i \otimes \ldots \otimes H_p^d \right]_+ \leq \liminf_{n\to\infty} \int_{X_1\times\ldots\times X_p} \tilde{h} \; d \left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^n \otimes \ldots \otimes H_p^d \right]_+ . \tag{4.2.42}$$

Fix an $i \in \{1, \ldots, p\}$. Let

$$d\mu_i^n := \tilde{h} \; d \left(-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^n \otimes \ldots \otimes H_p^d \right)$$

and

$$d\mu_i := \tilde{h} \; d \left(-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i \otimes \ldots \otimes H_p^d \right).$$

Then, we have that $d\mu_i^n$ converges weakly-* to $d\mu_i$. i.e. given a function $\phi \in C(X_1 \times \ldots \times X_p)$, we have

$$\begin{aligned} \int_{X_1\times\ldots\times X_p} \phi \; d\mu_i^n = & - \int_{X_1\times\ldots\times X_p} c\tilde{h}\phi \; dx_1 \ldots dx_p \\ & + \sum_{i=1}^{p} \int_{X_i} \left(\int_{X_1\times\ldots\times X_{i-1}\times X_{i+1}\times\ldots\times X_p} \tilde{h}\phi dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p \right) \; dU_i^n \\ \xrightarrow{n\to\infty} & - \int_{X_1\times\ldots\times X_p} c\tilde{h}\phi \; dx_1 \ldots dx_p \\ & + \sum_{i=1}^{p} \int_{X_i} \left(\int_{X_1\times\ldots\times X_{i-1}\times X_{i+1}\times\ldots\times X_p} \tilde{h}\phi dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_p \right) \; dU_i \end{aligned}$$

$$= \int_{X_1 \times \ldots \times X_p} \phi \, d\mu_i. \tag{4.2.43}$$

Now, we will decompose the measures μ_i^n and μ_i into their positive and negative parts as $\mu_i^n = \mu_i^n{}_+ - \mu_i^n{}_-$ and $\mu_i = \mu_{i+} - \mu_{i+}$. Since μ_i^n converges to μ_i weak-*, by plugging $\phi \equiv 1$ in (4.2.43), we get

$$\begin{aligned} &\mu_i^n{}_+(X_1 \times \ldots \times X_p) - \mu_i^n{}_-(X_1 \times \ldots \times X_p) = \mu_i^n(X_1 \times \ldots \times X_p) \\ &\qquad \xrightarrow{n\to\infty} \mu_i(X_1 \times \ldots \times X_p) = \mu_{i+}(X_1 \times \ldots \times X_p) - \mu_{i-}(X_1 \times \ldots \times X_p). \end{aligned} \tag{4.2.44}$$

Let $\tilde{X} = X_1 \times \ldots \times X_p$.

So, we have

$$\mu_{i+}(\tilde{X}) - \mu_{i+}(\tilde{X}) \leq \liminf_{n\to\infty} \{\mu_i^n{}_+(\tilde{X}) - \mu_i^n{}_-(\tilde{X})\}. \tag{4.2.45}$$

On the other hand, since the total variation norm $||\mu_i||_{\text{TV}}$ given by $|\mu_i| = \mu_{i+} + \mu_{i-}$ is lower semi-continuous w.r.t. weak-* convergence, we have

$$\begin{aligned} \mu_{i+}(\tilde{X}) + \mu_{i-}(\tilde{X}) &\leq \liminf_{n\to\infty} ||\mu_i^n||_{\text{TV}} \\ &= \liminf_{n\to\infty} \{\mu_i^n{}_+(\tilde{X}) + \mu_i^n{}_-(\tilde{X})\}. \end{aligned} \tag{4.2.46}$$

i.e.

$$\mu_{i+}(\tilde{X}) + \mu_{i-}(\tilde{X}) \leq \liminf_{n\to\infty} \{\mu_i^n{}_+(\tilde{X}) + \mu_i^n{}_-(\tilde{X})\}. \tag{4.2.47}$$

Now, by combining (4.2.45) and (4.2.47), we get

$$\mu_{i+}(\tilde{X}) \leq \liminf_{n\to\infty} \mu_i^n{}_+(\tilde{X}).$$

Note that, since we only have the positive part of the measures in (4.2.42), it is enough to get lower semi-continuity of $\mu_i^n{}_+(\tilde{X})$ only.

This completes the proof of Lemma 4.2.21. ■

Finally, we will present the main result for existence of dual maximizers.

Theorem 4.2.22. *Let f_i and $\tilde{h}$ be continuous and strictly positive on their compact supports $X_i \subseteq \mathbb{R}^d$ and $X_1 \times \ldots \times X_p$ respectively. Let $c \in L^1(X_1 \times \ldots \times X_p)$. Fix an $\eta > 1$ and* *assume that* $\Pi^{\tilde{h}/\eta}(f_1,\dots,f_p)$ *is non-empty. Then, there exist functions* $(u_1,\dots,u_p)$ *where* $u_i \in L^1(X_i)\ \forall i$*, such that*

$$\mathrm{OT}^{*'}_{CCMM} = J'(u_1,\dots,u_p).$$

Proof. Let $(u_1^n,\dots,u_p^n)_{n\in\mathbb{N}}$ be a maximizing sequence for $\mathrm{OT}^*_{\mathrm{CCMM}}$,

i.e. for each $i \in \{1,\dots,p\}$, $u_i^n \in L^1(X_i)$ and

$$\mathrm{OT}^{*'}_{\mathrm{CCMM}} = \lim_{n\to\infty} J'(u_1^n,\dots,u_p^n).$$

Fix $\varepsilon \le \min_{x_1,\dots,x_p\in\tilde{X}}\{f_1(x_1),\dots,f_p(x_p)\}$. Since $\tilde{h}$ is bounded away from zero, we may pick ε sufficiently small so that we have $\varepsilon f_1(x_1)\dots f_p(x_p) \le \tilde{h}$ for all $(x_1,\dots,x_p) \in X_1\times\dots\times X_p$. We can find a p-tuple of constants $(k_1^n,\dots,k_p^n)$ with $\sum_{i=1}^p k_i^n = 0$ such that adding k_i^n to each u_i^n ensures that $\overline{u_i^n f_i} = \overline{u_j^n f_j}$ for each $i \neq j$.

Now, by Proposition 4.2.19, for each i, we get a bound for $||u_i^n||_{L^1(X_i)}$ independent from n. For each i, let $U_i^n \in \mathcal{M}(X_i)$ be such that $dU_i^n(x_i) = du_i^n(x_i)\ dx_i$. Then, we have that $||U_i^n||_{\mathrm{TV}} = ||u_i^n||_{L^1(X_i)}$ is bounded. Thus, by Alaoglu's theorem, we can get a subsequence (without relabelling) $(U_1^n,\dots,U_p^n)$ that converges weakly-* to some $(U_1,\dots,U_p) \in M(X_1)\times\dots\times M(X_p)$.

So, we have that $\tilde{J}'(U_1^n,\dots,U_p^n) = J'(u_1^n,\dots,u_p^n) \xrightarrow{n\to\infty} \mathrm{OT}^*_{\mathrm{CCMM}}$. By Lemma 4.2.21, we have that

$$\mathrm{OT}^*_{\mathrm{CCMM}} \le \tilde{J}'(U_1,\dots,U_p).$$

Next, we will prove that $\mathrm{OT}^*_{\mathrm{CCMM}} \ge \tilde{J}'(U_1,\dots,U_p)$.

Let $C = cH_1^d\otimes\dots\otimes H_p^d$. By Lebesgue's decomposition theorem, for each i, the measures U_i can be written as $U_i = U_i^{\mathrm{ac}} + U_i^{\mathrm{s}}$, where $U_i^{\mathrm{ac}} \ll H^d$ and $U_i^{\mathrm{s}} \perp H^d$. On the other hand, by the Hahn-Jordan decomposition theorem, we can write $U_i = U_{i+} - U_{i-}$. Since the Hahn-Jordan decomposition and the Lebesgue decomposition commute, we have

$$U_{i+} = [U_i^{\mathrm{ac}}]_+ + [U_i^{\mathrm{s}}]_+\,,$$

and

$$U_{i-} = [U_i^{\mathrm{ac}}]_- + [U_i^{\mathrm{s}}]_-\,.$$

Now, observe that

$$\left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i \otimes \ldots \otimes H_p^d\right]_+$$
$$= \left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{ac}} \otimes \ldots \otimes H_p^d\right]_+ + \left[\sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{s}} \otimes \ldots \otimes H_p^d\right]_+ .$$

Thus,

$$\begin{aligned}
\tilde{J}'(U_1, \ldots, U_p) &= \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i - \int_{X_1 \times \ldots \times X_p} \tilde{h} \; d\left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i \otimes \ldots \otimes H_p^d\right]_+ \\
&= \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^{\mathrm{ac}} + \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^{\mathrm{s}} \\
&\quad - \int_{X_1 \times \ldots \times X_p} \tilde{h} \; d\left[-C + \sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{ac}} \otimes \ldots \otimes H_p^d\right]_+ \\
&\quad - \int_{X_1 \times \ldots \times X_p} \tilde{h} \; d\left[\sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{s}} \otimes \ldots \otimes H_p^d\right]_+ \\
&= \tilde{J}'(U_1^{\mathrm{ac}}, \ldots, U_p^{\mathrm{ac}}) + \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^{\mathrm{s}} \\
&\quad - \int_{X_1 \times \ldots \times X_p} \tilde{h} \; d\left[\sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{s}} \otimes \ldots \otimes H_p^d\right]_+ \\
&\leq J'(u_1, \ldots, u_p) + \sum_{i=1}^{p} \int_{X_i} f_i \; dU_i^{\mathrm{s}} \\
&\quad - \int_{X_1 \times \ldots \times X_p} h \; d\left[\sum_{i=1}^{p} H_1^d \otimes \ldots \otimes U_i^{\mathrm{s}} \otimes \ldots \otimes H_p^d\right] \\
&= J'(u_1, \ldots, u_p) \\
&\leq \mathrm{OT}^{*'}_{\mathrm{CCMM}} .
\end{aligned}$$

Here, for each i, u_i represents the Radon-Nikodym derivative of U_i^{ac} such that $u_i \; dx_i = dU_i^{\mathrm{ac}}$ and $h \in \Pi^{\tilde{h}/\eta}(f_1, \ldots, f_p)$.

So, we have proven that

$$\mathrm{OT}^{*'}_{\mathrm{CCMM}} \leq \tilde{J}'(U_1, \ldots, U_p) \leq J'(u_1, \ldots, u_p) \leq \mathrm{OT}^{*'}_{\mathrm{CCMM}} .$$

Thus, $(u_1, \ldots, u_p)$ is the desired maximizer. ■

CHAPTER 5

BARYCENTERS

5.1 Barycenters in the Wasserstein Space

5.1.1 Introduction

The problem of finding a barycenter in the Wasserstine space is a nonlinear interpolation between several probability measures. As an example in the discrete case, consider several coal mines, and the coal extracted has to be sent to a factory that is located centrally. The problem is to find the location to construct such a factory so that total transportation cost is minimized.

This notion of barycenters in Wasserstein space was first introduced by Agueh and Carlier in [1] and they have provided existence, uniqueness, characterizations of the minimizer and regularity of the barycenter, and relate it to the multi-marginal OT problem considered by Gangbo and Święch in [24]. This problem has a wide range of applications including economics [13] and data science [4, 8].

To elaborate a few applications of Wasserstein Barycenters; in image processing, Wasserstein barycenters have been used to generate "averaged" images from a set of input images. This technique is particularly useful for denoising and image reconstruction. "Image morphing" or "image interpolation" is such an application ([52, 56]). In Machine Learning, it can be used to generate representative data points from a set of input data, which can then be used to train machine learning models. This can be particularly useful in cases where the input data is noisy or incomplete.

Overall, Wasserstein barycenters are a powerful mathematical tool that have found many applications in different fields. Their properties and computational efficiency make them a popular choice for solving optimization problems involving probability measures.

To provide some background on the Wasserstein Barycenters, we will present some existing results in [1] in the next few sections.

5.1.2 The Primal Problem

Recall that we define the squared 2-Wasserstein distance between two probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ by

$$W_2^2(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|x-y|^2 d\gamma(x,y).$$

Let $p \geq 2$ be an integer. Given a p-tuple of probability measures $(\nu_1, \dots \nu_p)$ with each $\nu_i \in \mathcal{P}_2(\mathbb{R}^d)$ and a p-tuple of positive real numbers $(\lambda_1, \dots \lambda_p)$ with $\sum_{i=1}^p \lambda_i = 1$, we define the following minimization problem:

$$\mathrm{OT}_{BC} = \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \sum_{i=1}^p \lambda_i W_2^2(\nu_i, \nu) \right\}. \tag{5.1.1}$$

A solution of (5.1.1) is called the barycenter of the probabilities ν_i with weights λ_i.

Remark 5.1.1. *For $p = 2$ with $\lambda_1 = \lambda_2 = \frac{1}{2}$, this problem means finding the midpoint between the two measures ν_1 and ν_2 and such an interpolation is already known as the McCann's interpolation [40].*

Theorem 5.1.2. *([1], Proposition 2.3) Given an integer $p \geq 2$, a p-tuple of probability measures $(\nu_1, \dots \nu_p)$ with each $\nu_i \in \mathcal{P}_2(\mathbb{R}^d)$ and a p-tuple of positive real numbers $(\lambda_1, \dots \lambda_p)$ with $\sum_{i=1}^p \lambda_i = 1$, the Barycenter Problem given by (5.1.1) has a solution.*

The proof is given in [1] and it uses the direct method in Calculus of Variations.

To study the uniqueness and other properties of the barycenters, we require a dual formulation.

5.1.3 Duality

Define the space of continuous functions with at most quadratic growth,

$$Y := (1 + |\ .\ |^2) C_b(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) : \frac{f}{1 + |\ .|^2} \text{ is bounded} \right\}$$

that is equipped with the norm

$$||f||_Y := \sup_{x \in \mathbb{R}^d} \frac{|f(x)|}{1 + |x|^2}. \tag{5.1.2}$$

We define the dual of (5.1.1) as

$$\mathrm{OT}^*_{BC} := \sup_{f_i \in C_b(Y), \sum_{i=1}^p f_i = 0} \sum_{i=1}^p \int_{\mathbb{R}^d} \inf_{y \in \mathbb{R}^d} \left\{ \frac{\lambda_i}{2}|x-y|^2 - f_i(y) \right\} d\nu_i(x). \tag{5.1.3}$$

In [1], the authors have proven that the strong duality holds for the barycenter problem and dual maximizers for (5.1.3) exist.

Theorem 5.1.3. *([1], Proposition 2.2 & Proposition 2.3)*

$$\mathrm{OT}_{BC} = \mathrm{OT}^*_{BC},$$

and OT^*_{BC} *given by* (5.1.3) *has a solution.*

With the duality results, we can characterize the barycenters in several ways. Given below are few results from [1].

Proposition 5.1.4. *([1], Proposition 3.5) Assume that there is an index $i \in \{1, \dots, p\}$ such that ν_i vanishes on small sets. Then OT_{BC} admits a unique solution ν which is given by $\nu = \nabla\phi_{i\#}\nu_i$, where given a solution $(f_1, \dots, f_p)$ of OT^*_{BC}, ϕ_i is the convex potential defined by*

$$\lambda_i \phi_i(x) := \frac{\lambda_i}{2}|x|^2 - \inf_{y \in \mathbb{R}^d} \left\{ \frac{\lambda_i}{2}|x-y|^2 - f_i(y) \right\}. \tag{5.1.4}$$

Proposition 5.1.5. *([1], Proposition 3.8) Assume that ν_i vanishes on small sets for every $i \in \{1, \dots, p\}$, and let $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. Then the following conditions are equivalent:*

1. *ν solves OT_{BC}.*

2. *$\nu = \nabla\phi_{i\#}\nu_i$ for every i, where ϕ_i is defined by* (5.1.4).

3. *There exist convex potentials ψ_i such that $\nabla\psi_i$ is the Brenier's map transporting ν_i to ν, and a constant C such that*

$$\sum_{i=1}^p \lambda_i \psi_i^*(y) \le C + \frac{|y|^2}{2}, \forall y \in \mathbb{R}^d, \text{with equality } \nu\text{-a.e.} \tag{5.1.5}$$

Remark 5.1.6. *If ν and the potentials ψ_i satisfy the third statement of Proposition 5.1.5, then the support of ν is included in the contact set where the convex function $\phi := \sum_{i=1}^p \lambda_i \psi_i^*$ agrees with its quadratic majorant $C + \frac{|.|^2}{2}$. Note that, at such a point, we have*

$$\sum_{i=1}^{p} \lambda_i \partial\psi_i^*(x) \subset \partial\phi(x) \subset \{x\},$$

so that each potential ψ_i^* is differentiable at x. The potentials ψ_i^* are therefore differentiable on the support of ν and satisfy the relation

$$\sum_{i=1}^{p} \lambda_i \nabla\psi_i^* = id. \tag{5.1.6}$$

Remark 5.1.7. *If* (5.1.6) *holds everywhere for the Brenier's maps $\nabla\psi_i$ transporting ν_i to ν, then ν is optimal for* OT_{BC}.

In [1], the authors also have shown that OT_{BC} is equivalent to a linear programming problem of multi-marginal optimal transport type similar to the problem solved by Gangbo and Święch in [24]. Now we will list two main Theorems which describe this relation.

Theorem 5.1.8. *([1], Theorem 4.1) Assume that ν_i vanishes on small sets for $i = 1, \ldots, p$. Then the multi-marginal problem given by*

$$\sup\left\{ \int_{(\mathbb{R}^d)^p} \Big(\sum_{1\leq i\leq j\leq p} \lambda_i\lambda_j x_i \cdot x_j \Big)\, d\gamma(x_1, \ldots, x_p),\ \gamma \in \Pi(\nu_1, \ldots, \nu_p) \right\}. \tag{5.1.7}$$

admits a unique solution $\gamma \in \Pi(\nu_1, \ldots, \nu_p)$. Moreover, γ is of the form $\gamma = (T_1^1, \ldots, T_p^1)_{\#}\nu_1$ with $T_i^1 = \nabla u_i^* \circ \nabla u_1$ for $i = 1, \ldots, p$ where u_i are the strictly convex potentials defined by

$$u_i(x) := \frac{\lambda_i}{2}|x|^2 + \frac{g_i(x)}{\lambda_i},\ \forall x \in \mathbb{R}^d, \tag{5.1.8}$$

and $(g_1, \ldots, g_p)$ are the convex potentials that solve the dual of (5.1.7) *given by*

$$\inf\left\{ \sum_{i=1}^{p} \int_{\mathbb{R}^d} g_i\, d\nu_i, \sum_{i=1}^{p} g_i(x_i) \geq \sum_{1\leq i\leq j\leq p} \lambda_i\lambda_j x_i \cdot x_j,\ \forall x \in (\mathbb{R}^d)^p \right\}. \tag{5.1.9}$$

Proposition 5.1.9. *([1], Proposition 4.2) Assume that ν_i vanishes on small set for $i = 1, \dots, p$. Then the solution of* OT_{BC} *given by* (5.1.1) *is given by $\nu = T_{\#}\gamma$, where γ is the solution of* (5.1.7) *and T is defined by*

$$T(x) := \sum_{i=1}^{p} \lambda_i x_i, \quad \forall x := (x_1, \dots, x_p) \in (\mathbb{R}^d)^p. \tag{5.1.10}$$

Remark 5.1.10. *Note that the Euclidean barycenter $T(x)$ defined in* (5.1.10) *is characterized by the property*

$$\sum_{i=1}^{p} \lambda_i |x_i - T(x)|^2 = \inf_{y \in \mathbb{R}^d} \left\{ \lambda_i |x_i - y|^2 \right\}. \tag{5.1.11}$$

5.2 Capacity Constrained Barycenter Problem

In this section, we introduce the notion of capacity constrained barycenters in Wasserstein space which is a generalization of the barycenter problem (5.1.1). As the name suggests, the capacity constrained barycenter problem introduces capacities to each of the two marginal problems associated. Under certain assumptions on the capacities, we have proven that the problem attains a minimizer and obtained duality results.

5.2.1 The Primal Problem

Given an integer $p \geq 2$, a p-tuple of probability measures $(\nu_1,...,\nu_p)$ each in $\mathcal{P}_2(\mathbb{R}^d)$ and a p-tuple of positive real numbers $(\lambda_1,...,\lambda_p)$ with $\sum_{i=1}^{p} \lambda_i = 1$, we define the following minimizing problem:

$$\mathrm{OT}_{\mathrm{CCBC}} := \inf_{\nu \in \mathcal{P}'} \left\{ J(\nu) := \sum_{i=1}^{p} \lambda_i \widetilde{W}_2^2(\nu_i, \nu) \right\} \tag{5.2.1}$$

where

$$\mathcal{P}' = \mathcal{P}_2(\mathbb{R}^d) \cap \left\{ \nu' : \Pi^{\tilde{\gamma}_i}(\nu_i, \nu') \neq \emptyset, \ \forall i \in \{1, 2, \dots p\} \right\}, \tag{5.2.2}$$

and

$$\widetilde{W}_2^2(\nu_i, \nu) = \inf_{\gamma_i \in \Pi^{\tilde{\gamma}_i}(\nu_i, \nu)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} |x - y|^2 \, d\gamma_i(x, y) \right\}. \tag{5.2.3}$$

Here, $\{\tilde{\gamma}_i\}_{i=1}^{p} \subseteq \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of *capacities* of the two marginal problems and $\Pi^{\tilde{\gamma}_i}(\nu_i, \nu)$ is the set of transport plans from ν_i to ν bounded by $\tilde{\gamma}_i$ (see Definition 3.4.2).

We call the problem (5.2.1), the Capacity Constrained Barycenter (CCBC) problem between the measures $\nu_1, \dots, \nu_p$.

We will consider this minimization problem under two assumptions.

- Assumption 1: We assume that $\{\tilde{\gamma}_i\}_{i=1}^p$ are compactly supported finite measures that are absolutely continuous w.r.t. Lebesgue measure with bounded densities for all $i \in \{1, 2, \dots p\}$.
- Assumption 2: We assume that the set $\mathcal{P}'$ is non-empty.

Recall that the non-empty condition of the set $\Pi^{\tilde{\gamma}_i}(\nu_i, \nu)$ is given by the following theorem.

Theorem 5.2.1. *([46], Corollary 4.6.15) Let X, Y be compact sets and for given Borel probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a finite Borel measure $\tilde{\gamma}$ on $X \times Y$, we have that $\Pi^{\tilde{\gamma}}(\mu, \nu) \neq \emptyset$ if and only if*

$$\mu(A) + \nu(B) \leq \tilde{\gamma}(A \times B) + 1, \quad \forall A \in \mathcal{B}(X) \text{ and } \forall B \in \mathcal{B}(Y).$$

Remark 5.2.2. *Note that the assumption 2 is not a very restrictive assumption. For instance, we could pick the capacities $\tilde{\gamma}_i$ such that $\tilde{\gamma}_i = \nu_i \otimes \xi$ for some probability measure $\xi \in \mathcal{P}_2(\mathbb{R}^d)$ so that the set $\mathcal{P}'$ becomes non-empty.*

Lemma 5.2.3. *$\widetilde{W}_2^2$ is weakly lower semi-continuous.*

Proof. Let $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\tilde{\nu}_n\}_{n\in\mathbb{N}}$ be two sequences of probability measures in $\mathcal{P}'$ such that $\mu_n \rightharpoonup \mu^*$ and $\tilde{\nu}_n \rightharpoonup \nu^*$.
Since $\{\mu_n\}$ and $\{\tilde{\nu}_n\}$ are tight $\bigcup \Pi(\mu_n, \tilde{\nu}_n)$ is also tight (see [53], Lemma 4.4).
Now, let $\gamma_n^* \in \Pi(\mu_n, \tilde{\nu}_n)$ be optimal such that $\gamma_n^* \leq \tilde{\gamma}$. Since $\{\gamma_n^*\}$ is also tight, there exists a $\gamma^* \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\gamma_n^* \rightharpoonup \gamma^*$ and $\gamma^* \in \Pi(\mu^*, \nu^*)$. We also get that $\gamma^* \leq \tilde{\gamma}$ (see the proof of (5.2.8) in Theorem 5.2.4 below).

Now,

$$\widetilde{W}_2^2(\mu^*, \nu^*) = \inf_{\gamma \in \Pi^{\tilde{\gamma}}(\mu^*, \nu^*)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|x - y|^2 \, d\gamma(x, y) \right\}$$

$$
\begin{aligned}
&\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|x-y|^2 \, d\gamma^*(x,y) \\
&\leq \liminf_n \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|x-y|^2 \, d\gamma_n^*(x,y) \qquad \text{(see [53], Lemma 4.3)} \\
&= \liminf_n \inf_{\gamma_n\in\Pi^{\tilde{\gamma}}(\mu_n,\nu_n)} \left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}|x-y|^2 \, d\gamma_n(x,y) \right\} \\
&= \liminf_n \widetilde{W}_2^2(\mu_n,\tilde{\nu}_n).
\end{aligned}
$$

This proves that $\widetilde{W}_2^2$ is weakly lower semi-continuous. ■

Now, we will show that the CCBC problem has a solution.

Theorem 5.2.4. *Under the assumptions 1 and 2, the CCBC problem given by* (5.2.1) *has a solution.*

Proof. Let $\{\tilde{\nu}_n\}_{n\in\mathbb{N}} \subseteq \mathcal{P}'$ be a minimizing sequence of $\mathrm{OT}_{\mathrm{CCBC}}$.
i.e. $\lim_{n\to\infty} J(\tilde{\nu}_n) = \inf_\nu J(\nu)$. Then, we can find an $M > 0$ such that $J(\tilde{\nu}_n) \leq M$ for all n. Thus, for each $1 \leq i \leq p$ and for each $n \in \mathbb{N}$, $\lambda_i \widetilde{W_2}^2(\nu_i, \tilde{\nu}_n) \leq M$.

Using the duality and the assumption that ν_i's have finite second moments, we can show that (see Appendix 3)

$$
\sup_n \int |x|^2 \, d\tilde{\nu}_n \leq C, \qquad \text{for some constant C.}
$$

Hence, $\{\tilde{\nu}_n\}$ is tight (see Appendix 4).

Then, by Prokhorov's theorem, there exists a subsequence $\{\tilde{\nu}_n\}$ (not relabeled), that converges weakly to some $\nu^* \in \mathcal{P}(Y)$.

Since, $\int |x|^2 \, d\nu^* \leq \liminf_n \int |x|^2 \, d\tilde{\nu}_n \leq C$ (see [53], Lemma 4.3), we have that $\nu^* \in \mathcal{P}_2(Y)$.

Now, we will prove that $\forall 1 \leq i \leq p$, there exists a $\gamma_i \in \Pi^{\tilde{\gamma}_i}(\tilde{\nu}_n, \nu^*)$, so that $\nu^* \in \mathcal{P}'$.

Fix an $i \in \{1, \dots, p\}$ and $n \in \mathbb{N}$. By assumption 2, there exists some $\gamma_{i,n} \in \Pi(\nu_i, \tilde{\nu}_n)$ such that $\gamma_{i,n} \leq \tilde{\gamma}_i$. Since $\{\gamma_{i,n}\}$ is tight, there exists a subsequence $\{\gamma_{i,n}\}$ (not relabeled), that weakly converges to some $\gamma_i^* \in \mathcal{P}(X \times Y)$.

Let $E \subset \mathbb{R}^d$ be an open set and $\{\phi_l\}_{l=1}^{\infty} \subset C_b(\mathbb{R}^d)$ be a sequence of functions such that $0 \leq \phi_l \leq \mathbb{1}_E$ and $\phi_l \nearrow \mathbb{1}_E$ pointwise.

Then,

$$\int_{\mathbb{R}^d} \phi_l \; d\tilde{\nu}_n = \int_{\mathbb{R}^d} \phi_l \; d(\mathrm{Proj}_y(x,y)_{\#}\gamma_{i,n}) = \int_{\mathbb{R}^d} \phi_l \circ \mathrm{Proj}_y(x,y) \; d\gamma_{i,n}.$$

Letting $n \to \infty$, we get

$$\int_{\mathbb{R}^d} \phi_l \; d\nu^* = \int_{\mathbb{R}^d} \phi_l \circ \mathrm{Proj}_y(x,y) \; d\gamma_i^*. \tag{5.2.4}$$

In (5.2.4), on the left hand side, we used the fact that $\tilde{\nu}_n \rightharpoonup \nu^*$ and on the right hand side, we used the fact that $\gamma_{i,n} \rightharpoonup \gamma_i^*$. Now, letting $l \to \infty$ in 5.2.4, by monotone convergence theorem, we get

$$\int_{\mathbb{R}^d} \mathbb{1}_E \; d\nu^* = \int_{\mathbb{R}^d} \mathbb{1}_E \circ \mathrm{Proj}_y(x,y) \; d\gamma_i^*.$$

$$\text{i.e.} \qquad \nu^*(E) = \mathrm{Proj}_y(x,y)_{\#}\gamma_i^*(E). \tag{5.2.5}$$

Now, let $B \subset \mathbb{R}^d$ be any Borel set. Since Borel measures are outer regular, we have that

$$\nu^*(B) = \mathrm{Proj}_y(x,y)_{\#}\gamma_i^*(B). \tag{5.2.6}$$

Similarly, we can prove that, for any Borel set $A \in \mathbb{R}^d$,

$$\nu_i(A) = \mathrm{Proj}_x(x,y)_{\#}\gamma_i^*(A). \tag{5.2.7}$$

By (5.2.6) and (5.2.7), we can conclude that $\gamma_i^* \in \Pi(\nu_i, \nu^*)$.

Now, let $E, F \subset \mathbb{R}^d$ be two open sets and $\{\phi_l\}_{l=1}^{\infty} \subset C_b(\mathbb{R}^d \times \mathbb{R}^d)$ be a sequence of functions such that $0 \leq \phi_l \leq \mathbb{1}_{E\times F}$ and $\phi_l \nearrow \mathbb{1}_{E\times F}$ pointwise.

Since $\gamma_{i,n} \leq \tilde{\gamma}_i$, we have

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_l \; d\gamma_{i,n} \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_l \; d\tilde{\gamma}_i.$$

Now, letting $n \to \infty$, gives us

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_l \; d\gamma_i^* \leq \int_{\mathbb{R}^d\times\mathbb{R}^d} \phi_l \; d\tilde{\gamma}_i.$$

Letting $l \to \infty$, the monotone convergence theorem gives us

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{E\times F}\ d\gamma_i^* \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{E\times F}\ d\tilde{\gamma}_i,$$

$$\text{i.e.} \qquad \gamma_i^*(E \times F) \leq \tilde{\gamma}_i(E \times F).$$

Hence, for given two Borel sets $A, B \subset \mathbb{R}^d$, we have that

$$\gamma_i^*(A \times B) \leq \tilde{\gamma}_i(A \times B). \tag{5.2.8}$$

Thus, $\gamma_i^* \in \Pi^{\tilde{\gamma}_i}(\nu_i, \nu^*)$, i.e. $\nu^* \in \mathcal{P}'$.

Now, we will claim that ν^* is the desired minimizer.

Observe that,

$$\begin{aligned}
J(\nu^*) &= \sum_{i=1}^{p} \lambda_i \widetilde{W}_2^2(\nu_i, \nu^*) \\
&\leq \sum_{i=1}^{p} \lambda_i \liminf_n \widetilde{W}_2^2(\nu_i, \tilde{\nu}_n) \qquad \text{(By Lemma 5.2.3)} \\
&\leq \liminf_n \sum_{i=1}^{p} \lambda_i \widetilde{W}_2^2(\nu_i, \tilde{\nu}_n) \\
&= \liminf_n J(\tilde{\nu}_n) \\
&= \mathrm{OT}_{\mathrm{CCBC}}\,.
\end{aligned}$$

This concludes that ν^* is a minimizer for CCBC Problem. ■

5.2.2 Duality

For the dual formulation, we will be adopting the duality results for two-marginal capacity constrained OT problem by Rachev and Rüschendorf in [46].

Let $K \times K$ be a compact subset of $\mathbb{R}^d \times \mathbb{R}^d$ containing supports of the capacities $\tilde{\gamma}_i$ for all i. For an integer $p \geq 2$ consider the dual problem:

$$\begin{aligned}
\mathrm{OT}^*_{\mathrm{CCBC}} = \sup_{\mathcal{A}} \Bigg\{ &\sum_{i=1}^{p} \int_K \inf_{z\in K} \left\{ \frac{\lambda_i}{2}|x-z|^2 - f_i(z) - w_i(x,z) \right\} d\nu_i(x) \\
&+ \sum_{i=1}^{p} \int_{K\times K} w_i(x,y)\ d\tilde{\gamma}_i(x,y) \Bigg\}.
\end{aligned} \tag{5.2.9}$$

Here, the supremum is taken over the set $\mathcal{A}$ of functions f_i and w_i satisfying $f_i \in C_b(K), w_i \in C_b(K \times K)$, $w_i \leq 0$, $\forall i \in \{1, 2, \dots p\}$ and $\displaystyle\sum_{i=1}^{p} f_i = 0$.

In the Theorem below, we will show that the strong duality result holds for the CCBC Problem. We will be using some similar arguments as in [1], Proposition 2.2.

Theorem 5.2.5. *Under the assumptions 1 and 2, the strong duality holds,*

$$i.e. \qquad \mathrm{OT}_{CCBC} = \mathrm{OT}^*_{CCBC}.$$

Proof. Let $\nu \in \mathcal{P}' \cap \mathcal{P}(K)$, $\gamma_i \in \Pi(\nu_i, \nu)$ such that $\gamma_i \leq \tilde{\gamma}_i$ and $(f_1, \dots, f_p, w_1, \dots, w_p)$ such that $\sum_{i=1}^{p} f_i = 0$ and $w_i \leq 0$.

Then, for each $(x, y) \in K \times K$, we have

$$\inf_{z \in K} \left\{ \frac{\lambda_i}{2} |x - z|^2 - f_i(z) - w_i(x, z) \right\} \leq \frac{\lambda_i}{2} |x - y|^2 - f_i(y) - w_i(x, y).$$

Hence,

$$\inf_{z \in K} \left\{ \frac{\lambda_i}{2} |x - z|^2 - f_i(z) - w_i(x, z) \right\} + f_i(y) + w_i(x, y) \leq \frac{\lambda_i}{2} |x - y|^2.$$

Now, by integrating w.r.t. γ_i over $K \times K$, we get

$$\int_K \inf_{z \in K} \left\{ \frac{\lambda_i}{2} |x - z|^2 - f_i(z) - w_i(x, z) \right\} d\nu_i(x) + \int_K f_i(y)\ d\nu(y) + \int_{K \times K} w_i(x, y)\ d\gamma_i(x, y)$$
$$\leq \int_{K \times K} \frac{\lambda_i}{2} |x - y|^2\ d\gamma_i(x, y). \quad (5.2.10)$$

Since $\gamma_i \leq \tilde{\gamma}_i$ and $w_i \leq 0$, we have

$$\int_{K \times K} w_i(x, y)\ d\tilde{\gamma}_i(x, y) \leq \int_{K \times K} w_i(x, y)\ d\gamma_i(x, y). \quad (5.2.11)$$

By combining (5.2.10) and (5.2.11), we get

$$\int_K \inf_{z \in K} \left\{ \frac{\lambda_i}{2} |x - z|^2 - f_i(z) - w_i(x, z) \right\} d\nu_i(x) + \int_K f_i(y)\ d\nu(y) + \int_{K \times K} w_i(x, y)\ d\tilde{\gamma}_i(x, y)$$
$$\leq \int_{K \times K} \frac{\lambda_i}{2} |x - y|^2\ d\gamma_i(x, y). \quad (5.2.12)$$

Now, we take the summation over i to get

$$\sum_{i=1}^{p} \int_K \inf_{z\in K} \left\{ \frac{\lambda_i}{2}|x-z|^2 - f_i(z) - w_i(x,z) \right\} d\nu_i(x) + \sum_{i=1}^{p} \int_{K\times K} w_i(x,y)\ d\tilde{\gamma}_i(x,y)$$
$$\leq \sum_{i=1}^{p} \int_{K\times K} \frac{\lambda_i}{2}|x-y|^2\ d\gamma_i(x,y). \quad (5.2.13)$$

Note that, in (5.2.13), we used the fact that $\sum_{i=1}^{p} f_i = 0$.

Now, by taking the infimum over $\gamma_i \in \Pi^{\tilde{\gamma}_i}(\nu_i, \nu)$, infimum over $\nu \in \mathcal{P}'$ and supremum over $(f_i, w_i) \in \mathcal{A}$ from both sides, we get

$$\mathrm{OT}^*_{\mathrm{CCBC}} \leq \mathrm{OT}_{\mathrm{CCBC}}\,. \quad (5.2.14)$$

Next, we will prove that

$$\mathrm{OT}^*_{\mathrm{CCBC}} \geq \mathrm{OT}_{\mathrm{CCBC}}\,.$$

For $i \in \{1, \ldots, p\}$, define $H_i : C_b(K) \times C_b(K \times K) \mapsto \mathbb{R} \cup \{\infty\}$ by

$$H_i(f_i, w_i) = \begin{cases} -\displaystyle\int_K \inf_{z\in K} \left\{ \frac{\lambda_i}{2}|x-z|^2 - f_i(z) - w_i(x,z) \right\} d\nu_i(x) & \\ \qquad - \int_{K\times K} w_i(x,y)\ d\tilde{\gamma}_i(x,y) & \text{if } w_i \leq 0 \\ \infty & \text{otherwise} \end{cases} . \quad (5.2.15)$$

Now, we take the Legendre transform of $H_i(f_i, w_i)$ in both variables to get

$$H_i^*(\nu, \xi) = \sup_{f_i \in C_b(K), w_i \leq 0} \left\{ \int_K f_i\ d\nu + \int_{K\times K} w_i\ d\xi - H_i(f_i, w_i) \right\}. \quad (5.2.16)$$

Here, $\nu \in \mathcal{M}(K)$ and $\xi \in \mathcal{M}(K \times K)$.

Letting $\xi = 0$ in (5.2.16), we get

$$H_i^*(\nu, 0) = \sup_{f_i \in C_b(K), w_i \leq 0} \left\{ \int_K \inf_{z\in K} \left\{ \frac{\lambda_i}{2}|x-z|^2 - f_i(z) - w_i(x,z) \right\} d\nu_i(x) + \int_K f_i(y)\ d\nu(y) \right.$$
$$\left. + \int_{K\times K} w_i(x,y)\ d\tilde{\gamma}_i(x,y) \right\}. \quad (5.2.17)$$

First, we will show that if $\nu \in \mathcal{M}(K) \setminus \mathcal{P}(K)$, then $H_i^*(\nu, 0) = \infty$.

Case 1: Suppose $\exists A \in \mathcal{B}(K)$ such that $\nu(A) < 0$.

Then, there exists a function $f \in C_b(K)$ such that $f \leq 0$ and $\int_K f \mathrm{d}\nu > 0$. Let $t \geq 0$ be arbitrary and choose $f_i = tf$ and $w_i = 0$. Then,

$$\begin{aligned} H_i^*(\nu, 0) &\geq \int_K \inf_{y \in K} \left\{ \frac{\lambda_i}{2}|x-y|^2 - tf(y) \right\} d\nu_i + \int_K tf \; d\nu \\ &\geq t \int_K f \; d\nu, \quad \forall t \geq 0. \end{aligned}$$

Hence, $H_i^*(\nu, 0) = \infty$.

Case 2: Suppose $\nu(K) < 1$.

Let $t > 0$ be arbitrary and choose $f_i = -t$ and $w_i = 0$. Then,

$$\begin{aligned} H_i^*(\nu, 0) &\geq \int_K \inf_{y \in K} \left\{ \frac{\lambda_i}{2}|x-y|^2 + t \right\} d\nu_i - \int_K t \; d\nu \\ &\geq t \left(\int_K d\nu_i - \int_K d\nu \right), \quad \forall t \geq 0 \\ &= (1 - \nu(K))t, \quad \forall t \geq 0. \end{aligned}$$

Hence, $H_i^*(\nu, 0) = \infty$.

Case 3: Suppose $\nu(K) > 1$.

Let $t > 0$ be arbitrary and choose $f_i = t$ and $w_i = 0$. Then,

$$\begin{aligned} H_i^*(\nu, 0) &\geq \int_K \inf_{y \in K} \left\{ \frac{\lambda_i}{2}|x-y|^2 - t \right\} d\nu_i + \int_K t \; d\nu \\ &\geq t \left(-\int_K d\nu_i + \int_K d\nu \right), \quad \forall t \geq 0 \\ &= (-1 + \nu(K))t, \quad \forall t \geq 0. \end{aligned}$$

Hence, $H_i^*(\nu, 0) = \infty$.

Therefore, whenever $\nu \in \mathcal{M}(K) \setminus \mathcal{P}(K)$, we have $H_i^*(\nu, 0) = \infty$.

Now, let $\nu \in \mathcal{P}(K)$. By the duality of the two-marginal CCOT Problem (Theorem 3.4.1), we get

$$H_i^*(\nu, 0) = \lambda_i \widetilde{W}_2^2(\nu_i, \nu). \tag{5.2.18}$$

Hence, we have

$$\mathrm{OT}_{\mathrm{CCBC}} = \inf_{\nu} \sum_{i=1}^{p} \lambda_i \widetilde{W}_2^2(\nu_i, \nu) = \inf_{\nu} \sum_{i=1}^{p} H_i^*(\nu, 0). \tag{5.2.19}$$

Now, consider the Legendre transform of $\sum_{i=1}^{p} H_i^*(\cdot, 0)$, restricted to $C_b(K)$ viewed as a subspace of $C_b(K)^{**}$.

$$\left(\sum_{i=1}^{p} H_i^*(\cdot, 0)\right)^*(f) = \sup_{\nu} \left\{ \int_K f \, d\nu - \sum_{i=1}^{p} H_i^*(\nu, 0) \right\}.$$

where $f \in C_b(K)$ and $\nu \in \mathcal{M}(K)$.

Letting $f = 0$, we get

$$\left(\sum_{i=1}^{p} H_i^*(\cdot, 0)\right)^*(0) = \sup_{\nu} \left\{ -\sum_{i=1}^{p} H_i^*(\nu, 0) \right\}.$$

Hence,

$$-\left(\sum_{i=1}^{p} H_i^*(\cdot, 0)\right)^*(0) = \inf_{\nu} \left\{ \sum_{i=1}^{p} H_i^*(\nu, 0) \right\}. \tag{5.2.20}$$

By combining (5.2.19) and (5.2.20), we get

$$\mathrm{OT}_{\mathrm{CCBC}} = -\left(\sum_{i=1}^{p} H_i^*(\cdot, 0)\right)^*(0). \tag{5.2.21}$$

Now, we define the infimal convolution of $H_i's$ as

$$H(f) = \inf \left\{ \sum_{i=1}^{p} H_i(f_i, w_i) : \sum_{i=1}^{p} f_i = f, \ w_i \le 0 \right\}. \tag{5.2.22}$$

Then,

$$H(0) = \inf \left\{ \sum_{i=1}^{p} H_i(f_i, w_i) : \sum_{i=1}^{p} f_i = 0, \ w_i \le 0 \right\}.$$

So,

$$\begin{aligned} -H(0) &= \sup \left\{ -\sum_{i=1}^{p} H_i(f_i, w_i) : \sum_{i=1}^{p} f_i = 0, \ w_i \le 0 \right\}. \\ &= \sup_{\substack{\sum_{i=1}^{p} f_i = 0, \\ w_i \le 0}} \left\{ \sum_{i=1}^{p} \int_K \inf_{z \in K} \left\{ \frac{\lambda_i}{2} |x - z|^2 - f_i(z) - w_i(x, z) \right\} d\nu_i(x) \right. \\ &\qquad \left. + \sum_{i=1}^{p} \int_{K \times K} w_i(x, y) \, d\tilde{\gamma}_i(x, y) \right\}. \end{aligned}$$

Thus, we have

$$-H(0) = \mathrm{OT}^*_{\mathrm{CCBC}}\,. \tag{5.2.23}$$

Also note that,

$$\begin{aligned}
H^*(\nu) &= \sup_f \left\{ \int_K f \; d\nu - H(f) \right\} \\
&= \sup_f \left\{ \int_K f \; d\nu - \inf \left\{ \sum_{i=1}^p H_i(f_i, w_i) : \sum_{i=1}^p f_i = f, \;\; w_i \le 0 \right\} \right\} \\
&= \sup_f \left\{ \int_K f \; d\nu + \sup \left\{ -\sum_{i=1}^p H_i(f_i, w_i) : \sum_{i=1}^p f_i = f, \;\; w_i \le 0 \right\} \right\} \\
&= \sup_f \left\{ \sum_{i=1}^p \int_K f_i \; d\nu + \sup_{f_i \in C_b(K), w_i \le 0} \left\{ -\sum_{i=1}^p H_i(f_i, w_i) \right\} \right\} \\
&= \sup_{f_i \in C_b(K), w_i \le 0} \left\{ \sum_{i=1}^p \int_K f_i \; d\nu - \sum_{i=1}^p H_i(f_i, w_i) \right\} \\
&= \sum_{i=1}^p \sup_{f_i \in C_b(K), w_i \le 0} \left\{ \int_K f_i \; d\nu - H_i(f_i, w_i) \right\} \\
&= \sum_{i=1}^p H_i^*(\nu, 0).
\end{aligned}$$

Hence, we have

$$H^*(\nu) = \sum_{i=1}^p H_i^*(\nu, 0). \tag{5.2.24}$$

Now, by combining (5.2.21) and (5.2.24), we get

$$\mathrm{OT}_{\mathrm{CCBC}} = -H^{**}(0). \tag{5.2.25}$$

Since we have (5.2.23), it only remains to show that

$$H^{**}(0) = H(0). \tag{5.2.26}$$

Since H is convex, it is sufficient (see [20]) to show that H is continuous at 0 for the norm topology given by (5.1.2). To prove that, we rewrite H_i defined in (5.2.15) as

$$H_i(f_i, w_i) = \int_K \sup_{z \in K} \left\{ f_i(z) + w_i(x, z) - \frac{\lambda_i}{2}|x - z|^2 \right\} \, d\nu_i(x) - \int_{K \times K} w_i(x, y) \; d\tilde{\gamma}_i(x, y).$$

Plugging in $z = 0$, we get

$$H_i(f_i, w_i) \geq f_i(0) + \int_K w_i(x,0)\, d\nu_i(x) - \frac{\lambda_i}{2} \int_K |x|^2\, d\nu_i(x) - \int_{K\times K} w_i(x,y)\, d\tilde{\gamma}_i(x,y). \quad (5.2.27)$$

Since $\forall 1 \leq i \leq p$, $w_i \in C_b(K \times K)$, there exist some negative real numbers m_i such that $m_i \leq w_i(x, 0) \leq 0$, for all $x \in K$.

Thus, (5.2.27) becomes

$$H_i(f_i, w_i) \geq f_i(0) + m_i - \frac{\lambda_i}{2} \int_K |x|^2\, d\nu_i(x). \quad (5.2.28)$$

Note that, in (5.2.28), the last integral is non-negative since $w_i \leq 0$.

Now, taking the summation over i, we get

$$\sum_{i=1}^{p} H_i(f_i, w_i) \geq \sum_{i=1}^{p} f_i(0) + \sum_{i=1}^{p} m_i - \frac{\lambda_i}{2} \sum_{i=1}^{p} \int_K |x|^2\, d\nu_i(x).$$

Now, for all $f \in C_b(K)$ such that $\sum_{i=1}^{p} f_i = f$, we have

$$H(f) \geq f(0) + \sum_{i=1}^{p} m_i - \frac{\lambda_i}{2} \sum_{i=1}^{p} \int_K |x|^2\, d\nu_i(x). \quad (5.2.29)$$

By the finite second moment condition of each of the ν_i's, we get that,

$$H(f) > -\infty. \quad (5.2.30)$$

Now, let f be such that $||f||_Y \leq \frac{p}{4} \min\{\lambda_1, \ldots, \lambda_p\}$. Choosing $f_i = \frac{f}{p}$ and $w_i = 0$ in $H(f_i, w_i)$, we have

$$\begin{aligned}
H(f) &\leq \sum_{i=1}^{p} H_i\left(\frac{f}{p}, 0\right) \\
&\leq \sum_{i=1}^{p} \int_K \sup_{y\in K} \left\{\frac{f(y)}{p} - \frac{\lambda_i}{2}|x-y|^2\right\} d\nu_i \\
&\leq \sum_{i=1}^{p} \int_K \sup_{y\in K} \left\{\frac{\lambda_i}{4}(1+|y|^2) - \frac{\lambda_i}{2}|x-y|^2\right\} d\nu_i \\
&= \sum_{i=1}^{p} \int_K \left(\frac{\lambda_i}{4} + \frac{\lambda_i}{2}|x|^2\right) d\nu_i \\
&= \frac{1}{4} + \sum_{i=1}^{p} \frac{\lambda_i}{2} \int_K |x|^2\, d\nu_i.
\end{aligned}$$

Again, by the finite second moment condition of each of the ν_i's, we get that,

$$H(f) < \infty. \tag{5.2.31}$$

Thus, by (5.2.30) and (5.2.31), we have shown that the convex function H never takes the value $-\infty$ and is bounded from above in a neighborhood of 0. Thus, by a standard convex analysis result (see [20], Proposition 2.5), H is continuous at 0. Hence, $H^{**}(0) = H(0)$.

Thus, by combining (5.2.23) and (5.2.25), we get

$$\mathrm{OT}_{\mathrm{CCBC}} = \mathrm{OT}^*_{\mathrm{CCBC}}. \tag{5.2.32}$$

■

CHAPTER 6

ENTROPY REGULARIZATION

6.1 Entropy Regularized Optimal Transport (EROT) Problem

6.1.1 Introduction

Optimal transport offers an elegant solution to many theoretical and practical problems at the interface between probability, partial differential equations, and optimization. This success however comes at a high computational price, and the computation becomes prohibitive whenever the dimension exceeds a few hundred, since it requires the solution of a linear program over distributions on a product space. Entropic regularization provides us with an approximation of optimal transport, with lower computational complexity and easy implementation. It operates by adding an entropic regularization penalty to the original problem making it a strictly convex problem, hence guaranteeing a unique minimizer.

In this section, we explore the EROT problem and its duality.

The idea of smoothing the classical OTP with an entropic regularization term was first introduced by Cuturi in [16] and it has been shown that the resulting optimum can be computed through Sinkhorn's matrix scaling algorithm at a speed that is several orders of magnitude faster than that of transport solvers. Duality results and a characterization of the dual maximizers have been presented by Marino and Gerolin in [39, 17]. Some of the main results, definitions and remarks mentioned in this section are taken from [39, 45] .

6.1.2 The Primal Problem

Let (X, d_X) and (Y, d_Y) be Polish Spaces, $c : X \times Y \to \mathbb{R}$ be a cost function, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be probability measures and $\varepsilon > 0$. We consider the following minimization problem:

$$\mathrm{OT}_\varepsilon = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{X\times Y} c(x, y)d\gamma(x, y) + \varepsilon\,\mathrm{KL}(\gamma|\mu \otimes \nu). \tag{6.1.1}$$

The function $\mathrm{KL}(\gamma|\xi)$ in (6.1.1) is the Kullback-Leibler divergence between the probability measures γ and ξ in $\mathcal{P}(X \times Y)$ which is defined as

$$\mathrm{KL}(\gamma|\xi) = \begin{cases} \displaystyle\int_{X\times Y} \gamma \log(\gamma) d(\xi) & \text{if } \gamma \ll \xi \\ +\infty & \text{otherwise} \end{cases}.$$

Here, γ denotes the Radon-Nikodym derivative of γ with respect to the measure ξ.

The problem (6.1.1) can also be represented as

$$\mathrm{OT}_\varepsilon = \inf_{\gamma \in \Pi(\mu,\nu)} \varepsilon\, \mathrm{KL}(\gamma|\mathcal{K}) \tag{6.1.2}$$

where $\mathcal{K}$ represents the *Gibbs Kernel* for the cost c, given by:

$$\mathcal{K} = k(x,y)\mu \otimes \nu = e^{-\frac{c(x,y)}{\varepsilon}} \mu \otimes \nu. \tag{6.1.3}$$

The existence of a minimizer in (6.1.1) and its characterizations have been discussed by several authors under different settings ([9], [15], [29] [49]).

Under the definition (6.1.3) of $\mathcal{K}$ ([37], Proposition 2.3), a necessary and sufficient condition for a minimizer, γ_ε, of (6.1.1) to be unique is given by:

$$\gamma_\varepsilon = f_\varepsilon(x) g_\varepsilon(y) \mathcal{K}, \text{ where } f_\varepsilon, g_\varepsilon \text{ solve } \begin{cases} f_\varepsilon(x) \int_Y g_\varepsilon(y) k(x,y) d\nu(y) = 1 \\ g_\varepsilon(y) \int_X f_\varepsilon(y) k(x,y) d\mu(x) = 1 \end{cases}. \tag{6.1.4}$$

Here, the functions $f_\varepsilon(x)$ and $g_\varepsilon(y)$ are known as the *Entropic potentials* and they are unique up to the trivial transformation $f \mapsto f/\lambda$, $g \mapsto \lambda g$ for some $\lambda > 0$. The system solved by the Entropic potentials is called the *Schrödinger system.* The following theorem states that if we assume that μ and ν are positive everywhere and their entropy w.r.t. $\mathcal{K}$ is finite, then the minimizer of (6.1.1) takes a special form.

Theorem 6.1.1. *([9], Corollary 3,9) Let (X, d_X) and (Y, d_Y) be two Polish spaces, and $c : X \times Y \to [0, \infty)$ be a bounded cost function. Suppose that $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ are* *two probability measures that are absolutely continuous w.r.t. Lebesgue measure, such that* $\mu(x), \nu(y) > 0, \forall x \in X, y \in Y,$ *and* $\mathrm{KL}(\mu \otimes \nu | \mathcal{K}) < +\infty$ *for* $\mathcal{K} = e^{-\frac{c(x,y)}{\varepsilon}} \mu \otimes \nu$*. Then,* OT_ε *has a unique minimizer of the form*

$$\gamma_\varepsilon(x, y) = f_\varepsilon(x) g_\varepsilon(y) \mathcal{K}(x, y).$$

6.1.3 Duality

We define a dual functional for (6.2.1) as follows:

$$D_\varepsilon(u, v) = \int_X u(x) d\mu(x) + \int_Y v(y) d\nu(y) - \varepsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\mu \otimes \nu).$$

Then, the dual formulation of (6.2.1) is the following maximization problem:

$$\mathrm{OT}^*_\varepsilon = \sup_{u \in C_b(X), v \in C_b(Y)} D_\varepsilon(u, v). \tag{6.1.5}$$

Now, we will present the duality result for the EROT problem. This has been discussed in several articles, such as [14, 21, 28, 27, 39].

Theorem 6.1.2. *([39], Proposition 2.11) Let* $(X, d_X), (Y, d_Y)$ *be two Polish metric space,* $c : X \times Y \mapsto \mathbb{R}$ *be a bounded function and let* $\varepsilon > 0$*. Suppose that* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$ *are two probability measures that are absolutely continuous w.r.t. Lebesgue measure. Then,*

$$\mathrm{OT}_\varepsilon = \mathrm{OT}^*_\varepsilon + \varepsilon.$$

In [39], the authors also provide a direct proof of existence of maximizers in (6.1.5) following the direct method of Calculus of Variations. The main idea in the proof is to define a generalized version of c-transform (see Definition (2.2.1)), called the (c, ε)-transform.

6.1.3.1 Entropy-Transform and its properties

In this section, we will consider functions $c : X \times Y \to \mathbb{R}$ which are bounded.

First, we will define the space L^{exp}_ε, in which the functions admissible for the dual problem will lie.

Definition 6.1.3. *Given an $\varepsilon > 0$, a Polish space (X, d_X), and a probability measure $\mu \in \mathcal{P}(X)$, we define the set $L_\varepsilon^{exp}(X, d\mu)$ by*

$$L_\varepsilon^{exp}(X, d\mu) := \Big\{ u : X \to [-\infty, \infty) : u \text{ is a measurable function in } (X, d\mu) \text{ and } 0 < \int_X e^{u/\varepsilon}\, d\mu < \infty \Big\}.$$

For $u \in L_\varepsilon^{exp}(X, d\mu)$ we also define $\sigma_u := \varepsilon \log \left(\int_X e^{u/\varepsilon}\, d\mu \right)$.

Observe that $u \in L_\varepsilon^{\exp}(X, d\mu)$ may take the value $-\infty$ on a set of positive measure, but not μ-a.e., since we have a condition $\int_X e^{u/\varepsilon}\, d\mu > 0$.

Definition 6.1.4. *Given an $\varepsilon > 0$, two Polish spaces $(X, d_X), (Y, d_Y)$, two probability measures $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, and a bounded function $c : X \times Y \mapsto \mathbb{R}$, we define the (c, ε)-transform $\mathcal{F}^{(c,\varepsilon)} : L_\varepsilon^{exp}(X, d\mu) \to L^0(Y, d\nu)$ by*

$$\mathcal{F}^{(c,\varepsilon)}(u)(y) := -\varepsilon \log \left(\int_X e^{\frac{u(x) - c(x,y)}{\varepsilon}}\, d\mu(x) \right). \tag{6.1.6}$$

Similarly, we define the (c^*, ε)-transform $\mathcal{F}^{(c^*,\varepsilon)} : L_\varepsilon^{exp}(Y, d\nu) \to L^0(X, d\mu)$ by

$$\mathcal{F}^{(c^*,\varepsilon)}(v)(x) := -\varepsilon \log \left(\int_Y e^{\frac{v(y) - c(x,y)}{\varepsilon}}\, d\nu(y) \right). \tag{6.1.7}$$

For simplicity, we will use the notation $u^{(c,\varepsilon)} = \mathcal{F}^{(c,\varepsilon)}(u)$.

Remark 6.1.5. *Note that the definition of (c, ε)-transform is consistent with the definition of c-transform in the sense that as $\varepsilon \to 0$, $u^{(c,\varepsilon)} \to u^c$ ([26], Lemma 4.2).*

Now we will list some results about the (c, ε)-transform mentioned in [39] without proof.

Lemma 6.1.6. *([39], Lemma 2.3) Let (X, d_X), (Y, d_Y) be Polish spaces, $u \in L_\varepsilon^{exp}(X, d\mu)$, $v \in L_\varepsilon^{exp}(Y, d\nu)$ and $\varepsilon > 0$. Then,*

(i) $u^{(c,\varepsilon)} \in L^\infty(Y, d\nu)$ and $v^{(c,\varepsilon)} \in L^\infty(X, d\mu)$, satisfying the inequality

$$-\|c\|_\infty - \varepsilon \log \left(\int_X e^{\frac{u(x)}{\varepsilon}}\, d\mu \right) \le u^{(c,\varepsilon)}(y) \le \|c\|_\infty - \varepsilon \log \left(\int_X e^{\frac{u(x)}{\varepsilon}}\, d\mu \right).$$

(ii) $u^{(c,\varepsilon)} \in L^{exp}_{\varepsilon}(Y, d\nu)$ *and* $v^{(c,\varepsilon)} \in L^{exp}_{\varepsilon}(X, d\mu)$.

Furthermore, $|\sigma_{u^{(c,\varepsilon)}} + \sigma_u| \leq \|c\|_{\infty}$.

Given below are some more properties of the (c, ε)-transform.

Proposition 6.1.7. *([39], Prop 2.4) Let* (X, d_X) *and* (Y, d_Y) *be Polish metric spaces,* $c : X \times Y \to [0, \infty]$ *be a bounded function,* $\mu \in \mathcal{P}(X)$*,* $\nu \in \mathcal{P}(Y)$ *be probability measures,* $u \in L^{exp}_{\varepsilon}(X, d\mu)$ *and* $\varepsilon > 0$*. Then,*

(i) if c *is L-Lipschitz, then* $u^{(c,\varepsilon)}$ *is L-Lipschitz;*

(ii) if c *is* ω*-continuous, then* $u^{(c,\varepsilon)}$ *is* ω*-continuous;*

(iii) if $|c| \leq M$*, then* $|u^{(c,\varepsilon)} + \sigma_u| \leq M$*;*

(iv) if $|c| \leq M$*, then* $\mathcal{F}^{(c,\varepsilon)} : L^{\infty}(X, d\mu) \to L^p(Y, d\nu)$ *is a* 1*-Lipschitz compact operator.*

(v) if c *is* K*-concave with respect to* y*, then* $u^{(c,\varepsilon)}$ *is* K*-concave.*

6.1.3.2 Dual Problem

Recall that the dual functional of the EROT problem is given by

$$D_{\varepsilon}(u, v) = \int_X u(x) d\mu(x) + \int_Y v(y) d\nu(y) - \varepsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\mu \otimes \nu). \tag{6.1.8}$$

Now, we will state some results used to get the existence of dual maximizers.

Lemma 6.1.8. *([39], Lemma 2.6) Consider the functional* $D_{\varepsilon} : L^{exp}_{\varepsilon}(X, d\mu) \times L^{exp}_{\varepsilon}(Y, d\nu) \to \mathbb{R}$ *defined by* (6.1.8)*. Then,*

$$D_{\varepsilon}(u, u^{(c,\varepsilon)}) \geq D_{\varepsilon}(u, v) \qquad \forall v \in L^{exp}_{\varepsilon}(Y, d\nu), \tag{6.1.9}$$

$$D_{\varepsilon}(u, u^{(c,\varepsilon)}) = D_{\varepsilon}(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}. \tag{6.1.10}$$

In particular, $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_{\varepsilon}(u, v) \;:\; v \in L^{exp}_{\varepsilon}(Y, d\nu)\}$.

Lemma 6.1.9. *([39], Lemma 2.7) Given* $u \in L^{exp}_{\varepsilon}(X, d\mu)$ *and* $v \in L^{exp}_{\varepsilon}(Y, d\nu)$*, there exist* $u^* \in L^{exp}_{\varepsilon}(X, d\mu)$ *and* $v^* \in L^{exp}_{\varepsilon}(Y, d\nu)$ *such that*

- $D_{\varepsilon}(u, v) \leq D_{\varepsilon}(u^*, v^*)$,
- $\|v^*\|_{\infty} \leq 3\|c\|_{\infty}/2$,
- $\|u^*\|_{\infty} \leq 3\|c\|_{\infty}/2$.

Moreover we can choose $a \in \mathbb{R}$ *such that* $u^* = (v + a)^{(c,\varepsilon)}$ *and* $v^* = (u^*)^{(c,\varepsilon)}$.

The theorem given below states that the dual functional (6.1.8) attains a maximizer.

Theorem 6.1.10. *([39], Theorem 2.8) Let* (X, d_X), (Y, d_Y) *be Polish spaces,* $c : X \times Y \to \mathbb{R}$ *be a bounded function,* $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ *be probability measures and* $\varepsilon > 0$. *Then, the dual problem given by*

$$\sup \{D_{\varepsilon}(u, v) \; : \; u \in L^{exp}_{\varepsilon}(X, d\mu), v \in L^{exp}_{\varepsilon}(Y, d\nu)\} \tag{6.1.11}$$

attains the supremum for a unique couple (u_0, v_0) *(up to the trivial tranformation* $(u, v) \mapsto (u + a, v - a)$*). In particular we have*

$$u_0 \in L^{\infty}(X, d\mu) \quad \text{and} \quad v_0 \in L^{\infty}(Y, d\nu);$$

moreover, we can choose the maximizers such that $\|u_0\|_{\infty}, \|v_0\|_{\infty} \leq \frac{3}{2}\|c\|_{\infty}$.

The proof mainly uses Lemma 6.1.8 and Lemma 6.1.9 along with Banach-Alaoglu theorem.

6.2 Entropy Regularized Barycenter Problem

6.2.1 Introduction

Similar to finding optimizers of the classical OT problem, finding the Wasserstein barycenter is also not an easy task. By introducing the entropy regularization to the classical Wasserstein Barycenter problem, it becomes more tractable and flexible.

Two different notions of regularization exist in the literature. In [7, 12], the authors introduce a penalization term which is a function of the barycenter. On the other hand, in [38], the authors add a penalty term which is a function of the transport plans between the barycenter and the target measures. The main difference between these two approaches is the reference point that is used to regularize the resulting measure. The choice of regularization approach depends on the specific application and the desired properties of the resulting measure. However, the similarities or the differences of these approaches have not been discussed widely in the literature. In this section, as in [38] we will be considering the latter approach.

In [38], the authors introduce the regularized Wasserstein barycenter problem and its dual formulation. They prove that the strong duality holds and the existence of the primal problem via duality result. However, they do not discuss the existence of the maximizers of the dual functional.

In this section, we will provide a direct proof for the existence of a minimizer for the primal problem and the existence of dual maximizers.

6.2.2 The Primal Problem

Throughout this section, we will be working with compact subsets of $\mathbb{R}^d$ and symmetric, bounded cost functions. Recall that the entropy regularized OT problem is defined as

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon\,\mathrm{KL}(\gamma|\mu\otimes\nu). \tag{6.2.1}$$

By the duality result, we have

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \sup_{u\in C_b(X), v\in C_b(Y)} \int_X u(x)d\mu(x) + \int_Y v(y)d\nu(y) - \varepsilon\int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\mu\otimes\nu).$$

Also recall that given an integer $p\geq 2$, $X_1,\dots,X_p, Y$ compact subsets of $\mathbb{R}^d$, a p-tuple of probability measures $(\nu_1,...,\nu_p)$ each in $\mathcal{P}(X_i)$, a p-tuple of positive real numbers $(\lambda_1,...,\lambda_p)$ with $\sum_{i=1}^p \lambda_i{=}1$, and cost functions $c_i : X_i\times Y \mapsto \mathbb{R}$, we define the Wasserstein Barycenter problem as

$$\mathrm{OT}_{BC}(\nu_1,\dots,\nu_p) = \inf_{\nu\in\mathcal{P}(Y)} \sum_{i=1}^p \lambda_i \mathrm{OT}_{c_i}(\nu_i,\nu) \tag{6.2.2}$$

where $\mathrm{OT}_{c_i}(\nu_i, \nu)$ is the optimal transport cost between ν_i and ν with cost c_i given in (2.1.6).

In [1], Proposition 2.2, gives us the following duality result for the case $c_i(x, y) = \frac{1}{2}|x-y|^2$.

$$\mathrm{OT}_{BC}(\nu_1, \ldots, \nu_p) = \sup_{f_i \in C_b(Y), \sum_{i=1}^p f_i = 0} \sum_{i=1}^{p} \int_{X_i} \inf_{y \in R^d} \Big\{ \lambda_i c_i(x_i, y) - f_i(y) \Big\} \, d\nu_i(x_i). \tag{6.2.3}$$

Now we will introduce the Entropy Regularized Barycenter (ERBC) problem.

Given an integer $p \geq 2$, $X_1, \ldots, X_p, Y$ compact subsets of $\mathbb{R}^d$, a p-tuple of probability measures $(\nu_1,\ldots,\nu_p)$ each in $\mathcal{P}(X_i)$, a p-tuple of positive real numbers $(\lambda_1,\ldots,\lambda_p)$ with $\sum_{i=1}^{p} \lambda_i = 1$, and $\varepsilon > 0$, we define the ERBC problem as

$$\mathrm{OT}^{\varepsilon}_{BC}(\nu_1, \ldots, \nu_p) = \inf_{\nu \in \mathcal{P}(Y)} \sum_{i=1}^{p} \lambda_i \mathrm{OT}_{\varepsilon}(\nu_i, \nu). \tag{6.2.4}$$

Now we will prove that the minimization problem (6.2.4) has a minimizer. For simplicity we will assume that $c_i(x, y) = \frac{1}{2}|x - y|^2$ for each $1 \leq i \leq p$ in this proof, but we can prove that this result holds for more general costs. Also note that in [38], the authors have provided a proof for the existence of a minimizer via duality and here we provide a direct proof without assuming the duality result.

Lemma 6.2.1. $\mathrm{OT}_{\varepsilon}$ *is weakly lower semi-continuous.*

Proof. Let $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ be two sequences such that $\mu_n \rightharpoonup \mu^*$ and $\nu_n \rightharpoonup \nu^*$. We can pick subsequences (not relabeled) $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \mathrm{OT}_{\varepsilon}(\mu_n, \nu_n) = \liminf_{n\to\infty} \mathrm{OT}_{\varepsilon}(\mu_n, \nu_n).$$

Since $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ are tight $\bigcup_{n\in\mathbb{N}} \Pi(\mu_n, \nu_n)$ is also tight (see [53], Lemma 4.4). Now let $\{\gamma_n\}_{n\in\mathbb{N}}$ be a sequence with marginals μ_n and ν_n which is "close" to the optimal value of $\mathrm{OT}_{\varepsilon}(\mu_n, \nu_n)$,

i.e. given $\delta > 0$,

$$\int |x - y|^2 \, d\gamma_n + \varepsilon \, \mathrm{KL}(\gamma_n | \mu_n \otimes \nu_n) \leq \mathrm{OT}_{\varepsilon}(\mu_n, \nu_n) + \delta.$$

Since $\{\gamma_n\}_{n\in\mathbb{N}}$ is also tight, we can pick a subsequence (not relabeled) $\{\gamma_n\}_{n\in\mathbb{N}}$ such that $\gamma_n \rightharpoonup \gamma^*$ and $\gamma^* \in \Pi(\mu^*, \nu^*)$.

Now, observe that,

$$\begin{aligned}
\mathrm{OT}_\varepsilon(\mu^*,\nu^*) &\le \int |x-y|^2 \, d\gamma^* + \varepsilon \, \mathrm{KL}(\gamma^*|\mu^*\otimes\nu^*) \\
&\le \liminf_n \int |x-y|^2 \, d\gamma_n + \varepsilon \liminf_n \mathrm{KL}(\gamma_n|\mu_n\otimes\nu_n) \\
&\le \liminf_n \left\{ \int |x-y|^2 \, d\gamma_n + \varepsilon \, \mathrm{KL}(\gamma_n|\mu_n\otimes\nu_n) \right\} \\
&\le \liminf_n \mathrm{OT}_\varepsilon(\mu_n,\nu_n) + \delta \\
&= \lim_n \mathrm{OT}_\varepsilon(\mu_n,\nu_n) + \delta.
\end{aligned}$$

In the second inequality, we get the first term due to weak lower semi-continuity of the total cost (see [53], Lemma 4.3), and the second term due to the lower semi-continuity of the relative entropy, which is a well-known result (see [19], Lemma 1.4.3).

Finally, letting $\delta \to 0$, we get the lower semi-continuity result. ■

Theorem 6.2.2. *Let $X_1,\dots,X_p,Y$ be compact subsets of $\mathbb{R}^d$. Given an integer $p>0$, a p-tuple of probability measures $(\nu_1,\dots,\nu_p)$ each in $\mathcal{P}_2(X_i)$, a p-tuple of positive real numbers $(\lambda_1,\dots,\lambda_p)$ with $\sum_{i=1}^{p}\lambda_i=1$, and $\varepsilon>0$, there exists a minimizer for the ERBC problem given by* (6.2.4).

Proof. Let

$$I(\nu) := \sum_{i=1}^{p} \lambda_i \mathrm{OT}_\varepsilon(\nu_i,\nu) = \sum_{i=1}^{p} \inf_{\gamma_i\in\Pi(\nu_i,\nu)} \lambda_i \int_{X_i\times Y} \frac{1}{2}|x-y|^2 d\gamma_i(x,y) + \varepsilon \, \mathrm{KL}(\gamma_i|\nu_i\otimes\nu).$$

Also, we will denote the squared 2 - Wasserstein distance by

$$W_2^2(\nu_i,\nu) := \inf_{\gamma_i\in\Pi(\nu_i,\nu)} \int_{X_i\times Y} \frac{1}{2}|x-y|^2 d\gamma_i(x,y).$$

Then, (6.2.4) becomes

$$\mathrm{OT}^\varepsilon_{BC}(\nu_1,\dots,\nu_p) = \inf_{\nu\in\mathcal{P}(Y)} I(\nu).$$

Now, let $\{\nu^n\}_{n\in\mathbb{N}}$ be a minimizing sequence of $\mathrm{OT}^\varepsilon_{BC}$, i.e. $\lim_{n\to\infty} I(\nu_n) = \inf_\nu I(\nu)$. Then, we can find an $M>0$ such that $I(\nu_n)\le M$ for all n. Since $\varepsilon \, \mathrm{KL}(\gamma_i|\nu_i\otimes\nu)\ge 0$, we have for each n,

$$\sum_{i=1}^{p} \lambda_i W_2^2(\nu_i, \nu_n) \leq I(\nu_n) \leq M.$$

Thus, for each $1 \leq i \leq p$ and for each $n \in \mathbb{N}$, $\lambda_i W_2^2(\nu_i, \nu_n) \leq M$.

Using the duality of the Kantorovich problem and the assumption that ν_i's have finite second moments, we can show that (see Appendix 3)

$$\sup_n \int |x|^2 \, d\nu_n \leq C, \qquad \text{for some constant C.}$$

Hence, $\{\nu_n\}$ is tight (see Appendix 4).

Then, by Prokhorov's theorem, there exists a subsequence $\{\nu_n\}$ (not relabeled), that converges weakly to some $\nu^* \in \mathcal{P}(Y)$. Since, $\int |x|^2 \, d\nu^* \leq \liminf_n \int |x|^2 \, d\nu_n \leq C$, we have that $\nu^* \in \mathcal{P}_2(Y)$. Here, the first inequality is again due to the weak convergence of measures for lower semi-continuous bounded below costs ([19], Theorem A.3.12). Note that $\int_X |x|^2 \, d\nu = W_2^2(\nu, \delta_0)$ for any $\nu \in \mathcal{P}_2(Y)$.

Now, we will prove that ν^* is the desired minimizer.

Observe that,

$$\begin{aligned}
\inf_\nu I(\nu) &= \liminf_n I(\nu_n) \\
&= \liminf_n \sum_{i=1}^{p} \lambda_i \, \mathrm{OT}_\varepsilon(\nu_i, \nu_n) \\
&\geq \sum_{i=1}^{p} \liminf_n \lambda_i \, \mathrm{OT}_\varepsilon(\nu_i, \nu_n) \\
&\geq \sum_{i=1}^{p} \lambda_i \, \mathrm{OT}_\varepsilon(\nu_i, \nu^*) \qquad \text{(By Lemma 6.2.1)} \\
&= I(\nu^*).
\end{aligned}$$

This proves that (6.2.4) has a minimizer.

■

Remark 6.2.3. *Note that we assumed that $X_1, \ldots, X_p, Y$ are compact only to be consistent with the assumptions on the ERBC Problem in the original paper [38]. However, we do* *not require the compactness of the spaces in the proof, hence the Theorem 6.2.2 holds for* $X_1 = \ldots = X_p = Y = \mathbb{R}^d$.

6.2.3 Duality

We define the dual functional of the ERBC problem as

$$\mathrm{J}^{\varepsilon}_{BC}(\phi_1,\ldots,\phi_p,\psi_1,\ldots,\psi_p) = \sum_{i=1}^{p} \lambda_i \left(\int_{X_i} \phi_i d\nu_i - \varepsilon \int_{X_i\times Y} e^{\frac{(\phi_i(x_i)+\psi_i(y)-c_i(x_i,y))}{\varepsilon}} d\nu_i \otimes \nu^*(x_i,y) \right). \tag{6.2.5}$$

Since we already know the existence of a barycenter of the minimization problem (6.2.4), say ν^*, we will use ν^* for the following duality results.

Now we will present the duality result for the ERBC problem discussed in [38].

Theorem 6.2.4. *([38], Theorem 3.1) The dual formulation of* (6.2.4) *is*

$$\sup_{\substack{\{(\phi_i\in C_b(X_i),\psi_i\in C_b(Y))\}_{i=1}^p \\ \sum_{i=1}^p \lambda_i\psi_i=0}} \sum_{i=1}^{p} \lambda_i \left(\int_{X_i} \phi_i(x_i) d\nu_i(x_i) - \varepsilon \int_{X_i\times Y} e^{\frac{(\phi_i(x_i)+\psi_i(y)-c_i(x_i,y))}{\varepsilon}} d\nu_i \otimes \nu^*(x_i,y) \right). \tag{6.2.6}$$

Moreover, strong duality holds in the sense that the infimum of (6.2.4) *equals the supremum of* (6.2.6)*, and a solution to* (6.2.4) *exists. If* $\{(\phi_i,\psi_i)\}_{i=1}^p$ *solves* (6.2.6)*, then each* (ϕ_i,ψ_i) *is a solution to the dual formulation* (6.1.5) *of* $\mathrm{OT}_\varepsilon(\nu_i,\nu^*)$.

The proof relies on the convex duality theory of locally convex topological spaces.

Now, we will prove that dual maximizers for (6.2.6) exist.

Theorem 6.2.5. *Given an integer* $p > 2$*, let* $X_1,\ldots,X_p,Y$ *be compact subsets of* $\mathbb{R}^d$*,* $c_i : X_i \times Y \mapsto \mathbb{R}^+$ *be symmetric, bounded cost functions such that for each* $1 \le i \le p$*,* c_i *is* L_i*-Lipschitz,* $(\nu_1,\ldots,\nu_p)$ *be a p-tuple of probability measures each in* $\mathcal{P}(X_i)$*,* $(\lambda_1,\ldots,\lambda_p)$ *be a p-tuple of positive real numbers with* $\sum_{i=1}^{p} \lambda_i = 1$*, and* $\varepsilon > 0$*. Then, there exist functions* $\{(\phi_i \in L^1(X_i,\nu_i), \psi_i \in L^1(Y,\nu^*))\}_{i=1}^p$ *that solve* (6.2.6).

Proof. We will redefine the dual formulation as

$$\mathrm{J}^*_{BC} := \sup_{\substack{\{(\phi_i\in C_b(X_i),\psi_i\in C_b(Y))\}_{i=1}^p \\ \sum_{i=1}^p \lambda_i\psi_i=0}} \sum_{i=1}^p \lambda_i D(\phi_i,\psi_i), \tag{6.2.7}$$

where

$$D(\phi_i,\psi_i) := \int_{X_i}\phi_i d\nu_i + \int_Y \psi_i d\nu^* - \varepsilon\int_{X_i\times Y} e^{\frac{(\phi_i+\psi_i-c_i)}{\varepsilon}} d\nu_i\otimes\nu^*.$$

Let $(\phi_{1,n},\ldots,\phi_{p,n},\psi_{1,n},\ldots,\psi_{p,n})_{n\in\mathbb{N}}$ be a maximizing sequence. i.e. we have that for each $i=1,\ldots,p$, $\phi_{i,n}\in C_b(X_i),\psi_{i,n}\in C_b(Y)$ with $\sum_{i=1}^p \lambda_i\psi_{i,n}=0$ and $\lim_{n\to\infty}\sum_{i=1}^p \lambda_i D(\phi_{i,n},\psi_{i,n}) = \mathrm{J}^*_{BC}$.

For each $1\le i\le p$ define $\widetilde{\phi}_{i,n}=\psi_{i,n}^{c_i,\varepsilon}$. By Lemma 6.1.8, we have that

$$D(\phi_{i,n},\psi_{i,n}) \le D(\widetilde{\phi}_{i,n},\psi_{i,n}). \tag{6.2.8}$$

Now for each $1\le i\le p-1$, define $k_{i,n}=-\sigma_{\psi_{i,n}}$ and $k_{p,n}=-\frac{\sum_{i=1}^{p-1}\lambda_i k_{i,n}}{\lambda_p}$ (the definition of the softmax operator $\sigma_{\psi_{i,n}}$ is given in definition 6.1.3). Note that, we have $\sum_{i=1}^p \lambda_i k_{i,n}=0$.

By Proposition 6.1.7 part (iii), we have that $|\psi_{i,n}^{c_i,\varepsilon}+\sigma_{\psi_{i,n}}|\le ||c_i||_\infty$. A simple calculation gives us that $|\psi_{i,n}^{c_i,\varepsilon}-k_{i,n}|\le ||c_i||_\infty$ for each $1\le i\le p-1$.

Now, for each $1\le i\le p$, we define

$$\phi^*_{i,n}=\widetilde{\phi}_{i,n}-k_{i,n} \quad\text{and}\quad \psi^*_{i,n}=\psi_{i,n}+k_{i,n}.$$

Then for each $1\le i\le p-1$, we have that $||\phi^*_{i,n}||_\infty\le ||c_i||_\infty$. i.e.

$$\sup_n ||\phi^*_{i,n}||_\infty < \infty. \tag{6.2.9}$$

Observe that,

$$\sum_{i=1}^p \lambda_i\psi^*_{i,n} = \sum_{i=1}^p \lambda_i(\psi_{i,n}+k_{i,n}) = 0.$$

Since for each $1\le i\le p$, c_i is L_i-Lipschitz, by Proposition 6.1.7 part (i), for each $1\le i\le p$ and $n\in\mathbb{N}$, $\psi_{i,n}^{c_i,\varepsilon}$ is L_i-Lipschitz. Hence for each $1\le i\le p-1$, $\phi^*_{i,n}=\psi_{i,n}^{c_i,\varepsilon}-k_{i,n}$ is Lipschitz continuous with the same Lipschitz constant $\max_i L_i$.

Now, we will prove that for each $1\le i\le p$, $\sup_n ||\psi^*_{i,n}||_{L^1(Y,d\nu^*)}<\infty$.

Let $1 \leq i \leq p-1$ be arbitrary. Then,

$$\begin{aligned}
\int_Y e^{\frac{\psi^*_{i,n}}{\varepsilon}} d\nu^* &= \int_Y e^{\frac{\psi_{i,n}+k_{i,n}}{\varepsilon}} d\nu^* \\
&= \int_Y e^{\frac{\psi_{i,n}-\sigma_{\psi_{i,n}}}{\varepsilon}} d\nu^* \\
&= \int_Y e^{\frac{\psi_{i,n}-\varepsilon \log \int_Y e^{\frac{\psi_{i,n}}{\varepsilon}} d\nu^*}{\varepsilon}} d\nu^* \\
&= \int_Y e^{\frac{\psi_{i,n}}{\varepsilon}} e^{\log\left(\int_Y e^{\frac{\psi_{i,n}}{\varepsilon}} d\nu^*\right)^{-1}} d\nu^* \\
&= \int_Y e^{\frac{\psi_{i,n}}{\varepsilon}} \left(\int_Y e^{\frac{\psi_{i,n}}{\varepsilon}} d\nu^*\right)^{-1} d\nu^* \\
&= 1.
\end{aligned}$$

Since for a given $1 \leq q < \infty$, there is a constant c such that $ct^q \leq e^t$, for each $t > 0$, we have that for each $1 \leq q < \infty$,

$$\int_Y \left[\frac{\psi^*_{i,n}}{\varepsilon}\right]_+^q d\nu^* \leq \frac{1}{c} \int_Y e^{\frac{\psi^*_{i,n}}{\varepsilon}} d\nu^*.$$

Thus,

$$\int_Y \left[\psi^*_{i,n}\right]_+^q d\nu^* \leq \frac{\varepsilon^q}{c} \quad \text{for some constant } c.$$

So, $\sup_n || \left[\psi^*_{i,n}\right]_+ ||_{L^q(Y,d\nu^*)} < \infty$ for each $1 \leq q < \infty$ and for each $1 \leq i \leq p-1$.

In particular,

$$\sup_n || \left[\psi^*_{i,n}\right]_+ ||_{L^1(Y,d\nu^*)} < \infty, \qquad \forall 1 \leq i \leq p-1. \tag{6.2.10}$$

Now, since $(\phi^*_{i,n}, \psi^*_{i,n})_{1\leq i\leq p, n\in\mathbb{N}}$ is a maximizing sequence, there is some constant c_1 such that

$$\begin{aligned}
-c_1 &\leq \sum_{i=1}^{p} \lambda_i D(\phi^*_{i,n}, \psi^*_{i,n}) \\
&= \sum_{i=1}^{p} \lambda_i \int_{X_i} \phi^*_{i,n} d\nu_i - \varepsilon \sum_{i=1}^{p} \lambda_i \int_{X_i\times Y} e^{\frac{\phi^*_{i,n}+\psi^*_{i,n}-c_i}{\varepsilon}} d\nu_i \otimes \nu^* \\
&= \sum_{i=1}^{p-1} \lambda_i \int_{X_i} \phi^*_{i,n} d\nu_i + \lambda_p \int_{X_p} \phi^*_{p,n} d\nu_p - \varepsilon \sum_{i=1}^{p-1} \lambda_i \int_{X_i\times Y} e^{\frac{\phi^*_{i,n}+\psi^*_{i,n}-c_i}{\varepsilon}} d\nu_i \otimes \nu^* \\
&\quad - \varepsilon \lambda_p \int_{X_p\times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^*.
\end{aligned} \tag{6.2.11}$$

Recall that, for each $1 \le i \le p$, $\phi^*_{i,n} = \psi^{c_i,\varepsilon}_{i,n} - k_{i,n}$ and $\psi^*_{i,n} = \psi_{i,n} + k_{i,n}$.

Hence,

$$\begin{aligned}
\int_{X_i \times Y} e^{\frac{\phi^*_{i,n}+\psi^*_{i,n}-c_i}{\varepsilon}} d\nu_i \otimes \nu^* &= \int_{X_i \times Y} e^{\frac{\psi^{c_i,\varepsilon}_{i,n} - k_{i,n} + \psi_{i,n} + k_{i,n} - c_i}{\varepsilon}} d\nu_i \otimes \nu^* \\
&= \int_{X_i \times Y} e^{\frac{\psi^{c_i,\varepsilon}_{i,n} + \psi_{i,n} - c_i}{\varepsilon}} d\nu_i \otimes \nu^* \\
&= \int_{X_i} e^{\frac{\psi^{c_i,\varepsilon}_{i,n}}{\varepsilon}} \left(\int_Y e^{\frac{\psi_{i,n}-c_i}{\varepsilon}} d\nu^* \right) d\nu_i \qquad (6.2.12) \\
&= \int_{X_i} e^{\frac{\psi^{c_i,\varepsilon}_{i,n}}{\varepsilon}} \cdot e^{-\frac{\psi^{c_i,\varepsilon}_{i,n}}{\varepsilon}} d\nu_i \\
&= 1.
\end{aligned}$$

The penultimate equality holds since,

$$\psi^{c_i,\varepsilon}_{i,n} = -\varepsilon \log \int_Y e^{\frac{\psi_{i,n}-c_i}{\varepsilon}} d\nu^* \implies e^{-\frac{\psi^{c_i,\varepsilon}_{i,n}}{\varepsilon}} = \int_Y e^{\frac{\psi_{i,n}-c_i}{\varepsilon}} d\nu^*.$$

Also, since $||\phi^*_{i,n}||_\infty \le ||c_i||_\infty$ for each $1 \le i \le p-1$, $\sum_{i=1}^{p-1} \lambda_i \int_{X_i} \phi^*_{i,n} d\nu_i \le \max_i ||c_i||_\infty (1-\lambda_p)$. Now, the inequality (6.2.11) becomes

$$-c_1 \le \max_i ||c_i||_\infty (1-\lambda_p) + \lambda_p \int_{X_p} \phi^*_{p,n} d\nu_p - \varepsilon(1-\lambda_p) - \varepsilon\lambda_p \int_{X_p \times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^*.$$

Hence, there exists a constant c_2 such that

$$\frac{c_2}{\lambda_p} \le \int_{X_p} \phi^*_{p,n} d\nu_p - \varepsilon \int_{X_p \times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^*. \qquad (6.2.13)$$

Now, consider

$$\int_{X_p} \phi^*_{p,n} d\nu_p - \varepsilon \int_{X_p \times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^*.$$

Since $f(t) = t - \varepsilon e^{\frac{t-a}{\varepsilon}}$ is concave and it attains its maximum at $t = a$,

$$\begin{aligned}
\int_{X_p} \phi^*_{p,n} d\nu_p - \varepsilon \int_{X_p \times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^* &= \int_{X_p \times Y} \left(\phi^*_{p,n} - \varepsilon e^{\frac{\phi^*_{p,n} - (c_p - \psi^*_{p,n})}{\varepsilon}} \right) d\nu_p \otimes \nu^* \\
&\le \int_{X_p \times Y} (c_p - \psi^*_{p,n} - \varepsilon)\ d\nu_p \otimes \nu^* \\
&\le -\int_Y \psi^*_{p,n}\ d\nu^* + ||c_p||_\infty - \varepsilon.
\end{aligned}$$

Thus, by (6.2.13), there exists a constant c_3 such that

$$c_3 \leq -\int_Y \psi^*_{p,n}\ d\nu^*. \tag{6.2.14}$$

Since $\sum_{i=1}^p \lambda_i \psi^*_{i,n} = 0$, we get that

$$c_3 \leq \frac{1}{\lambda_p} \sum_{i=1}^{p-1} \lambda_i \int_Y \psi^*_{i,n}\ d\nu^*.$$

Hence,

$$\lambda_p c_3 \leq \sum_{i=1}^{p-1} \lambda_i \int_Y \left[\psi^*_{i,n}\right]_+\ d\nu^* - \sum_{i=1}^{p-1} \lambda_i \int_Y \left[\psi^*_{i,n}\right]_-\ d\nu^*.$$

Thus, by (6.2.10), there is some constant c_4 such that

$$\sum_{i=1}^{p-1} \lambda_i \int_Y \left[\psi^*_{i,n}\right]_-\ d\nu^* \leq c_4.$$

Since $\lambda_i \int_Y \left[\psi^*_{i,n}\right]_-\ d\nu^* \geq 0$, we have that $\lambda_i \int_Y \left[\psi^*_{i,n}\right]_-\ d\nu^* \leq c_4$ for each $1 \leq i \leq p-1$. This gives us that

$\sup_n || \left[\psi^*_{i,n}\right]_- ||_{L^1(Y,d\nu^*)} < \infty$ for each $1 \leq i \leq p-1$.

Hence, for each $1 \leq i \leq p-1$,

$$\int_Y |\psi^*_{i,n}|\ d\nu^* = \int_Y \left[\psi^*_{i,n}\right]_+\ d\nu^* + \int_Y \left[\psi^*_{i,n}\right]_-\ d\nu^* < \infty.$$

Thus we have for each $1 \leq i \leq p-1$, $\sup_n ||\psi^*_{i,n}||_{L^1} < \infty$.

Now observe that,

$$\begin{aligned}
||\psi^*_{p,n}||_{L^1(Y,d\nu^*)} &= || -\frac{1}{\lambda_p} \sum_{i=1}^{p-1} \lambda_i \psi^*_{i,n} ||_{L^1(Y,d\nu^*)} \\
&\leq \frac{1}{\lambda_p} \sum_{i=1}^{p-1} \lambda_i ||\psi^*_{i,n}||_{L^1(Y,d\nu^*)} \\
&\leq \max_i ||\psi^*_{i,n}||_{L^1(Y,d\nu^*)} \frac{1}{\lambda_p} \sum_{i=1}^{p-1} \lambda_i \\
&= \frac{1-\lambda_p}{\lambda_p} \max_i ||\psi^*_{i,n}||_{L^1(Y,d\nu^*)} \\
&< \infty.
\end{aligned}$$

Hence,

$$\sup_n ||\psi^*_{i,n}||_{L^1(Y,d\nu^*)} < \infty, \qquad \forall 1 \le i \le p. \tag{6.2.15}$$

Now we will prove that $\sup_n ||\phi^*_{p,n}||_{L^1(X_p,d\nu_p)} < \infty$.

Observe that, for a given $1 \le i \le p$,

$$\begin{aligned}
\phi^*_{i,n} &= (\psi^*_{i,n})^{c_i,\varepsilon} \\
&= -\varepsilon \log \int_Y e^{\frac{\phi^*_{i,n}-c_i}{\varepsilon}} \, d\nu^* \\
&\le -\varepsilon \int_Y \frac{\phi^*_{i,n} - c_i}{\varepsilon} \, d\nu^* \qquad \text{(by Jensen's inequality)} \\
&= \int_Y \left(c_i - \psi^*_{i,n}\right) \, d\nu^* \\
&\le \max_{1\le i\le p} ||c_i||_\infty - \int_Y \psi^*_{i,n} \, d\nu^* \\
&\le c_5, \qquad \text{for some constant } c_5.
\end{aligned}$$

The last inequality above holds by (6.2.15).

Hence, $\forall 1 \le i \le p$,

$$\phi^*_{i,n}(x) \le c_5, \qquad \forall x \in X_i. \tag{6.2.16}$$

In particular,

$$\int_{X_p} \phi^*_{p,n} \, d\nu_p \le c_5, \tag{6.2.17}$$

and

$$\int_{X_p} \left[\phi^*_{p,n}\right]_+ \, d\nu_p \le \max\{0, c_5\}. \tag{6.2.18}$$

In (6.2.13), since $-\varepsilon \int_{X_p\times Y} e^{\frac{\phi^*_{p,n}+\psi^*_{p,n}-c_p}{\varepsilon}} d\nu_p \otimes \nu^* \le 0$, we have that

$$\frac{c_2}{\lambda_p} \le \int_{X_p} \phi^*_{p,n} d\nu_p. \tag{6.2.19}$$

Combining (6.2.17) and (6.2.19), we get

$$\frac{c_2}{\lambda_p} \le \int_{X_p} \phi^*_{p,n} d\nu_p \le c_5.$$

$$\text{i.e.} \quad \frac{c_2}{\lambda_p} \leq \int_{X_p} \left[\phi_{p,n}^*\right]_+ \, d\nu_p - \int_{X_p} \left[\phi_{p,n}^*\right]_- \, d\nu_p \leq c_5.$$

Thus, by (6.2.18), there is some constant c_6 such that

$$\int_{X_p} \left[\phi_{p,n}^*\right]_- \, d\nu_p \leq c_6. \tag{6.2.20}$$

Thus, we have that

$$\int_{X_p} |\phi_{p,n}^*| \, d\nu_p = \int_{X_p} \left[\phi_{p,n}^*\right]_+ \, d\nu_p + \int_{X_p} \left[\phi_{p,n}^*\right]_- \, d\nu_p \leq \max\{0, c_5\} + c_6.$$

Hence, $\sup_n ||\phi_{p,n}^*||_{L^1(X_p, d\nu_p)} < \infty$.

For simplicity of the proof, we will just use the uniform L^1-boundedness of $\phi_{i,n}^*$ for each $1 \leq i \leq p-1$. Later, we will improve this proof using the uniform boundedness of them.

Now, since for each $1 \leq i \leq p$, $(\phi_{i,n}^*)_{n\in\mathbb{N}}$ is a sequence with $\sup_n ||\phi_{i,n}^*||_{L^1(X_i, d\nu_i)} < \infty$, by Komlós theorem ([32]), there is a subsequence $(\phi_{i,n_m}^*)_{m\in\mathbb{N}}$ such that $\left(\frac{1}{N}\sum_{m=1}^N \phi_{i,n_m}^*\right)$ converges pointwise ν_i-a.e. to some ϕ_i^* as $N \to \infty$ and the same is true for any further subsequence of (ϕ_{i,n_m}^*).

Similarly, since for each $1 \leq i \leq p$, $\sup_n ||\psi_{i,n}^*||_{L^1(Y, d\nu^*)} < \infty$, for each $1 \leq i \leq p$, we can find a subsequence $(\psi_{i,n_m}^*)_{m\in\mathbb{N}}$ such that $\left(\frac{1}{N}\sum_{m=1}^N \psi_{i,n_m}^*\right)$ converges pointwise ν^*-a.e. to some ψ_i^* as $N \to \infty$ and the same is true for any further subsequence of (ψ_{i,n_m}^*).

Recall that we have

$$\sup_{\phi_i, \psi_i} \sum_{i=1}^p \lambda_i D(\phi_i, \psi_i) = \lim_{n\to\infty} \sum_{i=1}^p \lambda_i D(\phi_{i,n}, \psi_{i,n}) = \lim_{m\to\infty} \sum_{i=1}^p \lambda_i D(\phi_{i,n_m}, \psi_{i,n_m}).$$

Now, fix $\delta > 0$. Then, there exists an integer N_0 such that for each $m > N_0$,

$$\begin{aligned}
\sup_{\phi_i, \psi_i} \sum_{i=1}^p \lambda_i D(\phi_i, \psi_i) - \delta &\leq \sum_{i=1}^p \lambda_i D(\phi_{i,n_m}, \psi_{i,n_m}) \\
&\leq \sum_{i=1}^p \lambda_i D(\widetilde{\phi}_{i,n_m}, \psi_{i,n_m}) \qquad \text{(By (6.2.8))} \\
&= \sum_{i=1}^p \lambda_i D(\phi_{i,n_m}^*, \psi_{i,n_m}^*).
\end{aligned}$$

Now, consider the subsequences $\{\phi^*_{i,n_{N_0+m}}\}_{m\in\mathbb{N}}$ and $\{\psi^*_{i,n_{N_0+m}}\}_{m\in\mathbb{N}}$. Note that, the averages $\{\frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}}\}_{m\in\mathbb{N}}$ and $\{\frac{1}{N}\sum_{m=1}^N \psi^*_{i,n_{N_0+m}}\}_{m\in\mathbb{N}}$ also converge to ϕ^*, ν_i-a.e. and ψ^*, ν^*-a.e. respectively.

Then, for any $N \geq 1$,

$$\begin{aligned}
\sup_{\phi_i,\psi_i}\sum_{i=1}^p \lambda_i D(\phi_i,\psi_i) - \delta &\leq \sum_{i=1}^p \frac{1}{N}\sum_{m=1}^N \lambda_i D(\phi^*_{i,n_{N_0+m}}, \psi^*_{i,n_{N_0+m}}) \\
&\leq \sum_{i=1}^p \lambda_i D\left(\frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}}, \frac{1}{N}\sum_{m=1}^N \psi^*_{i,n_{N_0+m}}\right) \\
&= \sum_{i=1}^p \lambda_i \int_{X_i} \frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}}\ d\nu_i \\
&\quad - \varepsilon \sum_{i=1}^p \lambda_i \int_{X_i\times Y} e^{\frac{\frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}} + \frac{1}{N}\sum_{m=1}^N \psi^*_{i,n_{N_0+m}} - c_i}{\varepsilon}} d\nu_i \otimes \nu^*.
\end{aligned} \tag{6.2.21}$$

Note that, on the first line above, we take the sum over m from 1 to N and divide by N on both sides, and on the second line, we use the concavity of the functional D (see Appendix 5).

Then,

$$\begin{aligned}
\sup_{\phi_i,\psi_i}\sum_{i=1}^p \lambda_i D(\phi_i,\psi_i) - \delta &\leq \limsup_{N\to\infty}\Bigg\{\sum_{i=1}^p \lambda_i \int_{X_i} \frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}}\ d\nu_i \\
&\quad - \varepsilon \sum_{i=1}^p \lambda_i \int_{X_i\times Y} e^{\frac{\frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}} + \frac{1}{N}\sum_{m=1}^N \psi^*_{i,n_{N_0+m}} - c_i}{\varepsilon}} d\nu_i \otimes \nu^*\Bigg\} \\
&\leq \sum_{i=1}^p \lambda_i \limsup_{N\to\infty} \int_{X_i} \frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}}\ d\nu_i \\
&\quad - \varepsilon \sum_{i=1}^p \lambda_i \liminf_{N\to\infty} \int_{X_i\times Y} e^{\frac{\frac{1}{N}\sum_{m=1}^N \phi^*_{i,n_{N_0+m}} + \frac{1}{N}\sum_{m=1}^N \psi^*_{i,n_{N_0+m}} - c_i}{\varepsilon}} d\nu_i \otimes \nu^*.
\end{aligned} \tag{6.2.22}$$

Now, we will consider each of the limits in (6.2.22) above.

Note that by (6.2.16), we have $\sup_m \sup_{x\in X_i} \phi^*_{i,n_{N_0+m}}(x) < \infty$. Hence, by Fatou's Lemma (lim sup version),

$$\begin{aligned}\limsup_{N\to\infty}\int_{X_i}\frac{1}{N}\sum_{m=1}^{N}\phi^*_{i,n_{N_0+m}}\ d\nu_i &\le \int_{X_i}\limsup_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\phi^*_{i,n_{N_0+m}}\ d\nu_i\\ &= \int_{X_i}\phi_i^*\ d\nu_i.\end{aligned} \tag{6.2.23}$$

Since $\forall 1\le i\le p$, $\left(\frac{1}{N}\sum_{m=1}^{N}\phi^*_{i,n_{N_0+m}}+\frac{1}{N}\sum_{m=1}^{N}\psi^*_{i,n_{N_0+m}}-c_i\right)$ converges pointwise $\nu_i\otimes\nu^*$-a.e. to $(\phi_i^*+\psi_i^*-c_i)$, by Fatou's Lemma,

$$\int_{X_i\times Y} e^{\frac{\phi_i^*+\psi_i^*-c_i}{\varepsilon}}d\nu_i\otimes\nu^*\le\liminf_{N\to\infty}\int_{X_i\times Y} e^{\frac{\frac{1}{N}\sum_{m=1}^{N}\phi^*_{i,n_{N_0+m}}+\frac{1}{N}\sum_{m=1}^{N}\psi^*_{i,n_{N_0+m}}-c_i}{\varepsilon}}d\nu_i\otimes\nu^*. \tag{6.2.24}$$

Thus, by combining (6.2.22), (6.2.23) and (6.2.24), we get

$$\begin{aligned}\sup_{\phi_i,\psi_i}\sum_{i=1}^{p}\lambda_i D(\phi_i,\psi_i)-\delta &\le \sum_{i=1}^{p}\lambda_i\int_{X_i}\phi_i^*\ d\nu_i-\varepsilon\sum_{i=1}^{p}\lambda_i\int_{X_i\times Y}e^{\frac{\phi_i^*+\psi_i^*-c_i}{\varepsilon}}d\nu_i\otimes\nu^*\\ &=\sum_{i=1}^{p}\lambda_i D(\phi_i^*,\psi_i^*).\end{aligned} \tag{6.2.25}$$

Since $\delta>0$ is arbitrary, letting $\delta\to 0$, we get that

$$\sup_{\phi_i,\psi_i}\sum_{i=1}^{p}\lambda_i D(\phi_i,\psi_i)\le\sum_{i=1}^{p}\lambda_i D(\phi_i^*,\psi_i^*).$$

Also, since $\sum_{i=1}^{p}\lambda_i\psi^*_{i,n_m}=0$, we have that

$$\lim_{N\to\infty}\sum_{i=1}^{p}\lambda_i\frac{1}{N}\sum_{m=1}^{N}\psi^*_{i,n_m}=\sum_{i=1}^{p}\lambda_i\psi_i^*=0\quad \nu^*\text{-a.e.}$$

Thus we can conclude that $\{\phi_i^*,\psi_i^*\}_{i=1}^{p}$ is a maximizer for (6.2.6). ■

Remark 6.2.6. *Note that, since N_0 depends on δ, $\forall 1\le i\le p$, the sets*

$$\mathcal{A}_{\phi_i}=\left\{x_i\in X_i:\{\phi^*_{i,n_{N_0+m}}\}\text{ does not converge to }\phi_i^*\right\}$$

and

$$\mathcal{A}_{\psi_i}=\left\{y\in Y:\{\psi^*_{i,n_{N_0+m}}\}\text{ does not converge to }\psi_i^*\right\}$$

depend on the choice of δ. However, since we only consider integrals against ν_i's and ν^*, these sets do not affect our calculations.

Remark 6.2.7. *Also observe that, even though we have strong duality results for* $\{(\phi_i \in C_b(X_i,\nu_i), \psi_i \in C_b(Y,\nu^*))\}_{i=1}^p$*, we get existence for* $\{(\phi_i \in L^1(X_i,\nu_i), \psi_i \in L^1(Y,\nu^*))\}_{i=1}^p$*. We may get a better regularity for* $\phi_i^*, \forall 1 \leq i \leq p-1$*, due to the uniform boundedness of the* $\phi_{i,n}^*, \forall 1 \leq i \leq p-1$ *(see* (6.2.9)*).*

BIBLIOGRAPHY

[1] M. Agueh and G. Carlier. "Barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924.

[2] L. Ambrosio, A. Bressan, D. Helbing, A. Klar, E. Zuazua, L. Ambrosio, and N. Gigli. "A user's guide to optimal transport". In: *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle* (2013), pp. 1–155.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.

[4] E. Anderes, S. Borgwardt, and J. Miller. "Discrete Wasserstein barycenters: Optimal transport for discrete data". In: *Mathematical Methods of Operations Research* 84.2 (2016), pp. 389–409.

[5] M. Beiglböck and W. Schachermayer. "Duality for Borel measurable cost functions". In: *Transactions of the American Mathematical Society* 363.8 (2011), pp. 4203–4224.

[6] J.-D. Benamou and Y. Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numerische Mathematik* 84.3 (2000), pp. 375–393.

[7] J. Bigot, E. Cazelles, and N. Papadakis. "Penalization of barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 51.3 (2019), pp. 2261–2285.

[8] S. Borgwardt and S. Patterson. "Improved linear programs for discrete barycenters". In: *Informs Journal on Optimization* 2.1 (2020), pp. 14–33.

[9] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum. "Entropy minimization, DAD problems, and doubly stochastic kernels". In: *Journal of Functional Analysis* 123.2 (1994), pp. 264–307.

[10] Y. Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808.

[11] Y. Brenier. "Optimal transportation of particles, fluids and currents". In: *Variational methods for evolving objects*. Vol. 67. Mathematical Society of Japan, 2015, pp. 59–86.

[12] G. Carlier, K. Eichinger, and A. Kroshnin. "Entropic-Wasserstein barycenters: PDE characterization, regularity, and CLT". in: *SIAM Journal on Mathematical Analysis* 53.5 (2021), pp. 5880–5914.

[13] G. Carlier and I. Ekeland. "Matching for teams". In: *Economic theory* 42.2 (2010),

pp. 397–418.

[14] G. Carlier and M. Laborde. “A differential approach to the multi-marginal Schrödinger system”. In: *SIAM Journal on Mathematical Analysis* 52.1 (2020), pp. 709–717.

[15] I. Csiszár. “I-divergence geometry of probability distributions and minimization problems”. In: *The annals of probability* (1975), pp. 146–158.

[16] M. Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: *Advances in neural information processing systems* 26 (2013).

[17] S. Di Marino and A. Gerolin. “Optimal transport losses and Sinkhorn algorithm with general convex regularization”. In: *arXiv preprint arXiv:2007.00976* (2020).

[18] A. N. Doledenok. “On Kantorovich multimarginal optimal transportation problems with density constraints”. In: 2018.

[19] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011.

[20] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1999.

[21] J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouvé, and G. Peyré. “Interpolating between optimal transport and mmd using sinkhorn divergences”. In: *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR. 2019, pp. 2681–2690.

[22] W. Gangbo and R. J. McCann. “Optimal maps in Monge’s mass transport problem”. In: *Comptes Rendus de l’Academie des Sciences-Serie I-Mathematique* 321.12 (1995), p. 1653.

[23] W. Gangbo and R. J. McCann. “The geometry of optimal transportation”. In: *Acta Math.* 177.2 (1996), pp. 113–161.

[24] W. Gangbo and A. Święch. “Optimal maps for the multidimensional Monge-Kantorovich problem”. In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 51.1 (1998), pp. 23–45.

[25] Z. Ge, S. Liu, Z. Li, O. Yoshie, and J. Sun. “Ota: Optimal transport assignment for object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 303–312.

[26] T. Geiger, D. Wachsmuth, and G. Wachsmuth. “Optimal control of ODEs with state suprema”. In: *Math. Control Relat. Fields* 11.3 (2021), pp. 555–578.

[27] A. Genevay, M. Cuturi, G. Peyré, and F. Bach. "Stochastic optimization for large-scale optimal transport". In: *Advances in neural information processing systems* 29 (2016).

[28] A. Gerolin, A. Kausamo, and T. Rajala. "Multi-marginal entropy-transport with repulsive cost". In: *Calculus of Variations and Partial Differential Equations* 59.3 (2020), p. 90.

[29] N. Gigli and L. Tamanini. "Second order differentiation formula on RCD (K, N) spaces". In: *Rendiconti Lincei* 29.2 (2018), pp. 377–386.

[30] L. V. Kantorovich. "On the translocation of masses". In: *Dokl. Akad. Nauk. USSR (NS)*. vol. 37. 1942, pp. 199–201.

[31] J. Kitagawa and B. Pass. "The multi-marginal optimal partial transport problem". In: *Forum of Mathematics, Sigma*. Vol. 3. Cambridge University Press. 2015.

[32] J. Komlós. "A generalization of a problem of Steinhaus". In: *Acta Mathematica Academiae Scientiarum Hungaricae* 18.1-2 (1967), pp. 217–229.

[33] J. Korman and R. McCann. "Optimal transportation with capacity constraints". In: *Transactions of the American Mathematical Society* 367.3 (2015), pp. 1501–1521.

[34] J. Korman, R. J. McCann, and C. Seis. "Dual potentials for capacity constrained optimal transport". In: *Calculus of Variations and Partial Differential Equations* 54.1 (2015), pp. 573–584.

[35] J. Korman, R. J. McCann, and C. Seis. "An elementary approach to linear programming duality with application to capacity constrained transport". In: *J. Convex Anal.* 22.3 (2015), pp. 797–808.

[36] H. Lavenant, S. Claici, E. Chien, and J. Solomon. "Dynamical optimal transport on discrete surfaces". In: *ACM Transactions on Graphics (TOG)* 37.6 (2018), pp. 1–16.

[37] C. Léonard. "A survey of the Schrödinger problem and some of its connections with optimal transport". In: *Discrete Contin. Dyn. Syst.* 34.4 (2014), pp. 1533–1574.

[38] L. Li, A. Genevay, M. Yurochkin, and J. M. Solomon. "Continuous regularized wasserstein barycenters". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 17755–17765.

[39] S. D. Marino and A. Gerolin. "An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm". In: *Journal of Scientific Computing* 85.2 (2020), pp. 1–28.

[40] R. J. McCann. "A convexity principle for interacting gases". In: *Advances in mathe-*

matics 128.1 (1997), pp. 153–179.

[41] G. Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Mem. Math. Phys. Acad. Royale Sci.* (1781), pp. 666–704.

[42] N. Papadakis. "Transport Optimal pour le Traitement d'Images". In: 2015.

[43] B. Pass. "Multi-marginal optimal transport: theory and applications". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 49.6 (2015), pp. 1771–1790.

[44] B. Pass. "On the local structure of optimal measures in the multi-marginal optimal transportation problem". In: *Calculus of Variations and Partial Differential Equations* 43.3 (2012), pp. 529–536.

[45] G. Peyré, M. Cuturi, et al. "Computational optimal transport: With applications to data science". In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607.

[46] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory.* Vol. 1. Springer Science & Business Media, 1998.

[47] R. T. Rockafellar. *Convex analysis, 2. printing ed.* 1972.

[48] H. L. Royden and P. Fitzpatrick. *Real analysis.* Vol. 32. Macmillan New York, 1988.

[49] L. Ruschendorf. "Convergence of the iterative proportional fitting procedure". In: *The Annals of Statistics* (1995), pp. 1160–1174.

[50] T. Salimans, H. Zhang, A. Radford, and D. Metaxas. "Improving GANs using optimal transport". In: *arXiv preprint arXiv:1803.05573* (2018).

[51] F. Santambrogio. "Optimal transport for applied mathematicians". In: *Birkäuser, NY* 55.58-63 (2015), p. 94.

[52] D. Simon and A. Aberdam. "Barycenters of natural images constrained wasserstein barycenters for image morphing". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2020, pp. 7910–7919.

[53] C. Villani. *Optimal transport: old and new.* Vol. 338. Springer, 2009.

[54] C. Villani. *Topics in optimal transportation.* Vol. 58. American Mathematical Soc., 2021.

[55] P. Zhao and Z.-H. Zhou. "Label distribution learning by optimal transport". In: *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 32. 1. 2018.

[56] L. Zhu, Y. Yang, S. Haker, and A. Tannenbaum. "An image morphing technique based on optimal mass preserving mapping". In: *IEEE transactions on image processing* 16.6 (2007), pp. 1481–1495.

APPENDIX

1 Alternative dual functional for CCOT Problem:

Let $X, Y \subseteq \mathbb{R}^d$. For given $f \in L^1(X), g \in L^1(Y)$ and $0 \le \tilde{h} \in L^\infty(X \times Y)$, consider

$$\mathrm{OT}^*_{\mathrm{CC}} := \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{h}}} J(u,v,w) \tag{1.1}$$

where

$$J(u,v,w) := \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy + \int_{X\times Y} w(x,y)\tilde{h}(x,y)\ dxdy, \tag{1.2}$$

$$\begin{aligned}\mathrm{Lip}_{c,\tilde{h}} := \Big\{&(u,v,w) : u \in L^1(X, fdx), v \in L^1(Y, gdy), w \in L^1(X\times Y, \tilde{h}dxdy),\\ &u(x)+v(y)+w(x,y) \le c(x,y), \text{and } w(x,y) \le 0\Big\},\end{aligned} \tag{1.3}$$

and

$$\mathrm{OT}^{*'}_{\mathrm{CC}} := \sup_{u\in L^1(X), v\in L^1(Y)} J'(u,v), \tag{1.4}$$

where

$$J'(u,v) := \int_X uf\ dx + \int_Y vg\ dy - \int_{X\times Y} [-c+u+v]_+\,\tilde{h}\ dxdy. \tag{1.5}$$

Proposition 1.1. $\mathrm{OT}^*_{CC} = \mathrm{OT}^{*'}_{CC}$.

Proof. Define $\mathcal{A} = \Big\{(u,v,w) : u \in L^1(X, fdx), v \in L^1(Y, gdy), \text{and } w = -[-c+u+v]_+\Big\}$.

Then,

$$\mathrm{OT}^{*'}_{\mathrm{CC}} := \sup_{(u,v,w)\in\mathcal{A}} J(u,v,w).$$

First, let $(u,v,w) \in \mathcal{A}$ be arbitrary. Then $w \le 0$, $w = -[-c+u+v]_+ \le -[-c+u+v]$, which gives $u+v+w \le c$, and $w \in L^1(X\times Y, \tilde{h}dxdy)$. Hence, $(u,v,w) \in \mathrm{Lip}_{c,\tilde{h}}$. Thus, $\mathcal{A} \subseteq \mathrm{Lip}_{c,\tilde{h}}$. So, we have

$$\sup_{(u,v,w)\in\mathcal{A}} J(u,v,w) \le \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{h}}} J(u,v,w), \tag{1.6}$$

i.e.

$$\mathrm{OT}^{*'}_{\mathrm{CC}} \le \mathrm{OT}^*_{\mathrm{CC}}. \tag{1.7}$$

Now observe that,

$$\mathrm{OT}^*_{\mathrm{CC}} = \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{h}}} \left\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy + \int_{X\times Y} w(x,y)\tilde{h}(x,y)\ dxdy \right\}$$

$$
\begin{aligned}
&= \sup_{(u,v,w)\in \mathrm{Lip}_{c,\tilde{h}}} \Bigg\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy \\
&\quad + \int_{\{(x,y):-c(x,y)+u(x)+v(y)\geq 0\}} w(x,y)\tilde{h}(x,y)\ dxdy \Bigg\} \\
&\quad + \int_{\{(x,y):-c(x,y)+u(x)+v(y)< 0\}} w(x,y)\tilde{h}(x,y)\ dxdy \Bigg\} \\
&\leq \sup_{(u,v)\in L^1(X)\times L^1(Y)} \Bigg\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy \\
&\quad + \int_{\{(x,y):-c(x,y)+u(x)+v(y)\geq 0\}} [c(x,y)-u(x)-v(y)]\tilde{h}(x,y)\ dxdy \Bigg\}.
\end{aligned} \tag{1.8}
$$

In the last line above, we used the fact that $w \leq c-u-v$, and $w \leq 0$.

On the other hand,

$$
\begin{aligned}
\mathrm{OT}^{*'}_{\mathrm{CC}} &= \sup_{(u,v)\in L^1(X)\times L^1(Y)} \left\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy - \int_{X\times Y} [-c+u+v]_+\, \tilde{h}\ dxdy \right\} \\
&= \sup_{(u,v)\in L^1(X)\times L^1(Y)} \Bigg\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy \\
&\quad - \int_{\{(x,y):-c(x,y)+u(x)+v(y)\geq 0\}} [-c+u+v]_+\, \tilde{h}(x,y)\ dxdy \\
&\quad - \int_{\{(x,y):-c(x,y)+u(x)+v(y)< 0\}} [-c+u+v]_+\, \tilde{h}(x,y)\ dxdy \Bigg\} \\
&= \sup_{(u,v)\in L^1(X)\times L^1(Y)} \Bigg\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy \\
&\quad - \int_{\{(x,y):-c(x,y)+u(x)+v(y)\geq 0\}} [-c+u+v]\, \tilde{h}(x,y)\ dxdy \Bigg\} \\
&= \sup_{(u,v)\in L^1(X)\times L^1(Y)} \Bigg\{ \int_X u(x)f(x)\ dx + \int_Y v(y)g(y)\ dy \\
&\quad + \int_{\{(x,y):-c(x,y)+u(x)+v(y)\geq 0\}} [c(x,y)-u(x)-v(y)]\, \tilde{h}(x,y)\ dxdy \Bigg\}.
\end{aligned} \tag{1.9}
$$

By combining (1.8) and (1.9), we get

$$
\mathrm{OT}^{*}_{\mathrm{CC}} \leq \mathrm{OT}^{*'}_{\mathrm{CC}}. \tag{1.10}
$$

The inequalities (1.7) and (1.10) conclude that

$$
\mathrm{OT}^{*}_{\mathrm{CC}} = \mathrm{OT}^{*'}_{\mathrm{CC}}.
$$

■

2 The relaxed CCOT Problem is strictly convex.

Given an integer $p > 2$, a lower semi-continuous function $c : \mathbb{R}^{pd} \mapsto \mathbb{R}$, $f \in L^1(\mathbb{R}^d)$ and $h \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$, consider the functional

$$I_c^\varepsilon(h) := \int_{\mathbb{R}^{pd}} ch \; d\hat{x} + \frac{1}{2\varepsilon} \sum_{i=1}^{p} || \langle h \rangle_{x_i} - f_i ||_2^2, \tag{2.1}$$

where $d\hat{x} = dx_1 \ldots dx_p$.

Proposition 2.1. *I_c^ε is strictly convex.*

Proof. Let $0 \leq h_1, h_2 \leq \tilde{h}$ be such that $h_1 \neq h_2$ and let $0 < \lambda < 1$.

Then, observe that

$$\begin{aligned}
I_c^\varepsilon(\lambda h_1 + (1-\lambda)h_2) &= \int_{\mathbb{R}^{pd}} c(\lambda h_1 + (1-\lambda)h_2) \mathrm{d}\hat{x} + \frac{1}{2\varepsilon} \sum_{i=1}^{p} || \langle (\lambda h_1 + (1-\lambda)h_2) \rangle_{x_i} - f_i ||_2^2 \\
&= \lambda \int_{\mathbb{R}^{pd}} ch_1 \; d\hat{x} + (1-\lambda) \int_{\mathbb{R}^{pd}} ch_2 \; d\hat{x} \\
&\quad + \frac{1}{2\varepsilon} \sum_{i=1}^{p} || \langle (\lambda h_1 + (1-\lambda)h_2) \rangle_{x_i} - (\lambda + (1-\lambda)) f_i ||_2^2 \\
&= \lambda \int_{\mathbb{R}^{pd}} ch_1 \; d\hat{x} + (1-\lambda) \int_{\mathbb{R}^{pd}} ch_2 \; d\hat{x} \\
&\quad + \frac{1}{2\varepsilon} \sum_{i=1}^{p} ||\lambda(\langle h_1 \rangle_{x_i} - f_i) + (1-\lambda)(\langle h_2 \rangle_{x_i} - f_i)||_2^2 \\
&< \lambda \int_{\mathbb{R}^{pd}} ch_1 \; d\hat{x} + (1-\lambda) \int_{\mathbb{R}^{pd}} ch_2 \; d\hat{x} \\
&\quad + \frac{\lambda}{2\varepsilon} \sum_{i=1}^{p} || \langle h_1 \rangle_{x_i} - f_i ||_2^2 + \frac{(1-\lambda)}{2\varepsilon} \sum_{i=1}^{p} || \langle h_2 \rangle_{x_i} - f_i ||_2^2 \\
&= \lambda I_c^\varepsilon(h_1) + (1-\lambda) I_c^\varepsilon(h_2).
\end{aligned}$$

Note that in the penultimate line, the strict inequality holds due to the strict convexity of the L^2 norm. ■

3 Uniform boundedness for probability measures with finite moment.

Proposition 3.1. *Given capacities $\{\tilde{\gamma}_i\}_{i=1}^p \subseteq \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)$, p probability measures $\nu_1, \ldots, \nu_p$ in $\mathcal{P}_2(\mathbb{R}^d)$, and a sequence of probability measures $\{\tilde{\nu}_n\}_{n \in \mathbb{N}} \in \mathcal{P}_2(\mathbb{R}^d)$ such that $\Pi^{\tilde{\gamma}_i}(\nu_i, \tilde{\nu}_n) \neq \emptyset, \forall 1 \leq i \leq p$, if there is an $M > 0$ such that $\widetilde{W_2}^2(\nu_i, \tilde{\nu}_n) \leq M, \forall i$ and n, we have*

$$\sup_n \int |x|^2 \; d\tilde{\nu}_n \leq C, \qquad \textit{for some constant } C.$$

Proof. Fix $i \in \{1, \ldots, p\}$ and $n \in \mathbb{N}$. Consider the two-marginal capacity constrained OT problem between the measures ν_i and $\tilde{\nu}_n$ with capacity $\tilde{\gamma}_i$ given by (5.2.3).

By the duality,

$$\widetilde{W_2}^2(\nu_i, \tilde{\nu}_n) = \sup_{\mathcal{B}_{i,n}} \left\{ \int_{\mathbb{R}^d} u(x)\ d\nu_i(x) + \int_{\mathbb{R}^d} v(y)\ d\tilde{\nu}_n(y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} w(x,y)\ d\tilde{\gamma}_i(x,y) \right\}, \tag{3.1}$$

where the supremum is taken over the set $\mathcal{B}_{i,n}$ of real-valued functions u, v, w satisfying $u \in L^1(\nu_i), v \in L^1(\tilde{\nu}_n)$, $w \in L^1(\tilde{\gamma}_i)$ and $w \leq 0$ with $u(x) + v(y) + w(x,y) \leq \frac{1}{2}|x-y|^2$.

Choose $u(x) = \inf_{y\in\mathbb{R}^d}\{\frac{1}{2}|x|^2 + \frac{1}{4}|y|^2 - x\cdot y\}, v(y) = \frac{1}{4}|y|^2, w(x,y) = 0$.

Then,

$u(x) + v(y) + w(x,y) \leq \frac{1}{2}|x|^2 + \frac{1}{4}|y|^2 - x\cdot y + \frac{1}{4}|y|^2 = \frac{1}{2}|x-y|^2$ for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$.

By plugging this choice of u, v, w in (3.1), we get

$$\int_{\mathbb{R}^d} \inf_{y\in\mathbb{R}^d} \left\{\frac{1}{2}|x|^2 + \frac{1}{4}|y|^2 - x\cdot y\right\} d\nu_i(x) + \int_{\mathbb{R}^d} \frac{1}{4}|y|^2\ d\tilde{\nu}_n(y) \leq \widetilde{W_2}^2(\nu_i, \tilde{\nu}_n). \tag{3.2}$$

Since $\nabla_y\left\{\frac{1}{2}|x|^2 + \frac{1}{4}|y|^2 - x\cdot y\right\} = \frac{1}{2}y - x$, we have

$\inf_{y\in\mathbb{R}^d}\left\{\frac{1}{2}|x|^2 + \frac{1}{4}|y|^2 - x\cdot y\right\} = -\frac{1}{2}|x|^2$.

Thus, (3.2) gives us,

$$\int_{\mathbb{R}^d} \frac{1}{4}|y|^2\ d\tilde{\nu}_n(y) \leq \widetilde{W_2}^2(\nu_i, \tilde{\nu}_n) + \int_{\mathbb{R}^d} \frac{1}{2}|x|^2\ d\nu_i(x).$$

Since $\widetilde{W_2}^2(\nu_i, \tilde{\nu}_n) \leq M$ and $\nu_i \in \mathcal{P}_2(\mathbb{R}^d)$, there is a constant $C > 0$, independent from n such that

$$\int_{\mathbb{R}^d} \frac{1}{4}|y|^2\ d\tilde{\nu}_n(y) \leq C.$$

■

4 An integral condition for tightness:

Proposition 4.1. *([3], Remark 5.1.5) Let $X \subseteq \mathbb{R}^d$ and $\{\nu_n\}_{n\in\mathbb{N}} \in \mathcal{P}_2(X)$ be a sequence of probability measures. Then, if $\sup_n \int |x|^2\ d\nu_n < +\infty$, then $\{\nu_n\}_{n\in\mathbb{N}}$ is tight.*

Proof. Let $\delta > 0$ be arbitrary and fix $n \in \mathbb{N}$. Suppose $\sup_n \int |x|^2\ d\nu_n < M$. Let R_δ be a real number such that $R_\delta > \sqrt{\frac{M}{\delta}}$. Define the compact set $K_\delta = \{x \in \mathbb{R}^d : |x| \leq R_\delta\}$. Then,

$$\nu_n(\{x \in \mathbb{R}^d : |x| > R_\delta\}) \leq \frac{1}{R_\delta^2} \int_{\mathbb{R}^d} |x|^2\ d\nu_n \leq \frac{M}{R_\delta^2} < \delta. \tag{4.1}$$

This proves that $\{\nu_n\}_{n\in\mathbb{N}}$ is tight.

Note that we get the first inequality in (4.1) by the Chebychev's inequality. ■

5 The entropic dual functional is strictly concave.

Given two Polish spaces X, Y, a bounded cost function $c : X \times Y \mapsto \mathbb{R}$, two probability measures $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(\nu)$, and two functions $\phi \in L^{\exp}_{\varepsilon}(X, d\mu), \psi \in L^{\exp}_{\varepsilon}(Y, \mathrm{d}\nu)$, consider the functional

$$D(\phi, \psi) := \int_X \phi d\mu + \int_Y \psi d\nu - \varepsilon \int_{X\times Y} e^{\frac{(\phi+\psi-c)}{\varepsilon}} d\mu \otimes \nu. \tag{5.1}$$

Proposition 5.1. *The functional D is strictly concave.*

Proof. Let $\phi_1, \phi_2 \in L^{\exp}_{\varepsilon}(X, d\mu)$, $\psi_1, \psi_2 \in L^{\exp}_{\varepsilon}(Y, \mathrm{d}\nu)$ and $0 < \lambda < 1$ be a real number. We need to show that

$$D(\lambda\phi_1 + (1-\lambda)\phi_2, \lambda\psi_1 + (1-\lambda)\psi_2) > \lambda D(\phi_1, \psi_1) + (1-\lambda)D(\phi_2, \psi_2).$$

i.e.

$$\begin{aligned}\lambda \int_X \phi_1\ d\mu + (1-\lambda)\int_X \phi_2\ d\mu + \lambda \int_Y \psi_1\ d\nu + (1-\lambda)\int_Y \psi_2\ d\nu \\ - \varepsilon \int_{X\times Y} e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu \\ > \lambda \int_X \phi_1\ d\mu + \lambda\int_Y \psi_1\ d\nu - \varepsilon\lambda \int_{X\times Y} e^{\frac{(\phi_1+\psi_1-c)}{\varepsilon}}\ d\mu\otimes\nu \\ + (1-\lambda)\int_X \phi_2\ d\mu + (1-\lambda)\int_Y \psi_2\ d\nu - \varepsilon(1-\lambda)\int_{X\times Y} e^{\frac{(\phi_2+\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu.\end{aligned}$$

So, it is sufficient to show that

$$\begin{aligned}- \varepsilon \int_{X\times Y} e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu \\ > -\varepsilon\lambda \int_{X\times Y} e^{\frac{(\phi_1+\psi_1-c)}{\varepsilon}}\ d\mu\otimes\nu - \varepsilon(1-\lambda)\int_{X\times Y} e^{\frac{(\phi_2+\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu.\end{aligned}$$

i.e.

$$\begin{aligned}\int_{X\times Y} e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu \\ < \lambda \int_{X\times Y} e^{\frac{(\phi_1+\psi_1-c)}{\varepsilon}}\ d\mu\otimes\nu + (1-\lambda)\int_{X\times Y} e^{\frac{(\phi_2+\psi_2-c)}{\varepsilon}}\ d\mu\otimes\nu.\end{aligned} \tag{5.2}$$

First, we will prove that

$$e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-c)}{\varepsilon}} < \lambda e^{\frac{(\phi_1+\psi_1-c)}{\varepsilon}} + (1-\lambda)e^{\frac{(\phi_2+\psi_2-c)}{\varepsilon}}. \tag{5.3}$$

Observe that

$$\begin{aligned}e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-c)}{\varepsilon}} &= e^{\frac{(\lambda\phi_1+(1-\lambda)\phi_2+\lambda\psi_1+(1-\lambda)\psi_2-\lambda c-(1-\lambda)c)}{\varepsilon}} \\ &= e^{\lambda\frac{(\phi_1+\psi_1-c)}{\varepsilon}+(1-\lambda)\frac{(\phi_2+\psi_2-c)}{\varepsilon}}\end{aligned}$$

$$< \lambda e^{\frac{(\phi_1+\psi_1-c)}{\varepsilon}} + (1-\lambda)e^{\frac{(\phi_2+\psi_2-c)}{\varepsilon}}.$$

Note that the last line above holds due to the strict convexity of the exponential function. Now, by integrating both sides of (5.3) with respect to $d\mu\otimes\nu$ on $X\times Y$, we get the inequality (5.2).

This completes the proof.

■