MULTI-STAGE OPTIMAL TRANSPORT

ABIHITH KOTHAPALLI

Abstract. Optimal transport (OT) provides a powerful mathematical framework for comparing probability distributions and modeling the movement of mass under cost constraints. In this thesis, we investigate a generalization of the classical OT problem in which mass must be transported through one or more intermediate stages, a problem which we refer to as multi-stage optimal transport (MSOT). We begin by formulating the discrete MSOT problem and demonstrate a reduction to the classical Kantorovich problem through a reduced cost matrix, allowing for efficient computation of solutions. We subsequently propose an algorithm with improved complexity guarantees compared to direct linear programming approaches. In the continuous setting, we explore the connections between MSOT and Wasserstein barycenters, and establish sufficient conditions for the existence of Monge mappings and Kantorovich transport plans. Our findings unify and extend existing OT theory, offering both theoretical insight and computational tools for transport problems involving multiple stages of transportation.

Contents

1. Introduction

Optimal transport (OT) is a fundamental problem in mathematics with wide-ranging applications in economics, machine learning, computer vision, and beyond. At its core, OT concerns how to most efficiently move mass from one probability distribution to another, given a cost of transportation. Originally formulated by Monge in the 18th century [Mon81] and rigorously developed in the 20th century by Kantorovich [Kan06a, Kan06b], OT

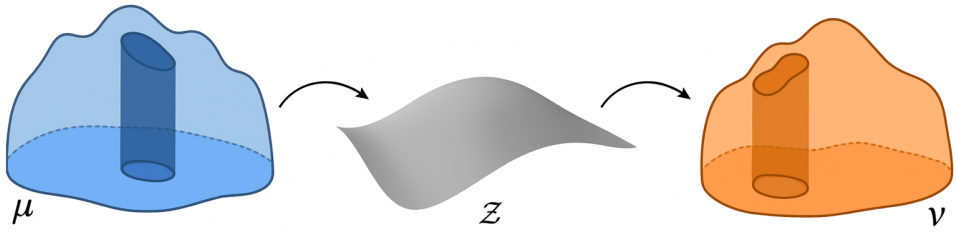

FIGURE 1. Illustration of the MSOT problem.

provides a rich mathematical framework for comparing distributions and modeling the movement of mass across space.

This thesis explores a generalization of the classical OT problem where mass must pass through one or more intermediate stages before reaching its destination. We call this generalized problem the multi-stage OT (MSOT) problem. Rather than transporting mass directly from a source measure μ to a target measure ν, we consider a setting where mass must first be transported from μ to an intermediate Borel space $\mathcal{Z}$, and then from this $\mathcal{Z}$ to ν. This framework reflects real-world transportation processes such as supply chains, multi-hop networks, or any scenario where direct movement is restricted or inefficient.

In this thesis, we develop a comprehensive theory of the MSOT problem. In the discrete setting, we prove that MSOT admits an exact reduction to an instance of the classical OT problem via a minimal-path cost formulation, and we use this result to derive a computationally efficient algorithm for computing optimal transport plans. We analyze the complexity of this algorithm and demonstrate its improvement over direct approaches based on linear programming. We extend our framework to handle an arbitrary number of intermediate stages, and we discuss settings where Monge-type solutions exist. In the continuous setting, we establish connections between MSOT and Wasserstein barycenters, and we prove existence and uniqueness results for both the Monge and Kantorovich MSOT formulations under appropriate assumptions. Finally, we show that under mild regularity assumptions, the MSOT problem can be reduced to an instance of a single-stage classical OT problem with a suitably defined cost function, thereby unifying our discrete and continuous analyses within a broader OT framework.

2. NOTATIONS

We adopt the following notations throughout the paper.

- $[\![m]\!]$: the set of integers $\{1, \dots, m\}$.
- $\mathbb{1}_m$: m-dimensional vector with all entries identically equal to 1.
- $\langle \cdot, \cdot \rangle$: the usual Euclidean dot-product between vectors, or the Frobenius dot-product for two matrices of the same size.

- $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$: Borel spaces. Typically, $\mathcal{X}$ will be the source space, $\mathcal{Y}$ the target space, and $\mathcal{Z}$ an intermediate space. If there are multiple intermediate spaces, we label them $\mathcal{Z}_1, \dots, \mathcal{Z}_N$. In the discrete setting, $\mathcal{X} = \{x_i\}_{i=1}^m$, $\mathcal{Y} = \{y_j\}_{j=1}^n$, $\mathcal{Z} = \{z_k\}_{k=1}^p$.
- $\mathcal{P}(\mathcal{X})$: the space of Borel probability measures on $\mathcal{X}$. $\mathcal{P}_p(\mathcal{X})$ denotes the space of Borel probability measures with finite pth moment.
- Σ_m: probability simplex with m bins. That is, the set of probability vectors in $\mathbb{R}_+^m$, $\Sigma_m := \left\{\mathbf{a} \in \mathbb{R}_+^m : \|\mathbf{a}\|_1 = 1\right\}$.
- μ, ν: (probability) measures supported on $\mathcal{X}, \mathcal{Y}$. In the discrete setting, $\mu = \sum_{i=1}^m a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^n b_j \delta_{y_j}$.
- $\mathbf{a}, \mathbf{b}$: histograms in the simplices Σ_m, Σ_n. We let $\mathbf{a} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T$, $\mathbf{b} = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T$.
- $\otimes$: product measure or vector outer product. That is, for measures μ and ν, $\mu \otimes \nu$ is their product measure, whereas for histograms $\mathbf{a}, \mathbf{b}$ (or more generally, Euclidean vectors), $\mathbf{a} \otimes \mathbf{b}$ is their outer product.
- $\pi_{\mathcal{X}}$: canonical projection mapping from product space onto $\mathcal{X}$. By abuse of notation, we use $\pi_{\mathcal{X}}$ to refer to both the maps $\mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ given by $(x, y) \mapsto x$ and $\mathcal{X} \times \mathcal{Z} \to \mathcal{Z}$ given by $(x, z) \mapsto x$ depending on context. Other canonical projection maps are denoted similarly.
- $T_{\#}\mu$: the push-forward measure of μ through the map $T \colon \mathcal{X} \to \mathcal{Y}$. The push-forward measure is defined such that $T_{\#}\mu(A) = \mu(T^{-1}(A))$.
- $\Gamma(\mu, \nu)$: the set of transport plans/couplings between μ, ν. That is, the set $\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : (\pi_{\mathcal{X}})_{\#}\gamma = \mu, (\pi_{\mathcal{Y}})_{\#}\gamma = \nu\}$. When transporting through an intermediate space $\mathcal{Z}$, we denote the set of two-step transport plans as $\Gamma_{\mathcal{Z}}(\mu, \nu)$ (see Equation (7.4)).
- $U(\mathbf{a}, \mathbf{b})$: the set of transport plans/couplings between histograms $\mathbf{a}, \mathbf{b}$. That is, the set $U(\mathbf{a}, \mathbf{b}) := \{P \in \mathbb{R}_+^{m \times n} : P\mathbb{1}_p = \mathbf{a}, P^T\mathbb{1}_m = \mathbf{b}\}$. When transporting through an intermediate space $\mathcal{Z}$, we denote the set of two-step transport plans as $U_{\mathcal{Z}}(\mathbf{a}, \mathbf{b})$ (see Equation (5.4)).
- $c(x, y)$: transportation cost between x and y. In the MSOT problem, $c_1(x, z)$ is the transportation cost between x and z, and $c_2(z, y)$ is the transportation cost between z and y. In the discrete setting, $C = \left[c(x_i, y_j)\right]_{i,j} \in \mathbb{R}_+^{m \times n}$, and in the discrete MSOT problem, C_1, C_2 are defined similarly.
- $M(\mu, \nu; c)$: the Monge problem between μ and ν given the cost function $c(\cdot, \cdot)$. In the discrete setting, $M(\mathbf{a}, \mathbf{b}; C)$ denotes the Monge problem between histograms $\mathbf{a}$ and $\mathbf{b}$ given cost matrix C. The Monge formulation of the MSOT problem with intermediate space $\mathcal{Z}$ is similarly denoted by $M_{\mathcal{Z}}(\mu, \nu; c)$ and $M_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C)$.
- $K(\mu, \nu; c)$: the Kantorovich problem between μ and ν given the cost function $c(\cdot, \cdot)$. In the discrete setting, $K(\mathbf{a}, \mathbf{b}; C)$ denotes the Kantorovich problem between histograms $\mathbf{a}$ and $\mathbf{b}$ given cost matrix C. The Kantorovich formulation of the MSOT problem with intermediate space $\mathcal{Z}$ is similarly denoted by $K_{\mathcal{Z}}(\mu, \nu; c)$ and $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C)$.

3. Monge and Kantorovich Problems

OT first originated from Gaspard Monge's formulation in the 18th century [Mon81], which was motivated by the problem of moving the mass from a pile of dirt or sand to construct a target structure in the most cost efficient manner. Formally, let $\mathcal{X}, \mathcal{Y}$ be two Borel spaces, and consider the probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Then, given a cost function $c\colon \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$, the Monge problem is given as

$$M(\mu, \nu; c) = \inf_{T\colon \mathcal{X} \to \mathcal{Y}} \left\{ \int_{\mathcal{X}} c(x, T(x)) \, \mathrm{d}\mu(x) : T_{\#}\mu = \nu \right\}, \tag{3.1}$$

where $T\colon \mathcal{X} \to \mathcal{Y}$ can be any Borel measurable map. Here, the notation $T_{\#}\mu$ indicates the push-forward of μ through the map T, which is defined such that $T_{\#}\mu(A) = \mu(T^{-1}(A))$. Equivalently, $T_{\#}\mu$ is defined such that

$$\int_{\mathcal{Y}} \psi \, \mathrm{d}(T_{\#}\mu) = \int_{\mathcal{X}} (\psi \circ T) \, \mathrm{d}\mu \tag{3.2}$$

for every measurable function ψ.

Definition 3.3. A mapping T that solves the Monge problem in Equation (3.1), should it exist, is called a Monge mapping between μ and ν.

In the discrete setting, we consider histograms $\mathbf{a} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T \in \Sigma_m$ and $\mathbf{b} = \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix}^T \in \Sigma_n$, corresponding to the probability measures $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$ for $\mathcal{X} = \{x_i\}_{i=1}^m$ and $\mathcal{Y} = \{y_j\}_{j=1}^n$. We can now encode the transport map $T\colon \{x_i\}_{i=1}^m \to \{y_j\}_{j=1}^n$ as a map between indices $\sigma\colon [\![m]\!] \to [\![n]\!]$ with $\sigma(i) = j$ corresponding to $T(x_i) = y_j$. Additionally, we can define a cost matrix given by

$$C := \begin{bmatrix} c(x_i, y_j) \end{bmatrix}_{i,j} \in \mathbb{R}_+^{m \times n}. \tag{3.4}$$

The Monge problem in Equation (3.1), then, becomes

$$M(\mathbf{a}, \mathbf{b}; C) = \inf_{\sigma\colon [\![m]\!] \to [\![n]\!]} \left\{ \sum_{i=1}^m a_i \cdot C_{i,\sigma(i)} : b_j = \sum_{i \in \sigma^{-1}(j)} a_i \right\}. \tag{3.5}$$

Thus, in this setting, we can interpret the Monge problem as seeking to find the optimal rearrangement or optimal assignment of mass with minimal cost.

The Monge problem is a traditionally difficult problem to solve between arbitrary measures μ and ν, particularly due to the fact that a mapping T satisfying $T_{\#}\mu = \nu$ may not always exist. In particular, we can consider the case when $n > m$ in the discrete setting of the problem: if all $a_i, b_j > 0$, then the map T must be surjective, but this is not possible if $n > m$. A similar issue arises if μ is a discrete measure while ν is continuous. The following theorem characterizes sufficient conditions under which at least one transport map is guaranteed to exist.

Theorem 3.6 ([San15, Corollary 1.29]). *If $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and μ is atomless, then there exists a transport map T such that $T_{\#}\mu = \nu$.*

More generally, however, the constraint on the mappings T is not closed under weak convergence, making it impossible to prove the existence of a minimizer to Equation (3.1) in general [San15]. Furthermore, from a computational perspective, the optimization problem in Equation (3.5) is non-convex, making it hard to solve in practice. However, thanks to the work of Brenier [Bre91] and later Gangbo and McCann [GM96], we have the following theorem on the existence of solutions to the Monge problem. Note that this theorem handles all costs of the form $c(x, y) = |x - y|^p$ for $p > 1$.

Theorem 3.7. *Let $\Omega \subset \mathbb{R}^d$ be a compact domain with $\mu, \nu \in \mathcal{P}(\Omega)$, and let $c(x, y) = h(x - y)$ for h strictly convex. If μ is absolutely continuous with resepct to the Lebesgue measure, then there exists a unique Monge mapping T between μ and ν.*

Following Monge's formulation of the OT problem, Leonid Kantorovich would later independently propose a relaxed formulation of the OT problem in the 20th century [Kan06a, Kan06b]. Intuitively, while Monge's formulation seeks a deterministic transport map between two measures μ and ν such that mass in the source measure cannot be split, Kantorovich's formulation seeks a *transportation plan* between μ and ν which allows source mass to be distributed between multiple target locations.

As before, we consider probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ for Borel spaces $\mathcal{X}, \mathcal{Y}$ and a cost function $c\colon \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$. The set of transportation plans between μ and ν is defined as

$$\Gamma(\mu, \nu) := \{\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : (\pi_{\mathcal{X}})_{\#}\gamma = \mu, (\pi_{\mathcal{Y}})_{\#}\gamma = \nu\}. \tag{3.8}$$

In this formulation, given measurable sets $A \subseteq \mathcal{X}, B \subseteq \mathcal{Y}$, $\gamma(A, B)$ is the amount of mass being transported from A to B. Note that $\Gamma(\mu, \nu)$ is always non-empty, as we trivially have that the product measure $\mu \otimes \nu \in \Gamma(\mu, \nu)$. The Kantorovich problem, then, is given as

$$K(\mu, \nu; c) = \min_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, \mathrm{d}\gamma(x, y). \tag{3.9}$$

We also consider the discrete setting of the Kantorovich problem between histograms $\mathbf{a} \in \Sigma_m$ and $\mathbf{b} \in \Sigma_n$, as before. In this setting, we can instead represent a transportation plan with an $m \times n$ matrix $P \in \mathbb{R}_+^{m\times n}$ where $P_{i,j}$ encodes the amount of mass being transported from the point x_i to y_j. We now denote the set of transport plans as

$$U(\mathbf{a}, \mathbf{b}) := \{P \in \mathbb{R}_+^{m\times n} : P\mathbb{1}_n = \mathbf{a}, P^T\mathbb{1}_m = \mathbf{b}\}. \tag{3.10}$$

Thus, using the cost matrix as defined in Equation (3.4), the Kantorovich problem becomes

$$K(\mathbf{a}, \mathbf{b}; C) = \min_{P \in U(\mathbf{a},\mathbf{b})} \langle C, P\rangle. \tag{3.11}$$

Unlike the Monge problem, the Kantorovich problem in Equation (3.9) will always admit a solution under mild regularity conditions on the underlying spaces $\mathcal{X}, \mathcal{Y}$ and the cost function $c(\cdot, \cdot)$, as demonstrated by the following theorems.

Theorem 3.12 ([San15, Theorem 1.4])**.** *Let $\mathcal{X}, \mathcal{Y}$ be compact metric spaces with $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Then, the Kantorovich problem in Equation (3.9) admits a solution if $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is continuous.*

Theorem 3.13 ([San15, Theorem 1.5])**.** *Let $\mathcal{X}, \mathcal{Y}$ be compact metric spaces with $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Then, the Kantorovich problem in Equation (3.9) admits a solution if $c\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and bounded from below.*

Theorem 3.14 ([San15, Theorem 1.7])**.** *Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces (complete and separable metric spaces) with $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Then, the Kantorovich problem in Equation (3.9) admits a solution if $c\colon \mathcal{X} \times \mathcal{Y} \to [0, +\infty]$ is lower semi-continuous.*

In the discrete setting, we observe that Equation (3.11) is now simply a linear programming problem; since the set of transport plans is non-empty, as the outer product $\mathbf{a} \otimes \mathbf{b} \in U(\mathbf{a}, \mathbf{b})$, and the objective function is bounded below by the minimum entry of C, Equation (3.11) always has an optimal solution. Hence, Equation (3.11) can be solved efficiently for any two histograms $\mathbf{a}, \mathbf{b}$ using standard linear programming methods (e.g., network simplex) [Vil08, PC18]. This is a particularly important fact, as in computational settings, one can often approximate non-discrete measures μ, ν with discrete ones (e.g., through sampling), and it follows that if μ^k, ν^k are sequences of measures that weakly converge to μ and ν, respectively, and γ^k is an optimal transport plan between μ^k, ν^k, then, up to subsequences, γ^k will converge weakly to the optimal transport plan γ^* between μ, ν [Vil08, Theorem 5.20].

Finally, we note that the Kantorovich problem is indeed a relaxation of the Monge problem [San15, Theorem 1.33]. In particular, any valid transportation map $T\colon \mathcal{X} \to \mathcal{Y}$ satisfying $T_{\#}\mu = \nu$ induces a transport plan as follows:

$$\gamma = (\mathrm{id}, T)_{\#}\mu \in \Gamma(\mu, \nu). \tag{3.15}$$

Furthermore, if T is a Monge mapping between μ and ν (i.e., T solves the Monge problem in Equation (3.1)), then $\gamma^* = (\mathrm{id}, T)_{\#}\mu$ is an optimal transportation plan for the Kantorovich problem in Equation (3.9). While the Kantorovich problem may have multiple minimizers in general, under the assumptions of Theorem 3.7, there is a unique optimal transport plan, which is the transport plan induced by the unique Monge mapping given in Theorem 3.7.

4. Wasserstein Space

Our discussion of the Monge and Kantorovich formulations of the classical OT problem leads naturally to a discussion of Wasserstein distances and Wasserstein space. In what follows, $(\mathcal{X}, d)$ will denote a compact metric space (or alternatively, a Polish space), $p \in [1, \infty)$, and $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$ will be Borel probability measures with finite pth moment. That is,

$$\mathcal{P}_p(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(\mathcal{X}) : \int_{\mathcal{X}} d(x, x_0)^p \, \mathrm{d}\mu(x) < \infty \text{ for some } x_0 \in \mathcal{X} \right\}. \tag{4.1}$$

The p-Wasserstein distance between measures μ, ν is then defined as

$$W_p(\mu, \nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} d(x, y)^p \, \mathrm{d}\gamma(x, y) \right)^{1/p}. \tag{4.2}$$

One can immediately observe that computing the p-Wasserstein distance between the measures μ, ν corresponds to an instance of the Kantorovich problem $K(\mu, \nu; d^p)$. The Wasserstein distance serves as a natural way to define a metric on the space $\mathcal{P}_p(\mathcal{X})$ based on the metric of the underlying space $\mathcal{X}$. In fact, it follows from this formulation that $W_p(\delta_x, \delta_y) = d(x, y)$ for any $x, y \in \mathcal{X}$. We formalize the metric properties of W_p in the following proposition.

Proposition 4.3 ([Vil08, Chapter 6])**.** *W_p is a metric on the space $\mathcal{P}_p(\mathcal{X})$. If $\mathcal{X}$ is a compact metric space (resp., a Polish space), then $\mathcal{P}_p(\mathcal{X})$ equipped with the p-Wasserstein metric is a compact metric space (resp., a Polish space).*

The preceding proposition motivates the following definition.

Definition 4.4. When equipped with the W_p metric, $\mathcal{P}_p(\mathcal{X})$ is referred to as Wasserstein space, and is denoted as $\mathbb{W}_p(\mathcal{X})$.

We can observe that for $p \leq q$, a straightforward application of Jensen's inequality or Hölder's inequality yields that $W_p \leq W_q$. It also happens that in the topology induced by W_p on $\mathcal{P}_p(\mathcal{X})$, convergence under W_p is equivalent to the weak convergence of probability measures in $\mathcal{P}_p(\mathcal{X})$. This result is given in the following theorem.

Theorem 4.5 ([Vil08, Theorem 6.9])**.** *Let $(\mu_k)_{k=1}^{\infty} \subset \mathcal{P}_p(\mathcal{X})$ be a sequence of probability measures and let $\mu \in \mathcal{P}_p(\mathcal{X})$. Then, $\mu_n \rightharpoonup \mu$ in $\mathcal{P}_p(\mathcal{X})$ if and only if $W_p(\mu_n, \mu) \to 0$.*

Finally, we will discuss the characterization of Wasserstein space as a geodesic space, and explore the relationship between these geodesics and the notion of Wasserstein barycenters. First, recall that for a metric space $(\mathcal{X}, d)$, a constant speed geodesic between two points $x_0, x_1 \in \mathcal{X}$ is a continuous curve $x\colon [0, 1] \to \mathcal{X}$ such that $d(x_s, x_t) = |s - t| d(x_0, x_1)$ for any $s, t \in [0, 1]$, where we are using the notation $x(t)$ and x_t interchangeably. The following theorem characterizes geodesics in Wasserstein space.

Theorem 4.6 ([San15, Theorem 5.27]). *Let $(\mathcal{X}, d)$ be a compact and convex metric space, and let $\mu, \nu \in \mathbb{W}_p(\mathcal{X})$. Let $\gamma \in \Gamma(\mu, \nu)$ be an optimal transport plan between μ, ν. Then, define the map $\chi_t \colon \mathcal{X} \times \mathcal{X} \to \mathcal{X}$ such that $(x, y) \mapsto (1-t)x + ty$. Then the curve $\omega(t) := (\chi_t)_{\#}\gamma$ is a constant speed geodesic in $\mathbb{W}_p(\mathcal{X})$ connecting μ and ν.*

Corollary 4.7. *Under the assumptions of Theorem 4.6, if there exists a Monge mapping T between μ and ν, then the constant speed geodesic given in Theorem 4.6 is equivalent to the curve $\omega(t) = ((1-t)\,\mathrm{id} + tT)_{\#}\mu$.*

The following theorem is derived from the preceding results and characterizes $\mathbb{W}_p(\mathcal{X})$ as a geodesic space.

Theorem 4.8 ([San15, Theorem 5.27]). *Under the assumptions of Theorem 4.6, for any $\mu, \nu \in \mathbb{W}_p(\mathcal{X})$, there exists (at least one) constant speed geodesic between μ and ν. Thus, $\mathbb{W}_p(\mathcal{X})$ is a geodesic space.*

Finally, we define the notion of Wasserstein barycenters and their relation to constant speed geodesics in $\mathbb{W}_p(\mathcal{X})$. Consider a collection of measures $\mu_1, \ldots, \mu_N \in \mathbb{W}_p(\mathcal{X})$ and weights $\lambda_1, \ldots, \lambda_n \in \mathbb{R}_+$. The Wasserstein barycenter of this collection of measures is defined to be the measure μ^* such that

$$\mu^* = \operatorname*{arg\,min}_{\mu \in \mathbb{W}_p(\mathcal{X})} \sum_{i=1}^{N} \lambda_i W_p^p(\mu, \mu_i). \tag{4.9}$$

Hence, the Wasserstein barycenter can be thought of as the Fréchet mean in Wasserstein space, generalizing the notion of geodesics. In fact, in the following remark, we characterize the relationship between Wasserstein barycenters between $N = 2$ measures and the constant speed geodesic between the measures.

Remark 4.10. By Theorem 4.8, since $\mathbb{W}_p(\mathcal{X})$ is a geodesic space for $\mathcal{X}$ compact and convex, we observe that in the special case with $N = 2$ measures, the Wasserstein barycenter of two measures in $\mathbb{W}_p(\mathcal{X})$ must lie on the geodesic between them. That is, consider the measures $\mu_0, \mu_1 \in \mathbb{W}_p(\mathcal{X})$; without loss of generality, we may assume that the weights for the Wasserstein barycenter are given by $(\lambda, 1-\lambda) \in \Sigma_2$ since the barycenter is the same up to rescaling/normalization of the weights. Then, the barycenter μ^* of μ_0, μ_1 lies on the constant speed geodesic between μ_0, μ_1. In particular, given an optimal transport map $\gamma \in \Gamma(\mu, \nu)$, the barycenter μ^* is given by

$$\mu^* = (\chi_{1-\lambda})_{\#}\gamma. \tag{4.11}$$

5. Discrete Multi-Stage Optimal Transport

In this section, we begin to formalize the MSOT problem, which is motivated by real-world transport processes requiring transportation across multiple stages. As examples, one can consider scenarios where materials may need to travel from extraction sites to factories before finally reaching

consumers; data may need to pass through several processing nodes in a network before arriving at its final destination; and goods in a supply chain might have to be routed through distribution hubs on their way from warehouses to retail locations. These scenarios involve intermediate structure, where the source and target distributions cannot be directly linked without accounting for intervening space, cost, or infrastructure. To model such systems, we introduce the MSOT problem, a natural generalization of classical OT that incorporates one or more intermediate stages through which transport must occur.

Here, we will focus our attention to the discrete formulation of the MSOT problem, but we discuss the continuous formulation in Section 7. In the discrete setting, we consider a discrete source domain $\mathcal{X} = \{x_i\}_{i=1}^m$ and discrete target domain $\mathcal{Y} = \{y_j\}_{j=1}^n$. As before, we wish to study transport between the measures $\mu = \sum_{i=1}^n a_i \delta_{x_i}$ and $\nu = \sum_{j=1}^m b_j \delta_{y_j}$, or equivalently, between histograms $\mathbf{a} \in \Sigma_m$ and $\mathbf{b} \in \Sigma_n$. In the MSOT problem, however, we cannot directly transport mass from the source and target; instead, there is some set of intermediate locations $\mathcal{Z} = \{z_k\}_{k=1}^p$, and we must transport mass from the source domain $\mathcal{X}$ to this intermediate domain $\mathcal{Z}$, and then from the intermediate domain $\mathcal{Z}$ to the target domain $\mathcal{Y}$.

We consider a cost function $c_1 \colon \mathcal{X} \times \mathcal{Z} \to [0, +\infty)$ giving the transport cost between $\mathcal{X}$ and $\mathcal{Z}$, and a cost function $c_2 \colon \mathcal{Z} \times \mathcal{Y} \to [0, +\infty)$ giving the transport cost between $\mathcal{Z}$ and $\mathcal{Y}$. The transport costs can be encoded in the matrices

$$C_1 := \big[c_1(x_i, z_k)\big]_{i,k} \in \mathbb{R}_+^{m \times p} \tag{5.1}$$

$$C_2 := \big[c_2(z_k, y_j)\big]_{k,j} \in \mathbb{R}_+^{p \times n}, \tag{5.2}$$

analogous to Equation (3.4) from the discrete formulation of the classical OT problem. The Monge formulation of the MSOT problem, then, is given as

$$M_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2) = \inf_{\substack{\sigma_1 \colon [\![m]\!] \to [\![p]\!] \\ \sigma_2 \colon [\![p]\!] \to [\![n]\!]}} \left\{ \sum_{i=1}^m a_i \cdot \big((C_1)_{i,\sigma_1(i)} + (C_2)_{\sigma_1(i),(\sigma_2 \circ \sigma_1)(i)}\big) : b_j = \sum_{i \in (\sigma_2 \circ \sigma_1)^{-1}(j)} a_i \right\}. \tag{5.3}$$

Notably, in the preceding formulation, there is no constraint on what the distribution looks like on the intermediate space $\mathcal{Z}$; instead, the intermediate distribution can take whatever form minimizes the cost of the overall transport process. This adds additional complexity to the problem, as enforcing that the distribution follow a particular histogram, say $\mathbf{c} \in \Sigma_p$, on the intermediate space would simply reduce the problem to solving $M(\mathbf{a}, \mathbf{c}; C_1) + M(\mathbf{c}, \mathbf{b}; C_2)$.

Under the same setup as the Monge formulation of the MSOT problem, we can similarly define a Kantorovich formulation of the MSOT problem. The Kantorovich formulation of the MSOT problem, as is the case with the classical OT problem, relaxes the Monge formulation and allows for mass splitting to occur in both stages of the transportation process. We now consider the set of two-step transportation plans,

$$U_{\mathcal{Z}}(\mathbf{a},\mathbf{b}) := \left\{ (P_1,P_2) \in \mathbb{R}_+^{m\times p} \times \mathbb{R}_+^{p\times n} : \begin{array}{c} P_1\mathbb{1}_p = \mathbf{a}, \\ P_1^T\mathbb{1}_m = P_2\mathbb{1}_n, \\ P_2^T\mathbb{1}_p = \mathbf{b} \end{array} \right\}, \tag{5.4}$$

which extends the definition of transportation plans in Equation (3.10) from the classical OT problem with the additional requirement that P_1, P_2 agree on their shared marginal. The Kantorovich formulation of the MSOT problem thus follows as

$$K_{\mathcal{Z}}(\mathbf{a},\mathbf{b};C_1,C_2) = \min_{(P_1,P_2)\in U_{\mathcal{Z}}(\mathbf{a},\mathbf{b})} \langle C_1,P_1\rangle + \langle C_2,P_2\rangle. \tag{5.5}$$

Just as in the classical OT problem, the Kantorovich formulation is indeed a relaxation of the Monge formulation in the MSOT setting. We discuss this in further detail in Remark 7.6 in Section 7. We briefly note that while we only consider one intermediate stage in these formulations, one can easily extend both Equations (5.3) and (5.5) to two or more intermediate stages. In this regard, we also direct the reader to Remarks 6.1 and 7.27.

We now discuss the solutions to the discrete MSOT problem, beginning with the following lemma regarding the existence of a minimizer.

Lemma 5.6. *The Kantorovich formulation of the discrete MSOT problem in Equation (5.5) admits a minimizer.*

Proof. Similar to the Kantorovich formulation of the classical OT problem (see Equation (3.11)), we observe that Equation (5.5) is a linear programming problem. Hence, it suffices to show that the constraint set $U_{\mathcal{Z}}(\mathbf{a},\mathbf{b})$ is always non-empty and the objective function is bounded below. Indeed, we observe that for all $\mathbf{c} \in \Sigma_p$, we have that $(\mathbf{a}\otimes\mathbf{c}, \mathbf{c}\otimes\mathbf{b}) \in U_{\mathcal{Z}}(\mathbf{a},\mathbf{b})$. Then, since $\sum_{i,k}(P_1)_{i,k} = \sum_{k,j}(P_2)_{k,j} = 1$, it follows that the objective function is bounded below by $\min_{i,k}(C_1)_{i,k} + \min_{k,j}(C_2)_{k,j} > -\infty$. □

In the remainder of this section, we will further characterize the form of solutions to the discrete MSOT problem, which will motivate the development of an optimal algorithm for solving the discrete MSOT problem in Section 6. In particular, we will show that the Kantorovich formulation of the discrete MSOT problem can be reduced to an instance of the classical Kantorovich problem (i.e., without intermediate stages), and we will then invert this reduction to derive the solutions to the original MSOT problem. We then discuss the conditions under which the solution to the Kantorovich formulation of the MSOT problem is induced by Monge mappings in Remark

5.22. We first prove the following proposition characterizing the solutions to the Kantorovich MSOT problem.

Proposition 5.7. *For any instance of the Kantorovich MSOT problem $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$ given in Equation (5.5), there exists an optimal solution (P_1^*, P_2^*) that satisfies the following condition:*

(∗) *For any $i \in [\![m]\!], j \in [\![n]\!]$, there is at most one $k \in [\![p]\!]$ for which both $(P_1^*)_{i,k} > 0$ and $(P_2^*)_{k,j} > 0$. Moreover, this k must satisfy*

$$k \in \arg\min_{\ell \in [\![p]\!]} \Big[(C_1)_{i,\ell} + (C_2)_{\ell,j}\Big]. \tag{5.8}$$

Proof. Suppose that (P_1, P_2) is optimal in Equation (5.5), but does not satisfy the condition (∗). That is, there exists i, j and $k_1, \dots, k_C$ such that the terms $(P_1)_{i,k_\alpha}, (P_2)_{k_\alpha,j}$ are non-zero for $\alpha \in [\![C]\!]$. Then, for $\alpha \in [\![C]\!]$, we define $\zeta(\alpha) := (C_1)_{i,k_\alpha} + (C_2)_{k_\alpha,j}$, which is the transportation cost between i, j going through k_α, and $\rho(\alpha) := \min\{(P_1)_{i,k_\alpha}, (P_2)_{k_\alpha,j}\}$, which is the mass being transported between i, j though k_α.

Without loss of generality, we can assume that $\zeta(1) \le \zeta(2) \le \cdots \le \zeta(C)$ (i.e., by reordering the $k_1, \dots, k_C$). We also denote the gap $\delta(\alpha) := \zeta(1) - \zeta(\alpha) \le 0$. Then, we may construct (P_1^*, P_2^*) from (P_1, P_2) such that

$$(P_1^*)_{i,k_1} = (P_1)_{i,k_1} + \sum_{\alpha=2}^{C} \rho(\alpha) \tag{5.9}$$

$$(P_2^*)_{k_1,j} = (P_2)_{k_1,j} + \sum_{\alpha=2}^{C} \rho(\alpha), \tag{5.10}$$

with $(P_1^*)_{i,k_\alpha} = (P_1)_{i,k_\alpha} - \rho(\alpha)$ and $(P_2^*)_{k_\alpha,j} = (P_2)_{k_\alpha,j} - \rho(\alpha)$ for $\alpha \in [\![C]\!] \setminus \{1\}$ and all other entries equal to P_1, P_2. Then, $(P_1^*, P_2^*) \in U_{\mathcal{Z}}(\mathbf{a}, \mathbf{b})$ since $P_1^* \mathbb{1}_p = P_1 \mathbb{1}_p = \mathbf{a}$, $(P_2^*)^T \mathbb{1}_p = P_2^T \mathbb{1}_p = \mathbf{b}$, and $(P_1^*)^T \mathbb{1}_m = (P_2^*)\mathbb{1}_n$ by construction. However, (P_1^*, P_2^*) has cost

$$\begin{aligned} \langle C_1, P_1^* \rangle + \langle C_2, P_2^* \rangle &= \langle C_1, P_1 \rangle + \langle C_2, P_2 \rangle + \sum_{\alpha=2}^{C} \rho(\alpha)\delta(\alpha) \\ &\ge \langle C_1, P_1 \rangle + \langle C_2, P_2 \rangle, \end{aligned} \tag{5.11}$$

where the rightmost inequality comes from the fact that we assume that (P_1, P_2) is optimal. Thus, for all $\alpha \in [\![C]\!] \setminus \{1\}$, since $\rho(\alpha) > 0$ and $\delta(\alpha) \le 0$, we must have that $\delta(\alpha) = 0$, from which it follows that (P_1^*, P_2^*) is also optimal. This must further imply that

$$k_1, \dots, k_C \in \arg\min_{\ell \in [\![p]\!]} \Big[(C_1)_{i,\ell} + (C_2)_{\ell,j}\Big], \tag{5.12}$$

as otherwise the preceding argument would contradict the optimality of (P_1, P_2) (i.e., we could similarly redistribute mass through the k given in Equation (5.8) and get a lower cost). Hence, since either $(P_1^*)_{i,k_\alpha} = 0$ or

$(P_2^*)_{k_\alpha,j} = 0$ for $\alpha \in [\![C]\!] \setminus \{1\}$, (P_1^*, P_2^*) satisfies condition $(*)$, completing the proof. □

We now utilize the result of Proposition 5.7 to prove the following theorem, which demonstrates the aforementioned reduction of the discrete Kantorovich MSOT problem to an instance of the classical Kantorovich problem.

Theorem 5.13. *Consider an instance of the discrete Kantorovich MSOT problem $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$. We define a reduced cost matrix $C \in \mathbb{R}_+^{m \times n}$ given by*

$$C_{i,j} := \min_{\ell \in [\![p]\!]} \Big[(C_1)_{i,\ell} + (C_2)_{\ell,j} \Big]. \tag{5.14}$$

It follows that $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2) = K(\mathbf{a}, \mathbf{b}; C)$.

Proof. Suppose that $(P_1^*, P_2^*) \in U_{\mathcal{Z}}(\mathbf{a}, \mathbf{b})$ is optimal for $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$. By Proposition 5.7, we may assume that (P_1^*, P_2^*) satisfies the condition $(*)$. For fixed i, j we will denote

$$k_{i,j} := \arg\min_{\ell \in [\![p]\!]} \Big[(C_1)_{i,\ell} + (C_2)_{\ell,j} \Big]. \tag{5.15}$$

If there are multiple minimizers to Equation (5.15), we can choose any one at random. We now construct a solution P^* for $K(\mathbf{a}, \mathbf{b}; C)$ as follows:

$$P^*_{i,j} = \min\Big\{ (P_1^*)_{i,k_{i,j}}, (P_2^*)_{k_{i,j},j} \Big\}. \tag{5.16}$$

It is clear that $P^* \in U(\mathbf{a}, \mathbf{b})$ by construction, and we observe that

(5.17)

$$\begin{aligned}
\langle C, P^* \rangle = \sum_{i,j} C_{i,j} (P^*)_{i,j} &= \sum_{i,j} \left(\min_{\ell} \Big[(C_1)_{i,\ell} + (C_2)_{\ell,j} \Big] \right) (P^*)_{i,j} \\
&= \sum_{i,j} \big((C_1)_{i,k_{i,j}} + (C_2)_{k_{i,j},j} \big) (P^*)_{i,j} \\
&= \sum_{i,j} (C_1)_{i,k_{i,j}} \min\Big\{ (P_1^*)_{i,k_{i,j}}, (P_2^*)_{k_{i,j},j} \Big\} \\
&\qquad + \sum_{i,j} (C_2)_{k_{i,j},j} \min\Big\{ (P_1^*)_{i,k_{i,j}}, (P_2^*)_{k_{i,j},j} \Big\} \\
&= \sum_{i,k} (C_1)_{i,k} (P_1^*)_{i,k} + \sum_{k,j} (C_2)_{k,j} (P_2^*)_{k,j} \\
&= \langle C_1, P_1^* \rangle + \langle C_2, P_2^* \rangle.
\end{aligned}$$

We now show that P^* is optimal for $K(\mathbf{a}, \mathbf{b}; C)$. Suppose for the sake of contradiction that P^* is not optimal, and instead there is an optimal

solution $P' \neq P^*$. We can now define P'_1, P'_2 as follows:

$$(P'_1)_{i,\ell} = \sum_{j:k_{i,j}=\ell} P'_{i,j} \tag{5.18}$$

$$(P'_2)_{\ell,j} = \sum_{i:k_{i,j}=\ell} P'_{i,j}. \tag{5.19}$$

We can immediately verify that $(P'_1, P'_2) \in U_{\mathcal{Z}}(\mathbf{a}, \mathbf{b})$. Moreover, we have that

$$\begin{aligned} \langle C_1, P'_1\rangle + \langle C_2, P'_2\rangle &= \sum_{i,\ell}(C_1)_{i,\ell}(P'_1)_{i,\ell} + \sum_{\ell,j}(C_2)_{\ell,j}\big(P'_2\big)_{\ell,j} \\ &= \sum_{i,\ell}(C_1)_{i,\ell}\sum_{j:\ell=k_{i,j}} P'_{i,j} + \sum_{\ell,j}(C_2)_{\ell,j}\sum_{i:\ell=k_{i,j}} P'_{i,j} \\ &= \sum_{i,j}\big[(C_1)_{i,k_{i,j}} + (C_2)_{k_{i,j},j}\big]P'_{i,j} \\ &= \sum_{i,j} C_{i,j}P'_{i,j} \\ &= \langle C, P'\rangle. \end{aligned} \tag{5.20}$$

Since we assumed that $P' \neq P^*$ is optimal, we thus have that

$$\langle C_1, P'_1\rangle + \langle C_2, P'_2\rangle = \langle C, P'\rangle < \langle C, P^*\rangle = \langle C_1, P^*_1\rangle + \langle C_2, P^*_2\rangle. \tag{5.21}$$

This contradicts the optimality of (P^*_1, P^*_2), yielding the desired result. □

We now provide the following remark regarding when the Monge and Kantorovich MSOT problems coincide.

Remark 5.22. We consider the setting under which the optimal solution to $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$ is induced by Monge mappings, thus giving a solution to $M_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$. We first require that the reduced problem is induced by a Monge mapping, giving $K(\mathbf{a}, \mathbf{b}; C) = M(\mathbf{a}, \mathbf{b}; C)$, with C as defined in Equation (5.14). Let this Monge mapping be denoted by σ. Then, $K_{\mathcal{Z}}(\mathbf{a}, \mathbf{b}; C_1, C_2)$ is induced by Monge mappings if and only if for all $j_1 \neq j_2$, $k_{\alpha,j_1} \neq k_{\beta,j_2}$ for $\alpha \in \sigma^{-1}(j_1), \beta \in \sigma^{-1}(j_2)$, where we are using the notation given in Equation (5.15). Note that there may be multiple choices for $k_{i,j}$ for a given i, j, so it suffices if there exists any series of choices that satisfies the preceding condition. It trivially must hold that $m \geq n$ and $p \geq n$ for Monge mappings to exist, though we do not necessarily require any such relation between m and p a priori.

6. Algorithmic Solution and Complexity Analysis

In this section, using the result of Theorem 5.13, we outline an algorithm to compute solutions to the Kantorovich MSOT problem and analyze its computational complexity. The algorithm is provided in Algorithm 1. As

Algorithm 1: Algorithm to compute solutions to Kantorovich MSOT problem.

Data: Source histogram $\mathbf{a} \in \Sigma_m$, target histogram $\mathbf{b} \in \Sigma_n$, and cost matrices $C_1 \in \mathbb{R}_+^{m\times p}, C_2 \in \mathbb{R}_+^{p\times n}$.
Result: Optimal transportation plans $(P_1^*, P_2^*) \in U_{\mathcal{Z}}(\mathbf{a}, \mathbf{b})$.

```
1  begin
2      C ← 0 ∈ ℝ^{m×n}
3      K ← 0 ∈ ℝ^{m×n}
4      for i = 1, ..., m do
5          for j = 1, ..., n do
6              η ← arg min_{ℓ∈⟦p⟧} [(C_1)_{i,ℓ} + (C_2)_{ℓ,j}]
7              K_{i,j} ← η
8              C_{i,j} ← (C_1)_{i,η} + (C_2)_{η,j}
9          end
10     end
11     P* ← Kantorovich(a, b; C)
12     P_1^* ← 0 ∈ ℝ^{m×p}
13     P_2^* ← 0 ∈ ℝ^{p×n}
14     for i = 1, ..., m do
15         for j = 1, ..., n do
16             η ← K_{i,j}
17             (P_1^*)_{i,η} ← (P_1^*)_{i,η} + (P*)_{i,j}
18             (P_2^*)_{η,j} ← (P_2^*)_{η,j} + (P*)_{i,j}
19         end
20     end
21     return P_1^*, P_2^*
22 end
```

before, we note that on Line 6 of the algorithm, if there exist multiple minimizers, η may be chosen from among them however desired. Furthermore, we assume the procedure Kantorovich$(\mathbf{a}, \mathbf{b}; C)$ invoked on Line 11 outputs the optimal transport plan between the histograms $\mathbf{a}$ and $\mathbf{b}$ given the reduced cost matrix C.

We now consider the computational complexity of Algorithm 1. The complexity of computing η on Line 6 is $\mathcal{O}(p)$. Hence, the total complexity of the nested loop in Lines 4-10 is $\mathcal{O}(mnp)$. Then, the complexity of solving the Kantorovich problem (e.g., by network flow) on Line 11 is $\mathcal{O}((m+n)mn\log(m+n))$. Finally, the nested loop in Lines 14-20 is $\mathcal{O}(mn)$. Hence, the complexity of Algorithm 1 altogether is $\mathcal{O}((m+n)mn\log(m+n)+mnp)$. Meanwhile, the complexity of solving the Kantorovich MSOT problem in its

original form using linear programming has complexity $\mathcal{O}((m+n+p)(mp+np)\log(m+n+p))$ [Tar97], demonstrating the improvement in complexity offered by Algorithm 1. We note that when $p \leq \max(m,n)$, Algorithm 1 has complexity equivalent to solving the reduced Kantorovich problem $K(\mathbf{a},\mathbf{b};C)$, and is thus "optimal" in some sense (i.e., assuming no better algorithm exists to solve the classical Kantorovich problem, no better algorithm can exist to solve the Kantorovich MSOT problem).

Remark 6.1. We additionally remark that the result given in Theorem 5.13, and as a consequence Algorithm 1, can easily be extended to the setting of two or more intermediate stages. Consider the (discrete) intermediate spaces $\mathcal{Z}_1,\dots,\mathcal{Z}_N$, containing $p_1,\dots,p_N$ points, respectively. Then, given the corresponding cost matrices $C_1,\dots,C_{N+1}$, we can similarly compute the reduced cost matrix as

$$C_{i,j} = \min_{(\ell_1,\dots,\ell_N)\in[\![p_1]\!]\times\cdots\times[\![p_N]\!]} \left[\sum_{\alpha=0}^{N}(C_\alpha)_{\ell_\alpha,\ell_{\alpha+1}}\right], \tag{6.2}$$

where we let $\ell_0 := i$ and $\ell_{N+1} := j$ for convenience.

Computing this reduced cost matrix, however, quickly becomes the computational bottleneck for the approach in Algorithm 1. In particular, denoting $p_0 := m$ and $p_{N+1} := n$ for convenience, the computational complexity of this approach becomes

$$\mathcal{O}\left((m+n)mn\log(m+n) + \prod_{\alpha=0}^{N+1} p_\alpha\right), \tag{6.3}$$

whereas the linear programming solution has complexity

$$\mathcal{O}\left(\left(\sum_{\alpha=0}^{N+1} p_\alpha\right)\left(\sum_{\alpha=0}^{N} p_\alpha p_{\alpha+1}\right)\log\left(\sum_{\alpha=0}^{N+1} p_\alpha\right)\right). \tag{6.4}$$

When $m = n = p_1 = \cdots = p_N$, the complexity of our approach becomes $\mathcal{O}(m^3\log m + m^{N+2})$, while the linear programming approach has complexity $\mathcal{O}(m^3\log m)$. However, while our approach may have seemingly worse complexity for $N \geq 2$, we observe that the computational bottleneck of this approach is simply the nested loop required to compute the reduced cost matrix in Equation 6.2, while the linear programming approach may have a large leading constant that is overlooked. Thus, our approach may still be faster for small values of $N \geq 2$. Additionally, the computation of Equation (6.2) can easily be parallelized; alternatively, in the setting of repeated comparisons of different histograms with the same series of cost matrices, this reduced cost matrix can be computed once and reused, amortizing the computational cost of this subprocedure. Hence, this algorithm may still present computational benefits when $N \geq 2$.

7. Continuous Multi-Stage Optimal Transport

We now formulate the MSOT problem in the continuous setting. We consider Borel spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ and probability measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Additionally, we let $c_1 \colon \mathcal{X} \times \mathcal{Z} \to [0, +\infty)$ denote the cost function between $\mathcal{X}$ and $\mathcal{Z}$, and $c_2 \colon \mathcal{Z} \times \mathcal{Y} \to [0, +\infty)$ denote the cost function between $\mathcal{Z}$ and $\mathcal{Y}$. The Monge formulation of the MSOT problem thus follows as

(7.1)

$$M_{\mathcal{Z}}(\mu, \nu; c_1, c_2) = \inf_{\substack{T_1 \colon \mathcal{X} \to \mathcal{Z} \\ T_2 \colon \mathcal{Z} \to \mathcal{Y}}} \left\{ \int_{\mathcal{X}} c_1(x, T_1(x)) + c_2(T_1(x), (T_2 \circ T_1)(x)) \, \mathrm{d}\mu(x) : (T_2 \circ T_1)_{\#}\mu = \nu \right\},$$

where $T_1 \colon \mathcal{X} \to \mathcal{Z}$ and $T_2 \colon \mathcal{Z} \to \mathcal{Y}$ can be any Borel measurable maps.

Similar to the Monge formulation of the classical OT problem, the Monge MSOT problem is difficult to solve for arbitrary measures μ and ν; crucially, it is unclear a priori when mappings T_1, T_2 exist such that $(T_2 \circ T_1)_{\#}\mu = \nu$. As with the classical OT problem, one can easily find that in the discrete setting, taking $n > p$ or $n > m$ immediately precludes the existence of Monge mappings. Alternatively, if μ is discrete while ν is continuous, we similarly find that such mappings cannot exist. Nevertheless, we can show the following result providing sufficient conditions under which transport maps for the Monge MSOT problem exist.

Theorem 7.2. *Let $\mathcal{X} = \mathbb{R}^{d_{\mathcal{X}}}, \mathcal{Y} = \mathbb{R}^{d_{\mathcal{Y}}}, \mathcal{Z} = \mathbb{R}^{d_{\mathcal{Z}}}$. If $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ and μ is atomless, then there exist mappings $T_1 \colon \mathcal{X} \to \mathcal{Z}, T_2 \colon \mathcal{Z} \to \mathcal{Y}$ such that $(T_2 \circ T_1)_{\#}\mu = \nu$.*

To prove this theorem, we first state the following lemma.

Lemma 7.3 ([San15, Lemma 1.28])**.** *There exists a Borel measurable map $\sigma_d \colon \mathbb{R}^d \to \mathbb{R}$ that is injective, whose image is a Borel subset of $\mathbb{R}$, and whose inverse map is Borel measurable.*

With this lemma, we can now prove Theorem 7.2.

Proof of Theorem 7.2. Let $\omega \in \mathcal{P}(\mathcal{Z})$ be atomless. By Theorem 3.6, there exists a transport map T_1' between $(\sigma_{d_{\mathcal{X}}})_{\#}\mu$ and $(\sigma_{d_{\mathcal{Z}}})_{\#}\omega$, and a transport map T_2' between $(\sigma_{d_{\mathcal{Z}}})_{\#}\omega$ and $(\sigma_{d_{\mathcal{Y}}})_{\#}\nu$. The desired mappings T_1, T_2 are then given by $T_1 := (\sigma_{d_{\mathcal{Z}}})^{-1} \circ T_1' \circ \sigma_{d_{\mathcal{X}}}$ and $T_2 := (\sigma_{d_{\mathcal{Y}}})^{-1} \circ T_2' \circ \sigma_{d_{\mathcal{Z}}}$. □

We will later discuss sufficient assumptions for the existence of Monge mappings (i.e., minimizers to Equation (7.1)) in Corollary 7.14. Just as before, we can relax the Monge formulation through the Kantorovich formulation of the MSOT problem. To do so, we first define the set of two-step

transportation plans as follows:

(7.4)
$$\Gamma_{\mathcal{Z}}(\mu,\nu) = \left\{ (\gamma_1,\gamma_2) \in \mathcal{P}(\mathcal{X}\times\mathcal{Z}) \times \mathcal{P}(\mathcal{Z}\times\mathcal{Y}) : \begin{array}{r} (\pi_{\mathcal{X}})_{\#}\gamma_1 = \mu, \\ (\pi_{\mathcal{Z}})_{\#}\gamma_1 = (\pi_{\mathcal{Z}})_{\#}\gamma_2, \\ (\pi_{\mathcal{Y}})_{\#}\gamma_2 = \nu \end{array} \right\}.$$

The Kantorovich formulation of the MSOT problem is then given as

$$K_{\mathcal{Z}}(\mu,\nu;c_1,c_2) = \min_{(\gamma_1,\gamma_2)\in\Gamma_{\mathcal{Z}}(\mu,\nu)} \left[\int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1(x,z) + \int_{\mathcal{Z}\times\mathcal{Y}} c_2(z,y)\,\mathrm{d}\gamma_2(z,y) \right]. \tag{7.5}$$

Remark 7.6. The Kantorovich formulation of the MSOT problem is indeed a relaxation of the Monge formulation, similar to the corresponding formulations of the classical OT problem. For any valid set of transportation maps $T_1 : \mathcal{X} \to \mathcal{Z}$ and $T_2 : \mathcal{Z} \to \mathcal{Y}$ such that $(T_2 \circ T_1)_{\#}\mu = \nu$, the transport plans induced by these maps, $\gamma_1 = (\mathrm{id}, T_1)_{\#}\mu$ and $\gamma_2 = (T_1, T_2 \circ T_1)_{\#}\mu$, are valid transport plans for the Kantorovich MSOT problem, in the sense that $(\gamma_1,\gamma_2) \in \Gamma_{\mathcal{Z}}(\mu,\nu)$. Additionally, if T_1, T_2 are Monge mappings (i.e., T_1, T_2 are optimal for $M_{\mathcal{Z}}(\mu,\nu;c_1,c_2)$), then the induced transport plans γ_1,γ_2 as defined above are optimal for $K_{\mathcal{Z}}(\mu,\nu;c_1,c_2)$.

We first consider solutions to the Kantorovich formulation of the MSOT problem in the special case where $(\mathcal{X},d)$ is a compact and convex metric space and all measures lie on the same underlying space (i.e., $\mathcal{X}=\mathcal{Y}=\mathcal{Z}$). Additionally, for $p \in [1,\infty)$, we assume that $c_1 = C_1 d^p$ and $c_2 = C_2 d^p$ for constants $C_1, C_2 \in \mathbb{R}_+$. Before we proceed any further, we first require the following lemma, demonstrating that the Kantorovich MSOT problem exhibits some notion of optimal substructure.

Lemma 7.7. *Let $\mathcal{X},\mathcal{Y},\mathcal{Z}$ be compact metric spaces, and let $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. Additionally, let $c_1 \colon \mathcal{X}\times\mathcal{Z} \to \mathbb{R}$ and $c_2 \colon \mathcal{Z}\times\mathcal{Y} \to \mathbb{R}$ be continuous. Suppose that $(\gamma_1^*,\gamma_2^*) \in \Gamma_{\mathcal{Z}}(\mu,\nu)$ are the optimal transport plans for $K_{\mathcal{Z}}(\mu,\nu;c_1,c_2)$. Then, if we denote $\omega := (\pi_Z)_{\#}\gamma_1^* = (\pi_Z)_{\#}\gamma_2^*$, it follows that $\gamma_1^* \in \Gamma(\mu,\omega)$ is an optimal transport plan for $K(\mu,\omega;c_1)$, and $\gamma_2^* \in \Gamma(\omega,\nu)$ is an optimal transport plan for $K(\omega,\nu;c_2)$.*

Proof. Suppose that γ_1^* was not optimal for $K(\mu,\omega;c_1)$. Then, by Theorem 3.12, there must exist some other optimal transport plan $\gamma_1' \in \Gamma(\mu,\omega)$ with lower transport cost,

$$\int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1'(x,z) < \int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1^*(x,z). \tag{7.8}$$

However, we observe that $(\gamma_1', \gamma_2^*) \in \Gamma_{\mathcal{Z}}(\mu, \nu)$ and has transport cost

$$\int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1'(x,z) + \int_{\mathcal{Z}\times\mathcal{Y}} c_2(z,y)\,\mathrm{d}\gamma_2^*(z,y) < \int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1^*(x,z) + \int_{\mathcal{Z}\times\mathcal{Y}} c_2(z,y)\,\mathrm{d}\gamma_2^*(z,y), \tag{7.9}$$

contradicting the optimality of (γ_1^*, γ_2^*). An identical argument shows the optimality of $\gamma_2^* \in \Gamma(\omega, \nu)$ for $K(\omega, \nu; c_2)$. □

We also require the additional lemma presented below.

Lemma 7.10. *Let $(\mathcal{X}, d)$ be a compact and convex metric space and let $p \in [1, \infty)$. Suppose $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$, and let $c_1 = C_1 d^p$ and $c_2 = C_2 d^p$ for $C_1, C_2 \in \mathbb{R}_+$. It follows that $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$ is equivalent to the Wasserstein barycenter problem.*

Proof. From Lemma 7.7, it follows that $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$ can be reformulated as follows:

$$\begin{aligned} K_{\mathcal{X}}(\mu, \nu; c_1, c_2) &= \min_{\omega \in \mathcal{P}_p(\mathcal{X})} \left[\min_{\gamma_1 \in \Gamma(\mu,\omega)} \int_{\mathcal{X}\times\mathcal{X}} c_1(x,z)\,\mathrm{d}\gamma_1(x,z) \right. \\ &\qquad \left. + \min_{\gamma_2 \in \Gamma(\omega,\nu)} \int_{\mathcal{X}\times\mathcal{X}} c_2(z,y)\,\mathrm{d}\gamma_2(z,y) \right] \\ &= \min_{\omega \in \mathcal{P}_p(\mathcal{X})} \Big[C_1 W_p^p(\mu, \omega) + C_2 W_p^p(\omega, \nu) \Big]. \end{aligned} \tag{7.11}$$

Thus, we indeed have that $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$ is equivalent to finding the Wasserstein barycenter of $\mu, \nu \in \mathbb{W}_p(\mathcal{X})$ with weights C_1, C_2. □

We are now prepared to prove the following theorem, which provides sufficient conditions for a solution to the Kantorovich MSOT problem to exist.

Theorem 7.12. *Under the assumptions of Lemma 7.10, $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$ admits a solution.*

Proof. By Lemma 7.10, we have that $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$ is equivalent to finding a Wasserstein barycenter. By Theorem 3.12, there exists some optimal transport plan $\gamma^* \in \Gamma(\mu, \nu)$ between μ and ν. If we let $\lambda := \frac{C_1}{C_1 + C_2}$, the desired barycenter in Equation (7.11) is

$$\omega = (\chi_{1-\lambda})_{\#}\gamma^*. \tag{7.13}$$

See Remark 4.10 for details. Hence the optimal transport plans $\gamma_1^* \in \Gamma(\mu, \omega)$ for $K(\mu, \omega; c_1)$ and $\gamma_2^* \in \Gamma(\omega, \nu)$ for $K(\omega, \nu; c_2)$ solve $K_{\mathcal{X}}(\mu, \nu; c_1, c_2)$, which are guaranteed to exist by Theorem 3.12. □

As a corollary, we can also extend this result to the setting of the Monge formulation of the MSOT problem.

Corollary 7.14. *Under the assumptions of Lemma 7.10, suppose additionally that $\mathcal{X} \subset \mathbb{R}^d$; μ is absolutely continuous and has bounded density with respect to the Lebesgue measure; and $\nu \in \mathcal{P}_2(\mathcal{X})$ and vanishes on small sets. Then, there exists a unique set of Monge mappings T_1, T_2 for $M_{\mathcal{X}}(\mu, \nu; c_1, c_2)$.*

Proof. By Theorem 3.7, there exists a unique Monge mapping T between μ and ν, and so by Corollary 4.7, $\omega = (\lambda \operatorname{id} + (1-\lambda)T)_{\#}\mu$ with $\lambda := \frac{C_1}{C_1+C_2}$ as before. By [AC11, Theorem 5.1], ω will also be absolutely continuous under the given assumptions, so there exists a unique Monge mapping T_2 between ω and ν. Hence, if we let $T_1 := \lambda \operatorname{id} + (1-\lambda)T$, then T_1, T_2 are unique Monge mappings for $M_{\mathcal{X}}(\mu, \nu; c_1, c_2)$. □

Remark 7.15. In Corollary 7.14, $\nu \in \mathcal{P}_2(\mathcal{X})$ is said to vanish on small sets if and only if $\nu(B) = 0$ for any Borel set $B \subset \mathcal{X} \subset \mathbb{R}^d$ with Hausdorff dimension less than or equal to $d-1$. If ν is also absolutely continuous and has bounded density with respect to the Lebesgue measure, then this condition is automatically satisfied.

Remark 7.16. We can directly extend the preceding results to the setting of two or more intermediate stages in the MSOT problem, assuming similar regularity assumptions as Lemma 7.10. In this setting, the optimal transportation plans will be such that the resulting measures on each intermediate stage all lie on the constant speed geodesic between the source and target measures μ, ν. In particular, each intermediate measure will be given by a Wasserstein barycenter with weights given by normalizing the coefficients of the cost functions. The total cost of the MSOT problem can then be found by computing the transport cost of each individual stage. Moreover, under the additional assumptions of Corollary 7.14, each barycenter will be absolutely continuous, and so we can guarantee the existence of a unique set of Monge mappings.

We now consider when the Kantorovich MSOT problem admits a solution under milder regularity assumptions. We aim to prove the following result.

Theorem 7.17. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be Polish spaces, and additionally assume that $\mathcal{Z}$ is compact. Furthermore, assume that the cost functions $c_1 \colon \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_+$ and $c_2 \colon \mathcal{Z} \times \mathcal{Y} \to \mathbb{R}_+$ are continuous. Under these conditions, for $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$, $K_{\mathcal{Z}}(\mu, \nu; c_1, c_2)$ admits a solution. In particular, $K_{\mathcal{Z}}(\mu, \nu; c_1, c_2)$ can be reduced to $K(\mu, \nu; c)$ for some cost function $c(\cdot, \cdot)$.*

We first prove the following lemma.

Lemma 7.18. *Under the assumptions of Theorem 7.17, we may define the function $c \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ given by*

$$c(x, y) = \min_{z \in \mathcal{Z}} \Big[c_1(x, z) + c_2(z, y) \Big]. \tag{7.19}$$

This function is well-defined and continuous.

Proof. Fix any $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Then, we have that the function $z \mapsto c_1(x, z) + c_2(z, y)$ is continuous. Since $\mathcal{Z}$ is compact, $c(\cdot, \cdot)$ is well-defined (i.e., a minimizer exists) by Weierstrass' theorem. Continuity of $c(\cdot, \cdot)$ results from Berge's maximum theorem [Ber63]. In particular, the function $(x, y, z) \mapsto c_1(x, z) + c_2(z, y)$ is continuous and the constant-valued correspondence $\Xi \colon \mathcal{X} \times \mathcal{Y} \rightrightarrows \mathcal{Z}$ given by $\Xi \equiv \mathcal{Z}$ is compact-valued and continuous. Thus, application of Berge's maximum theorem completes the proof. □

We additionally reference the following result, called the gluing lemma, from [Vil08].

Lemma 7.20 (Gluing Lemma, [Vil08])**.** *For Polish spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ with measures $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y}), \omega \in \mathcal{P}(\mathcal{Z})$, let $\gamma_1 \in \Gamma(\mu, \omega)$ and $\gamma_2 \in \Gamma(\omega, \nu)$. Then, there exists a coupling $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Z} \times \mathcal{Y})$ such that $(\pi_{\mathcal{X},\mathcal{Z}})_{\#}\gamma = \gamma_1$ and $(\pi_{\mathcal{Z},\mathcal{Y}})_{\#}\gamma = \gamma_2$, where $\pi_{\mathcal{X},\mathcal{Z}}$ and $\pi_{\mathcal{Z},\mathcal{Y}}$ are the canonical projections.*

We can now prove the following proposition.

Proposition 7.21. *Under the assumptions of Theorem 7.17 and with $c(\cdot, \cdot)$ as defined in Equation (7.19), we have that $K(\mu, \nu; c) \leq K_{\mathcal{Z}}(\mu, \nu; c_1, c_2)$.*

Proof. Let $(\gamma_1, \gamma_2) \in \Gamma_{\mathcal{Z}}(\mu, \nu)$. By the Gluing Lemma 7.20, there exists $\gamma \in \mathcal{P}(\mathcal{X} \times \mathcal{Z} \times \mathcal{Y})$ with $(\pi_{\mathcal{X},\mathcal{Z}})_{\#}\gamma = \gamma_1$ and $(\pi_{\mathcal{Z},\mathcal{Y}})_{\#}\gamma = \gamma_2$. We now have that

$$\int_{\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}} c_1(x, z) + c_2(z, y)\, \mathrm{d}\gamma(x, z, y) = \int_{\mathcal{X}\times\mathcal{Z}} c_1(x, z)\, \mathrm{d}\gamma_1(x, z) + \int_{\mathcal{Z}\times\mathcal{Y}} c_2(z, y)\, \mathrm{d}\gamma_2(z, y). \tag{7.22}$$

Let $\gamma' = (\pi_{\mathcal{X},\mathcal{Y}})_{\#}\gamma$ and note that $\gamma' \in \Gamma(\mu, \nu)$ by construction. By the definition of $c(\cdot, \cdot)$, we thus have that

$$c(x, y) \leq \int_{\mathcal{Z}} c_1(x, z) + c_2(z, y)\, \mathrm{d}\gamma(z|x, y). \tag{7.23}$$

Integrating both sides and substituting with Equation (7.22) implies the desired result:

$$\int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, \mathrm{d}\gamma'(x, y) \leq \int_{\mathcal{X}\times\mathcal{Z}} c_1(x, z)\, \mathrm{d}\gamma_1(x, z) + \int_{\mathcal{Z}\times\mathcal{Y}} c_2(z, y)\, \mathrm{d}\gamma_2(z, y). \tag{7.24}$$

□

We now have the machinery to prove Theorem 7.17.

Proof of Theorem 7.17. Given Proposition 7.21, it suffices to show that for an optimal transport plan $\gamma^* \in \Gamma(\mu, \nu)$ for $K(\mu, \nu; c)$, we can construct γ_1^*, γ_2^* achieving the same transport cost for $K(\mu, \nu; c_1, c_2)$. The result then follows from the fact that an optimal transport plan $\gamma^* \in \Gamma(\mu, \nu)$ is guaranteed to exist by Theorem 3.12, given the result of Lemma 7.18.

By the Kuratowski and Ryll-Nardzewski measurable selection theorem [Kec95], we can pick a measurable map $\Psi\colon \mathcal{X}\times\mathcal{Y}\to\mathcal{Z}$ which satisfies $c(x,y)=c_1(x,\Psi(x,y))+c_2(\Psi(x,y),y)$. If we define the function φ that maps $(x,y)\mapsto(x,\Psi(x,y),y)$, then we can define a coupling $\gamma\in\mathcal{P}(\mathcal{X}\times\mathcal{Z}\times\mathcal{Y})$ by $\gamma=\varphi_{\#}\gamma^*$.

We can now consider $\gamma_1^*:=(\pi_{\mathcal{X},\mathcal{Z}})_{\#}\gamma$ and $\gamma_2^*:=(\pi_{\mathcal{Z},\mathcal{Y}})_{\#}\gamma$, and we have that $(\gamma_1^*,\gamma_2^*)\in\Gamma_{\mathcal{Z}}(\mu,\nu)$. It follows that

$$\int_{\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}} c_1(x,z)+c_2(z,y)\,\mathrm{d}\gamma(x,z,y)=\int_{\mathcal{X}\times\mathcal{Z}} c_1(x,z)\,\mathrm{d}\gamma_1^*(x,z)+\int_{\mathcal{Z}\times\mathcal{Y}} c_2(z,y)\,\mathrm{d}\gamma_2^*(z,y), \tag{7.25}$$

and additionally, by construction we have that

$$\int_{\mathcal{X}\times\mathcal{Z}\times\mathcal{Y}} c_1(x,z)+c_2(z,y)\,\mathrm{d}\gamma(x,z,y)=\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\,\mathrm{d}\gamma^*(x,y). \tag{7.26}$$

Combining Equations (7.25) and (7.26) completes the proof. □

Remark 7.27. We conclude with one final remark on the existence of solutions to the Kantorovich MSOT problem. If we were to extend the Kantorovich MSOT problem to the setting of two or more intermediate stages, $\mathcal{Z}_1,\dots,\mathcal{Z}_N$ with corresponding cost functions between stages, then under similar regularity conditions as Theorem 7.17, we can extend our results to reduce the problem to an instance of the classical Kantorovich problem and guarantee the existence of solutions with similar gluing and cutting arguments.

Acknowledgments

I would like to thank my advisor Dr. Akram Aldroubi for his patient guidance and mentorship throughout my thesis work, as well as Dr. Rocío Díaz Martín for her continuous support and thoughtful suggestions, which have greatly strengthened this work. I would also like to thank Dr. Doug Hardin and Dr. Alexander Powell for generously serving on my thesis committee.

References

[AC11] Martial Agueh and Guillaume Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[Ber63] Claude Berge. *Topological Spaces: Including a Treatment of Multi-valued Functions, Vector Spaces and Convexity*. Macmillan, 1963.

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417, 1991.

[GM96] Wilfrid Gangbo and Robert McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, Sep 1996.

[Kan06a] Leonid Kantorovich. On a problem of Monge. *J. Math. Sci.(NY)*, 133:1383, 2006.

[Kan06b] Leonid Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4):1381–1382, 2006.

[Kec95] Alexander Kechris. *Classical Descriptive Set Theory*. Springer-Verlag, 1995.

[Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[PC18] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *arXiv preprint arXiv:1803.00567*, 2018.

[San15] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser Boston, 2015.

[Tar97] Robert Tarjan. Dynamic trees as search trees via euler tours, applied to the network simplex algorithm. *Mathematical Programming*, 78(2):169–177, Aug 1997.

[Vil08] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

Department of Mathematics, Vanderbilt University, Nashville, TN, USA
Email address: abi.kothapalli@vanderbilt.edu