Tropical Optimal Transport and Wasserstein Distances in Phylogenetic Tree Space

Wonjun Lee[1], Wuchen Li[1], Bo Lin[2], and Anthea Monod[3,†]

1 Department of Mathematics, University of California, Los Angeles, CA, USA
2 School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
3 Department of Applied Mathematics, Tel Aviv University, Israel

† **Corresponding e-mail: antheam@tauex.tau.ac.il**

Abstract

We study the problem of optimal transport on phylogenetic tree space from the perspective of tropical geometry, and thus define the Wasserstein-p distances for probability measures in this continuous metric measure space setting. With respect to the tropical metric—a combinatorial metric on the space of phylogenetic trees—the cases of $p = 1, 2$ are treated in detail, which give an efficient way to compute geodesics and also provide theoretical foundations for geometric insight a statistical framework on tree space. We construct explicit algorithms for the computation of the tropical Wasserstein-1 and 2 distances, and prove their convergence. Our results provide the first study of the Wasserstein distances and optimal transport on sets of phylogenetic trees. Several numerical examples are provided.

Keywords: Optimal transport; Phylogenetic tree space; Tropical metric; Wasserstein distances.

1 Introduction

Evolutionary relationships in many disciplines of biology are mathematically represented by a graph known as a phylogenetic tree. As data objects, they have been under active research for several decades. There has been extensive work on reconstructing phylogenetic trees from sequence alignments and the statistical implications of tree reconstruction based on the input alignments. However, there is an emerging need for new frameworks to study and compare sets of phylogenetic trees now that vast volumes of data can be generated relatively cheaply and quickly with modern technology. The biological implications are numerous: by passing from the genetic to the genomic scale, relationships between gene and genome evolution can now be studied, as well as the relationships between species and populations. This specific question of developing a statistically and computationally viable framework to study and compare datasets of phylogenetic trees was recently addressed by Monod et al. (2018) from the perspective of tropical geometry, where an explicit tropical metric was defined on the moduli space of phylogenetic trees, referred to as *palm tree space* (tropical tree space). We adopt this same framework in this paper and build upon it to provide a set of tools for statistical, probabilistic, and geometric studies using optimal transport theory in the space of phylogenetic trees.

Optimal transport theory arises from a question posed in economics, and specifically, in the allocation of resources. Its mathematical formulation was established in the 18th century and has been well-studied since, resulting in strong connections and mutual implications between the domains of analysis and geometry. It has also provided important results in applications and computational fields, such as computer science. An important concept arising from optimal transport is the *Wasserstein distances*, which are metrics on probability distributions. Intuitively, they measure the effort required to recover the probability mass of one distribution in terms of an efficient reconfiguration of the other. As such, Wasserstein distances broaden the scope of optimal transport theory to probability and optimization theory. Additionally, they have been exploited to move further beyond these realms to solve concrete problems in inferential statistics. Establishing Wasserstein distances for the space of phylogenetic trees thus provides a framework for a vast body of existing results in these related fields to be applicable to the important problem of statistical inference and data analysis on sets of phylogenetic trees by providing a setting for the study of probability measures

and distributions. Additionally, it provides an alternative means to study geometric aspects of the complex space of all phylogenetic trees.

Yet, there is very little previous work using Wasserstein distances in studies involving phylogenetic trees. Evans and Matsen (2012) use them to compare probability distributions giving rise to individual trees, while Kloeckner (2015) discusses geometric properties of measures on equidistant trees using Wasserstein distances. For finite spaces, Sommerfeld and Munk (2018) conduct statistical inference studies for empirical Wasserstein metrics computed from datasets. This paper offers a new contribution to these previous works in that we look beyond the setting of a single tree, and define Wasserstein distances for general sets of phylogenetic trees in an infinite metric measure space in the continuous setting—known to be a challenging framework for optimal transport theory. Numerical computations of optimal transport with various ground metrics has been recently studied in the continuous setting and shown to be efficient (Benamou et al., 2016; Li et al., 2018). We provide numerical computations and algorithms for the proposed tropical Wasserstein distances. The general case of optimal transport also provides a useful tool to study the geometry of underlying sample space; see Villani (2008) and many references therein. In particular, optimal transport theory provides a computational framework for the probability density space of phylogenetic trees, which also encodes the geometry of sample space (Lafferty, 1988; Otto and Villani, 2000; Otto, 2001). For example, geometric studies using Riemannian calculus in the optimal transport framework has been studied by Li (2018, 2019). Optimal transport in the context of algebraic geometry has been very recently studied by Çelik et al. (2019), where the Wasserstein distance between a probability distribution and an algebraic variety is minimized making use of transportation polytopes. Our work also offers a new contribution to the optimal transport domain with the study of an unusual metric that is combinatorial in nature and exhibits challenging properties unlike any previously studied.

The remainder of this paper is organized as follows. Section 2 gives an overview of tropical geometry and its coincidence with the space of all phylogenetic trees; we also review properties of the tropical metric, which endows this space with a metric structure, and give some variational forms. Section 3 overviews the problem of optimal transport and the role of the Wasserstein distances in this framework. We then define the tropical Wasserstein-p distance, with the tropical metric as a ground metric; we also give variational forms of the tropical Wasserstein distance. We study the specific cases of $p = 1$ and 2: the $p = 1$ case gives a method for computing all infinitely many tropical geodesics, while in the case of $p = 2$, the Wasserstein metric is amenable to statistical analysis by providing an inner product structure on probability measures on tree space. Section 4 gives algorithms to explicitly compute the tropical Wasserstein-p distances. We close the paper with a discussion in Section 6 on future research stemming from the work presented in this paper.

2 Tropical Geometry, Tree Space, and the Tropical Metric

In this section, we give details on our mathematical setting of the tropical geometric interpretation of phylogenetic tree space. We give the definition of the tropical metric on this space, and give alternative versions in terms of variational forms. This is the metric with respect to which we will define the tropical optimal transport problem and the tropical Wasserstein-p distances.

Tropical geometry may be seen as a subdiscipline of algebraic geometry. In the latter, the zero sets of systems of polynomial equations are studied using algebraic methods; in the former, these polynomials are defined via the *tropical semiring*, $(\mathbb{R} \cup \{-\infty\}, \boxplus, \odot)$ where addition between two elements is given by their max and multiplication is given by their sum:

$$a \boxplus b := \max(a, b),$$
$$a \odot b := a + b.$$

Notice that tropical subtraction is not defined, therefore resulting in a semiring, rather than a ring. Both operations of the semiring are commutative and associative; multiplication distributes over addition. *Tropicalization* refers to interpreting classical arithmetic operations with their tropical counterparts. Using these operations, lines, polynomials, and other more general mathematical constructions can be built, which will result in "skeletal" piecewise linear structures, hence their mathematical relevance to phylogenetic trees.

2.1 Phylogenetic Trees and Tree Space

A *phylogenetic tree* is an acyclic connected graph, $T = (V, E)$, where the set of vertices V consists of labeled terminal nodes called *leaves*, among which there is no vertex of degree 2; E is the set of *edges* or *branches* that each have positive length, representing positive evolutionary time. They can be unrooted, with labels $\{1, 2, \ldots, N\}$ on the leaf set V, or rooted, by setting the endpoint of the unique edge connecting to the leaf label 0 as the root.

Since a tree can be explicitly defined by all pairwise distances between leaves $d(i, j)$, the setting of metrics in working with trees is natural. Specifically, tree metrics are given by the set of all pairwise distances between leaves are given by the tropical product (or sum, in classical arithmetic) of positive (nonnegative) branch lengths between leaf i and leaf j, where any two leaves i and j are connected by a unique path. Note that tree metrics are more rigorous than general metrics, since they must also satisfy the following tropically quadratic inequality, in addition to the usual conditions for metrics:

$$d(i, j) \odot d(k, l) \leq d(i, k) \odot d(j, l) \boxplus d(i, l) \odot d(j, k)$$

for all distinct leaves $i, j, k, l \in \{1, 2, \ldots, N\}$. The condition for a tree metric may be equivalently expressed as follows: a metric is a tree metric if and only if the maximum among the following tropically quadratic *Plücker relations*

$$\begin{aligned} &d(i, j) \odot d(k, l) \\ &d(i, k) \odot d(j, l) \\ &d(i, l) \odot d(j, k) \end{aligned}$$

is achieved at least twice for $1 \leq i < j < k < l \leq N$. This defining condition for a tree metric is known as the *four-point condition* (Buneman, 1974). The space of all phylogenetic trees with N leaves $\mathcal{T}_N$ is the set of all $\binom{N}{2}$-tuples $\{d(i, j)\}_{1 \leq i < j \leq N}$ which satisfies the four-point condition. See Monod et al. (2018) for further details.

2.2 The Tropical Grassmannian and the Tropical Projective Torus

An observation from tropical geometry on the condition for tree metrics expressed in terms of the tropically quadratic Plücker relations is that the space of all phylogenetic trees is a tropical hypersurface. Speyer and Sturmfels (2004) develop this equivalence and identify a homeomorphism between the space of all phylogenetic trees with N leaves $\mathcal{T}_N$ and a tropical version of the Grassmannian of 2-planes in N dimensions. For $k < N$, the Grassmannian is the space of all k-dimensional subspaces of an N-dimensional vector space. It can be mapped to a particular set zero set of polynomials—specifically, a projective variety—via the Plücker embedding. The tropicalization of the Plücker embedding for the Grassmannian of 2-planes in N dimensions recovers the four-point condition, explicitly defining tree metrics for phylogenetic trees with N leaves. This important results essentially endows $\mathcal{T}_N$ with a tropical structure, which is a natural setting for increased computational efficiency (Maclagan and Sturmfels, 2015; Monod et al., 2018).

The space of all phylogenetic trees $\mathcal{T}_N$ is contained within an ambient space, known as the *tropical projective torus*, $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$, where $n + 1 := \binom{N}{2}$. The tropical projective torus is the quotient space generated by the following equivalence relation $\sim$ on $\mathbb{R}^{n+1}$:

$$x \sim y \Leftrightarrow x_1 - y_1 = x_2 - y_2 = \cdots = x_{n+1} - y_{n+1}.$$

Mathematically, $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ is constructed in the same manner as the complex torus: take a lattice $\Lambda \in \mathbb{C}^{n+1}$ as a real vector space, then the complex torus is $\mathbb{C}^{n+1}/\Lambda$. For $x \in \mathbb{R}^{n+1}$, let $\bar{x}$ be its image in $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$. The tropical projective torus identifies with $\mathbb{R}^n$ by taking representatives of the equivalence classes whose last coordinate is zero:

$$\bar{x} \mapsto (x_1 - x_{n+1},\, x_2 - x_{n+1}, \ldots,\, x_n - x_{n+1}). \tag{1}$$

We denote the image of $\bar{x}$ as $\mathbf{x}$ in $\mathbb{R}^n$. The implication of the larger ambient space of the tropical projective torus on tree space is that evolutionary time between trees may be normalized.

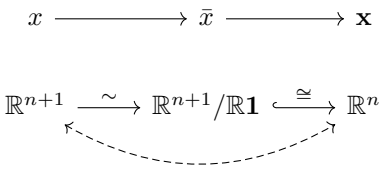


Figure 1: Diagram illustrating embedding of and relationships between Euclidean spaces and the tropical projective torus. The dashed arrow represents the isometry of the tropical metric between all three spaces.

2.3 The Tropical Metric

The tropical projective torus $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ becomes a metric space when endowed with a *generalized Hilbert projective metric* function (Cohen et al., 2004; Akian et al., 2011), which is tropical in nature and, within the context of tropical geometry, has been referred to as the *tropical metric* in recent literature (Lin et al., 2017; Monod et al., 2018). Our work here is based on the ambient tree space given by the tropical projective torus endowed with the tropical metric.

Definition 1. For a point $x \in \mathbb{R}^{n+1}$, denote its coordinates by $x_1, x_2, \ldots, x_{n+1}$ and its representation in the tropical projective torus $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ by $\bar{x}$. The *tropical metric* on $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ is given by

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) := \max_{1 \le i \le n+1} (x_i - y_i) - \min_{1 \le i \le n+1} (x_i - y_i)$$

Lemma 2. *On $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$, we have the following alternate expression for the tropical metric:*

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) = \max_{1 \le i \le j \le n+1} |(x_i - y_i) - (y_j - x_j)|.$$

Proposition 3. *(Monod et al., 2018, Proposition 17) $d_{\mathrm{tr}}(\cdot, \cdot)$ is a well-defined metric function on $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$.*

When considering the representatives of the equivalence classes as in (1), the tropical metric translates to the following on $\mathbb{R}^n$: for $\bar{x}, \bar{y} \in \mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$,

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) := \max \left\{ \max_{1 \le i < j \le n} \left|(\mathbf{x}_i - \mathbf{y}_i) - (\mathbf{x}_j - \mathbf{y}_j)\right|, \max_{1 \le i \le n} |\mathbf{x}_i - \mathbf{y}_i| \right\}.$$

Figure 1 illustrates the relationship where $\mathbb{R}^{n+1}$ identifies with $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ by the equivalence relation $\sim$; $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ then embeds into $\mathbb{R}^n$. The metric d_{tr} is defined on $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ and has a representation in $\mathbb{R}^n$; it is an isometry from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{n+1}/\mathbf{1}$ to $\mathbb{R}^n$. We denote an element in $\mathbb{R}^{n+1}$ by x, an element in $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ by $\bar{x}$, and an element in $\mathbb{R}^n$ by $\mathbf{x} = (x_1, \ldots, x_n)$.

2.4 Variational Forms of the Tropical Metric

It turns out that the tropical metric may be considered in terms of unknown functions and corresponding differential equations, which provides an alternative formulation for the tropical metric in terms of a variational form. Variational forms are useful in computational studies, since numerically, it is often easier to find solutions to variational problems rather than differential equations. As we will see further on, this turns out to be an important advantage in explicit computations of the tropical Wasserstein distances and associated results.

Proposition 4. *For $\bar{x}, \bar{y} \in \mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$, we have*

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) = \begin{pmatrix} \text{minimize} & \displaystyle\int_0^1 L_{\mathrm{tr}}\big(\mathbf{v}(t)\big)\, dt, \\ \text{subject to:} & \dfrac{d\mathbf{z}}{dt} = \mathbf{v}(t),\ \mathbf{z}(0) = \mathbf{x},\ \mathbf{z}(1) = \mathbf{y} \end{pmatrix}, \tag{2}$$

where $\mathbf{v}, \mathbf{z} : [0,1] \to \mathbb{R}^n$ *and we define the* tropical Lagrangian $L_{\mathrm{tr}}(\cdot)$ *as the tropical norm for* $\mathbf{a} \in \mathbb{R}^n$ *as follows:*

$$L_{\mathrm{tr}}(\mathbf{a}) = \|\mathbf{a}\|_{\mathrm{tr}} = \max\left(\max_{1\le i\le n}(\mathbf{a}_i), 0\right) - \min\left(\min_{1\le i\le n}(\mathbf{a}_i), 0\right) \quad \forall\ \mathbf{a} \in \mathbb{R}^n. \tag{3}$$

Proof. By definition,

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) = d_{\mathrm{tr}}(\bar{y}, \bar{x}) = \max\left(\max_{1\le i\le n}(\mathbf{x}_i - \mathbf{y}_i), 0\right) - \min\left(\min_{1\le i\le n}(\mathbf{x}_i - \mathbf{y}_i), 0\right).$$

First, let $\mathbf{z}(t) = t \cdot \mathbf{y} + (1-t) \cdot \mathbf{x}$, then $\mathbf{v}(t)$ is the constant vector $\mathbf{y} - \mathbf{x}$, and the integral $\int_0^1 \|\mathbf{v}(t)\|_{\mathrm{tr}} dt$ becomes $L_{\mathrm{tr}}(\mathbf{y} - \mathbf{x}) = d_{\mathrm{tr}}(\bar{x}, \bar{y})$. Second, in order to show that

$$\int_0^1 \|\mathbf{v}(t)\|_{\mathrm{tr}} dt \ge d_{\mathrm{tr}}(\bar{x}, \bar{y}),$$

it suffices to show that the integral is always no less than any of $|\mathbf{y}_i - \mathbf{x}_i|$ and $|(\mathbf{y}_i - \mathbf{x}_i) - (\mathbf{y}_j - \mathbf{x}_j)|$ where $1 \le i, j \le n$.

For $1 \le i \le n$, by definition of L_{tr} we have

$$\|v(t)\|_{\mathrm{tr}} \ge |\mathbf{v}(t)_i - 0| = |\mathbf{v}(t)_i|.$$

Now consider the function $f_i : [0,1] \to \mathbb{R}$ given by $f_i(t) = \mathbf{x}(t)_i$. Then $\mathbf{v}(t)_i = \dfrac{df_i}{dt}(t)$, which gives

$$\int_0^1 \mathbf{v}(t)_i dt = f_i(1) - f_i(0) = \mathbf{y}_i - \mathbf{x}_i \tag{4}$$

and

$$\int_0^1 \|\mathbf{v}(t)\|_{\mathrm{tr}} dt \ge \int_0^1 |\mathbf{v}(t)_i| dt \ge \left|\int_0^1 \mathbf{v}(t)_i dt\right| = |\mathbf{y}_i - \mathbf{x}_i|.$$

Similarly, for any $1 \le i, j \le n$, by definition of L_{tr}, we have

$$\|\mathbf{v}(t)\|_{\mathrm{tr}} \ge |\mathbf{v}(t)_i - \mathbf{v}(t)_j|.$$

By (4), we get

$$\int_0^1 \|\mathbf{v}(t)\|_{\mathrm{tr}} dt \ge \int_0^1 |\mathbf{v}(t)_i - \mathbf{v}(t)_j| dt \ge \left|\int_0^1 (\mathbf{v}(t)_i - \mathbf{v}(t)_j)\, dt\right| = |(\mathbf{y}_i - \mathbf{x}_i) - (\mathbf{y}_j - \mathbf{x}_j)|.$$

□

Example 5. When $n = 2$,

$$L_{\mathrm{tr}}(\mathbf{a}) = \begin{cases} a_1, & \text{if } a_1 \ge a_2 \ge 0; \\ a_2, & \text{if } a_2 \ge a_1 \ge 0; \\ -a_1, & \text{if } 0 \ge a_2 \ge a_1; \\ -a_2, & \text{if } 0 \ge a_1 \ge a_2; \\ a_1 - a_2, & \text{if } a_1 \ge 0 \ge a_2; \\ a_2 - a_1, & \text{if } a_2 \ge 0 \ge a_1. \end{cases}$$

The above variational form (2) of $d_{\mathrm{tr}}(\cdot,\cdot)$ may be further generalized as follows.

Corollary 6. *For* $\bar{x}, \bar{y} \in \mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$*, let* L_{tr} *be the same as in Proposition 4. For* $p > 1$*, we have*

$$d_{\mathrm{tr}}(\bar{x}, \bar{y}) = \begin{pmatrix} \text{minimize} & \left(\displaystyle\int_0^1 L_{\mathrm{tr}}\big(\mathbf{v}(t)\big)^p dt\right)^{\frac{1}{p}} \\ \text{subject to:} & \dfrac{d\mathbf{z}}{dt} = \mathbf{v}(t),\ \mathbf{z}(0) = \mathbf{x},\ \mathbf{z}(1) = \mathbf{y} \end{pmatrix}. \tag{5}$$

Proof. When $\mathbf{z}(t) = t \cdot \mathbf{y} + (1-t) \cdot \mathbf{x}$, $\mathbf{v}(t)$ is still the constant $\mathbf{y} - \mathbf{x}$ and the equality still holds. In addition, by the Hölder inequality,

$$\left(\int_0^1 \left\| \mathbf{v}(t) \right\|_{\mathrm{tr}}^p dt \right)^{\frac{1}{p}} \geq \int_0^1 \left| \|\mathbf{v}(t)\|_{\mathrm{tr}} \right| dt.$$

Hence for any $\mathbf{z} : [0,1] \to \mathbb{R}^n$ and $\mathbf{v}(t) = \frac{d\mathbf{z}}{dt}$,

$$\left(\int_0^1 \left\| \mathbf{v}(t) \right\|_{\mathrm{tr}}^p dt \right)^{\frac{1}{p}} \geq d_{\mathrm{tr}}(\mathbf{x}, \mathbf{y}).$$

□

3 Optimal Transport and the Tropical Wasserstein-p Distances

The question underlying the theory of optimal transport can be posed in a very basic and intuitive manner as follows: What is the most efficient way to move a given pile of dirt from one location to another? The total volume of the dirt must remain intact, but the shape and form of the pile may change during transportation and arrive at its location in a differently-shaped pile. This problem has been recast mathematically in various formulations with various assumptions. There is a vast literature of historical as well as technical aspects and perspectives on the optimal transport problem; see for example Villani (2003, 2008); Ambrosio and Gigli (2013) for detailed discussions. In viewing the pile of dirt as a probability measure to be transported over a space—or alternatively, one probability distribution to be transformed into another—we obtain a probabilistic and statistical perspective on the problem.

A key factor in solving the optimal transport problem is the *cost function*, which gives the cost of moving the pile of dirt, or the transporting the probability measure. Mathematically, this is generally a function of two variables—an origin or "start" location and destination or "end" location—which maps to the positive real line to give the cost, and may take into account any number of factors. In the simplest, case, however, when the cost of moving the pile of dirt from its origin to destination is nothing more than the distance between the origin and destination, the solution to the optimal transport problem yields the *Wasserstein distance* (for a fixed dimension). Intuitively, the Wasserstein distance gives the minimum cost of transforming one probability distribution into another. This minimum cost is simply the "amount of dirt" to be transported, multiplied by the mean distance it must be moved. In the case of probability distributions that contain a total mass of 1, the minimum cost is therefore simply the mean distance it must be moved. More precisely, the Wasserstein distance is a distance function for probability distributions defined on a given metric space, and thus provides a means for comparing distributions.

Although other metrics for probability distributions exist in the literature on mathematical statistics, the Wasserstein distance possesses desirable computational and intuitive properties. To illustrate a few such properties, let us consider random variables $X, Y \in \mathbb{R}^d$ distributed as $X \sim P$ and $Y \sim Q$ with densities p and q, respectively. Three commonly-used measures for distances between P and Q are total variation, $\frac{1}{2} \int |p - q|$; Hellinger, $\sqrt{\int (\sqrt{p} - \sqrt{q})^2}$; and L_2, $\int (p-q)^2$.

When comparing one discrete versus one continuous distribution, these distances yield results that are not very informative. Let P be uniform on $[0,1]$, and let Q be uniform on $\{0, 1/n, 2/n, \ldots, 1\}$. The total variation distance between these distributions is 1, which is the total size of each of the two sets, and the largest that any distance can be, while the Wasserstein distance is $1/n$.

These distances also do not take into account the underlying geometry of the space on which the distributions are defined. Consider the three densities p_1, p_2 and p_3 shown in Figure 2. We have

$$\int |p_1 - p_2| = \int |p_1 - p_3| = \int |p_2 - p_3|,$$

and similar results for the Hellinger and L_2 distances, however, intuitively, we would like to think of p_1 and p_2 being more similar and closer each other than to p_3. The Wasserstein distance is able to make this distinction.

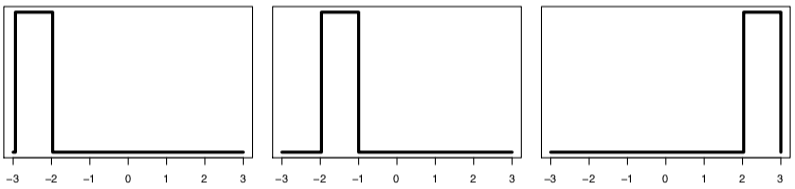


Figure 2: Three example densities p_1, p_2, p_3. This figure appears in Wasserman (2019). The total variation, Hellinger, and L_2 distances between these three densities are the same, while the Wasserstein distance between p_1 and p_2 is smaller than that between either p_1 or p_2 and p_3.

In computing a distance between distributions, we arrive at some measure of their similarity or dissimilarity, but the total variation, Hellinger, L_2, and other distances do not provide any information on how or why the distributions are qualitatively different. Perhaps the most helpful property of the Wasserstein distance is that, in addition to a measure of distance between the distributions, we also obtain a map that describes how P morphs into Q. This map is known as a *transport plan*.

In addition to the illustrative examples discussed above, there are other desirable computational and statistical properties of the Wasserstein distance, such as stability to small perturbations and a well-behaved and intuitive Wasserstein Fréchet mean. Further details and more complete discussions on statistical aspects of the Wasserstein distance can be found in Panaretos and Zemel (2019); Wasserman (2019).

Aside from statistical aspects, there also exist other analytic advantages of the Wasserstein distances, depending on the context. For instance, the Wasserstein distances' intimate connection to optimal transport problems inherently make them natural tools in these and other settings with foundations in partial differential equations.

Formalizing the Optimal Transport Problem and Defining the Wasserstein-p Distances. Monge (1781) is largely recognized to have provided the first mathematical formalization of the optimal transport problem described above, while the subsequent probabilistic reinterpretation by Kantorovich (1942) lead to a fundamental computational breakthrough that seeded the development of linear optimization. As such, the statement of the mathematical optimal transport problem is often referred to as the Monge–Kantorovich transport problem and presented in the setting of measure theory. We now give an overview of this presentation.

Definition 7. Let Ω and Ω' be separable metric spaces that are Radon spaces (that is, any probability measure on each space is a Radon measure). Let $c : \Omega \times \Omega' \to [0, \infty]$ be a Borel-measurable cost function. For $\rho^0 \in \Omega$ and $\rho^1 \in \Omega'$, the *Monge–Kantorovich transport problem* is to find a probability measure π on $\Omega \times \Omega'$ such that

$$\inf \left\{ \int_{\Omega \times \Omega'} c(x, y) \mathrm{d}\pi(x, y) \;\middle|\; \pi \in \Pi(\rho^0, \rho^1) \right\}$$

is achieved. Here, $\Pi(\rho^0, \rho^1)$ denotes the collection of all probability measures on $\Omega \times \Omega'$ with marginal measures ρ^0 on Ω and ρ^1 on Ω'.

When the cost function is lower semi-continuous, and given that Ω and Ω' are Radon spaces, $\Pi(\rho^0, \rho^1)$ is tight, and therefore a solution to the Monge–Kantorovich transport problem always exists under these conditions (e.g. Ambrosio et al., 2008). From this formulation, the Wasserstein-p distance may be defined as follows.

Definition 8. Let (Ω, d) be a separable metric Radon space. Let $p \geq 1$ and $\mathscr{P}_p(\Omega)$ be the collection of all probability measures μ on Ω with $\int_{\mathbb{R}^n} d(\mathbf{x}, \mathbf{x}_0)^p \mathrm{d}\mu(\mathbf{x}) < +\infty$; i.e. μ has finite pth moment for some $\mathbf{x}_0 \in \Omega$.

The *Wasserstein-p distance* between probability measures $\rho^0, \rho^1 \in \mathscr{P}_p(\Omega)$ is given by

$$W_p : \mathscr{P}_p(\Omega) \times \mathscr{P}_p(\Omega) \to [0, +\infty)$$
$$W_p(\rho^0, \rho^1) := \left(\inf_{\pi \in \Pi(\rho^0, \rho^1)} \int_{\Omega \times \Omega} d(\mathbf{x}, \mathbf{y})^p \mathrm{d}\pi(\mathbf{x}, \mathbf{y}) \right)^{1/p},$$

where, as before, $\Pi(\rho^0, \rho^1)$ is the collection of all probability measures on $\Omega \times \Omega$ with marginal measures ρ^0 and ρ^1 on the respective copies of Ω. Equivalently, we have

$$W_p(\rho^0, \rho^1)^p = \inf \mathbb{E}\big[d(X, Y)^p\big],$$

where $\mathbb{E}[\cdot]$ denotes the expectation, and the infimum is taken over all joint distributions of random variables X and Y with respective marginals ρ^0 and ρ^1. The metric d is referred to as the *ground metric*; the function π is known as the *transport plan*.

The transport plan $\pi(\mathbf{x}, \mathbf{y})$ is a function that describes a way to move the measure ρ^0 into ρ^1, and between locations $\mathbf{x}$ and $\mathbf{y}$; transport plans are not unique. Since the total mass moved out of a region around x must be equal to $\rho^0(\mathbf{x})\mathrm{d}\mathbf{x}$ and the total mass moved into a region around $\mathbf{x}$ must be $\rho^1(\mathbf{x})\mathrm{d}\mathbf{x}$, we have the following restrictions on a transport plan:

$$\int_{\mathbb{R}^n} \pi(\mathbf{x}, \mathbf{x}') d\mathbf{x}' = \rho^0(\mathbf{x});$$
$$\int_{\mathbb{R}^n} \pi(\mathbf{x}, \mathbf{x}') d\mathbf{x} = \rho^1(\mathbf{x}').$$

In other words, π is a joint probability distribution with marginals ρ^0 and ρ^1. The total infinitesimal mass which moves from $\mathbf{x}$ to $\mathbf{y}$, therefore, is $\pi(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$ and the cost of moving this amount of mass from $\mathbf{x}$ to $\mathbf{y}$ is $c(\mathbf{x}, \mathbf{y})\pi(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y}$. The total cost is then

$$C = \iint c(\mathbf{x}, \mathbf{y})\pi(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} = \int c(\mathbf{x}, \mathbf{y}) \mathrm{d}\pi(\mathbf{x}, \mathbf{y}).$$

The *optimal transport plan* is the π which achieves the minimal value of C:

$$C^* = \inf_{\pi \in \Pi(\rho^0, \rho^1)} \int c(\mathbf{x}, \mathbf{y}) \mathrm{d}\pi(\mathbf{x}, \mathbf{y}).$$

If the cost of a move $c(\mathbf{x}, \mathbf{y})$ is no more than the distance between the two points $d(\mathbf{x}, \mathbf{y})$, then the optimal cost value C^* is identically the Wasserstein-1 distance, W_1.

Remark 9. In the particular case where $p = 1$, the Wasserstein-1 distance is also referred to as the *Kantorovich–Rubinstein distance*, and the *earth mover's distance* (EMD) in the computer science literature.

Remark 10. The Wasserstein distances satisfy all conditions for a formal definition of a metric (e.g. Villani, 2008). If the condition of finite pth moment is relaxed, the Wasserstein distances may technically be infinite, and therefore not a metric in the strict sense.

Remark 11. For any $p \geq 1$, if (Ω, d) is a complete and separable metric space, then so too is $(\mathscr{P}_p(\Omega), W_p)$ (e.g. Villani, 2008).

A Time-Dependent Cost Function: Formulating a Hamiltonian. In formulating the above variational forms of the tropical metric (2) and (5), the notation with respect to t is not by coincidence and purposely alludes towards a dependence upon time. Within the setting of Wasserstein distances and their relation to the optimal transport problem where the ground metric is itself the cost function, intuitively, a time-dependent ground metric corresponds to a cost function where time is a cost factor.

Considering time dependence allows for a rich and alternate formulation of the optimal transport problem, which extends to the continuous displacement of measures—precisely the setting of the tropical metric on the tropical projective torus as a continuous metric measure space. However, there are certain instances where continuous displacement problems turn out to be equivalent to steady-state, time-independent problems with an alternate formulation that favors computational efficiency: this occurs when the Lagrangian L is homogeneous of degree 1 and convex.

Lemma 12. *The tropical Lagrangian L_{tr} defined in (3) is convex on $\mathbb{R}^n$. More specifically, for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $0 \le w \le 1$, we have*

$$(1-w)\|\mathbf{a}\|_{\mathrm{tr}} + w\|\mathbf{b}\|_{\mathrm{tr}} \ge \|(1-w)\mathbf{a} + w\mathbf{b}\|_{\mathrm{tr}}. \tag{6}$$

Proof. By definition,

$$\|(1-w)\mathbf{a} + w\mathbf{b}\|_{\mathrm{tr}} = \max_{1\le i\le n}\big((1-w)a_i + wb_i,\, 0\big) - \min_{1\le i\le n}\big((1-w)a_i + wb_i,\, 0\big).$$

So either there exist $1 \le j, k \le n$ such that

$$\|(1-w)\mathbf{a} + w\mathbf{b}\|_{\mathrm{tr}} = [(1-w)a_j + wb_j] - [(1-w)a_k + wb_k],$$

or there exists $1 \le j \le n$ such that

$$\|(1-w)\mathbf{a} + w\mathbf{b}\|_{\mathrm{tr}} = \|(1-w)a_j + wb_j\|_{\mathrm{tr}}.$$

Note that

$$\begin{aligned}
[(1-w)a_j + wb_j] - [(1-w)a_k + wb_k] =& (1-w)(a_j - a_k) + w(b_j - b_k)\\
\le& (1-w)\left(\max_{1\le i\le n}(a_i, 0) - \min_{1\le i\le n}(a_i, 0)\right) + w\left(\max_{1\le i\le n}(b_i, 0) - \min_{1\le i\le n}(b_i, 0)\right)\\
=& (1-w)\|\mathbf{a}\|_{\mathrm{tr}} + w\|\mathbf{b}\|_{\mathrm{tr}}.
\end{aligned}$$

We also have

$$\begin{aligned}
\|(1-w)a_j + wb_j\|_{\mathrm{tr}} \le& (1-w)|a_j| + w|b_j|\\
\le& (1-w)\left(\max_{1\le i\le n}(a_i, 0) - \min_{1\le i\le n}(a_i, 0)\right) + w\left(\max_{1\le i\le n}(b_i, 0) - \min_{1\le i\le n}(b_i, 0)\right)\\
=& (1-w)\|\mathbf{a}\|_{\mathrm{tr}} + w\|\mathbf{b}\|_{\mathrm{tr}}.
\end{aligned}$$

Hence Lemma 12 holds in either case. □

The convexity of the tropical Lagrangian L_{tr} then allows for the formulation of the *Hamiltonian* (Villani, 2008, Example 7.5) for $\mathbf{b} \in \mathbb{R}^n$ as follows:

$$\begin{aligned}
H(b) =& \sup_{\mathbf{a}\in\mathbb{R}^n} \mathbf{a}^\intercal\mathbf{b} - \frac{1}{p}\|\mathbf{a}\|_{\mathrm{tr}}^p\\
=& \sup_{\mathbf{a}}\left[\sum_{i=1}^n b_i a_i - \frac{1}{p}\left(\max\left(\max_{1\le i\le n}(a_i), 0\right) - \min\left(\min_{1\le i\le n}(a_i), 0\right)\right)^p\right].
\end{aligned} \tag{7}$$

We now explicitly compute the value of the Hamiltonian (7), which will provide concise formulations with regard to the tropical Wasserstein-p distances. For convenience, and identifying $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ with $\mathbb{R}^n$, for $\mathbf{b} \in \mathbb{R}^n$ we define

$$\zeta(\mathbf{b}) := \max_{S\subset\{1,2,\dots,n\}}\left|\sum_{i\in S} b_i\right|. \tag{8}$$

In other words, $\zeta(\mathbf{b})$ is the absolute values of the sum of either all positive b_i or all negative b_i. In particular, $\zeta(\mathbf{b}) = 0$ if and only if $p = 0$.

Example 13. When $n = 2$, we have $\mathbf{b} = (b_1, b_2)$ and

$$\zeta(\mathbf{b}) = \begin{cases} b_1 + b_2, & \text{if } b_1 \ge 0, b_2 \ge 0;\\ -b_1 - b_2, & \text{if } b_1 \le 0, b_2 \le 0;\\ b_1, & \text{if } b_1 \ge -b_2 \ge 0;\\ b_2, & \text{if } b_2 \ge -b_1 \ge 0;\\ -b_1, & \text{if } -b_1 \ge b_2 \ge 0;\\ -b_2, & \text{if } -b_2 \ge b_1 \ge 0. \end{cases}$$

Proposition 14. *The value of $H(\mathbf{b})$ is:*

(i) 0 *when* $\mathbf{b} = \mathbf{0}$, *or* $\zeta(\mathbf{b}) \le 1$ *and* $p = 1$;

(ii) ∞ *when* $\mathbf{b} \neq \mathbf{0}$ *and* $p < 1$, *or* $\zeta(\mathbf{b}) > 1$ *and* $p = 1$;

(iii) $\dfrac{p-1}{p}\zeta(\mathbf{b})^{\frac{p}{p-1}}$ *when* $b \neq 0$ *and* $p > 1$.

Proof. (i) When $\mathbf{b} = \mathbf{0}$, $\sum_{i=1}^n b_i a_i$ is always zero, and $L_{\mathrm{tr}}(\mathbf{a}) \ge 0$, so $H(\mathbf{b}) \le 0$. However, when $\mathbf{a} = \mathbf{0}$, the right-hand side of (7) is zero, so $H(\mathbf{0}) = 0$. When $\zeta(\mathbf{b}) \le 1$ and $p = 1$, let

$$u := \max_{1 \le i \le n} (a_i, 0) \ge 0 \quad \text{and} \quad v := \min_{1 \le i \le n} (a_i, 0) \le 0.$$

Then we have

$$\begin{aligned} \sum_{i=1}^n b_i a_i &= \sum_{b_i > 0} b_i a_i + \sum_{b_i < 0} b_i a_i \\ &\le \sum_{b_i > 0} b_i u + \sum_{b_i < 0} b_i v \\ &\le \zeta(\mathbf{b}) u + \zeta(\mathbf{b})(-v) \\ &= \zeta(\mathbf{b})(u - v) \le u - v. \end{aligned}$$

Hence $H(\mathbf{b}) \le 0$, and equality holds when $\mathbf{a} = \mathbf{0}$. So $H(\mathbf{b}) = 0$.

(ii) Now we may assume that $\mathbf{b} \neq \mathbf{0}$ and thus $\zeta(\mathbf{b}) > 0$. We may choose nonempty $S \subset \{1, 2, \ldots, n\}$ such that

$$\zeta(\mathbf{b}) = \left| \sum_{j \in S} b_j \right|.$$

For any $N > 0$ and each $1 \le i \le n$, we let

$$a_i = \begin{cases} \dfrac{b_i}{|b_i|} \cdot N, & \text{if } i \in S; \\ 0, & \text{if } i \notin S. \end{cases}$$

Then $\sum_{i=1}^n b_i a_i = \zeta(\mathbf{b}) \cdot N$ and the set $\{a_i \mid 1 \le i \le n\} \cup \{0\}$ is either $\{0, N\}$ or $\{0, -N\}$, so $L_{\mathrm{tr}}(\mathbf{a})$ is $N - 0$ or $0 - (-N)$, which is N. Since $\zeta(\mathbf{b}) > 0$, when $p < 1$, or $\zeta(\mathbf{b}) > 1$ and $p = 1$, we have

$$\lim_{N \to \infty} \left(\zeta(\mathbf{b}) N - \frac{1}{p} N^p \right) = \infty.$$

So $H(\mathbf{b}) = \infty$.

(iii) We denote u, v as in (i) above. Then

$$H(\mathbf{b}) \le \zeta(\mathbf{b})(u - v) - \frac{1}{p}(u - v)^p.$$

Let $s := u - v \ge 0$. We need to find the maximum of $\zeta(\mathbf{b}) s - \frac{1}{p} s^p$ when $s \ge 0$. The derivative of this function of s is

$$\zeta(\mathbf{b}) - s^{p-1}.$$

Hence the function is increasing when $0 \le s \le \zeta(\mathbf{b})^{\frac{1}{p-1}}$, and it is decreasing when $s \ge \zeta(\mathbf{b})^{\frac{1}{p-1}}$. So the maximum is attained when $s = \zeta(\mathbf{b})^{\frac{1}{p-1}}$, thus

$$H(\mathbf{b}) \le \zeta(\mathbf{b}) \cdot \zeta(\mathbf{b})^{\frac{1}{p-1}} - \frac{1}{p}\zeta(\mathbf{b})^{\frac{p}{p-1}} = \frac{p-1}{p}\zeta(\mathbf{b})^{\frac{p}{p-1}}.$$

Finally, as in (ii), we may choose nonempty $S \subset \{1, 2, \ldots, n\}$ such that

$$\zeta(\mathbf{b}) = \left| \sum_{j \in S} b_j \right|,$$

and the equality holds when

$$a_i = \begin{cases} \dfrac{b_i}{|b_i|} \cdot \zeta(\mathbf{b})^{\frac{1}{p-1}}, & \text{if } i \in S; \\ 0, & \text{if } i \notin S. \end{cases} \tag{9}$$

□

For notational convenience, we also define $\eta \colon \mathbb{R}^n \to \mathbb{R}^n$, where $\eta(\mathbf{b}) = (\eta(b)_i)_{i=1}^n$, with

$$\eta(\mathbf{b})_i := a_i = \begin{cases} \dfrac{b_i}{|b_i|} \cdot \zeta(\mathbf{b})^{\frac{1}{p-1}}, & \text{if } i \in S; \\ 0, & \text{if } i \notin S. \end{cases}$$

That is, $\eta(\mathbf{b})_i$ is defined by (9).

The Tropical Wasserstein-p Distances. We consider the tropical projective torus as a probability space with finite pth moment as follows:

$$\mathscr{P}_p(\mathbb{R}^n) = \left\{ \rho \in L^1(\mathbb{R}^n) \,:\, \int_{\mathbb{R}^n} \rho(\mathbf{x}) d\mathbf{x} = 1,\ \rho \ge 0 \right\}.$$

Within the optimal transport framework discussed above and as in Definition 8, the tropical Wasserstein-p distance is given as follows:

$$\begin{aligned} \tilde{W}_p^{\mathrm{tr}} &: \mathscr{P}_p(\mathbb{R}^n) \times \mathscr{P}_p(\mathbb{R}^n) \to [0, +\infty) \\ \tilde{W}_p^{\mathrm{tr}}(\rho^0, \rho^1)^p &:= \inf_{\pi \in \Pi(\rho^0, \rho^1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} d_{\mathrm{tr}}(\mathbf{x}, \mathbf{y})^p \mathrm{d}\pi(\mathbf{x}, \mathbf{y}), \end{aligned} \tag{10}$$

where the infimum is taken over the set of all possible joint distributions (transport plans) π with marginals ρ^0 and ρ^1, $\Pi(\rho^0, \rho^1)$. Here, the distance $\tilde{W}_p^{\mathrm{tr}}$ depends the choice of p in the linear programming formulation (10). The following alternative gives an equivalent definition of the tropical Wasserstein-p distances.

Definition 15 (Tropical Wasserstein-p distance)**.** The *tropical Wasserstein-p distance* is given by

$$W_p^{\mathrm{tr}}(\rho^0, \rho^1)^p = \inf_{\mathbf{v}, \rho} \int_0^1 \int_{\mathbb{R}^n} \left\| \mathbf{v}(t, \mathbf{x}) \right\|_{\mathrm{tr}}^p \rho(t, \mathbf{x}) d\mathbf{x} dt \tag{11a}$$

such that the following dynamical constraint or *continuity equations* hold:

$$\begin{aligned} \partial_t \rho(t, \mathbf{x}) + \nabla \cdot \big(\rho(t, \mathbf{x}) \mathbf{v}(t, \mathbf{x}) \big) &= 0, \\ \rho(0, \mathbf{x}) &= \rho^0(\mathbf{x}), \\ \rho(1, \mathbf{x}) &= \rho^1(\mathbf{x}). \end{aligned} \tag{11b}$$

Here $\| \cdot \|_{\mathrm{tr}}$ is the tropical norm, ρ^0, $\rho^1 \in \mathscr{P}_p(\mathbb{R}^n)$, ∇, $\nabla \cdot$ are gradient and divergence operators in $\mathbb{R}^n$, and the infimum is taken over all continuous density functions $\rho \colon [0, 1] \times \mathbb{R}^n \to \mathbb{R}$, and Borel vector fields $v \colon [0, 1] \times \mathbb{R}^n \to \mathbb{R}^n$.

Here, the formulation (11) given by the pairs (11a) and (11b) is known as the *Benamou–Brenier formula*, given by Benamou and Brenier (2000). As discussed in Chapter 8 of Villani (2003), when c satisfies suitable conditions, the linear programming formulation $\tilde{W}_p^{\mathrm{tr}}$ is equivalent to the dynamical formulation W_p^{tr}. In this work, we focus on the dynamical formulation (11) with $p = 1, 2$ for their concrete implications on computations of the tropical projective torus.

3.1 The Tropical Wasserstein-1 Distance

We first study the case $p = 1$. In this case, and given Lemma 12, it turns out that the tropical Wasserstein-1 distance W_1^{tr} may be recast as the following minimization problem.

Proposition 16 (Minimal Flux Formulation)**.** *By identifying $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$ with $\mathbb{R}^n$ as discussed in Section 2.3, the tropical Wasserstein-1 distance satisfies*

$$W_1^{\mathrm{tr}}(\rho^0, \rho^1) = \inf_{\mathbf{m}} \left\{ \int_{\mathbb{R}^n} \left\|\mathbf{m}(\mathbf{x})\right\|_{\mathrm{tr}} d\mathbf{x} : \rho^1(\mathbf{x}) - \rho^0(\mathbf{x}) + \nabla \cdot \mathbf{m}(\mathbf{x}) = 0 \right\}, \tag{12}$$

where the infimum is taken over all Borel flux functions $\mathbf{m}\colon \mathbb{R}^n \to \mathbb{R}^n$.

Proof. Denote

$$\mathbf{m}(\mathbf{x}) := \int_0^1 \mathbf{v}(t, \mathbf{x}) \rho(t, \mathbf{x}) dt,$$

then by Jensen's inequality, the minimizer is obtained by a time-independent solution. In other words,

$$\int_0^1 \int_{\mathbb{R}^n} \left\|\mathbf{v}(t, \mathbf{x})\right\|_{\mathrm{tr}} \rho(t, \mathbf{x}) d\mathbf{x} dt \geq \int_{\mathbb{R}^n} \left\| \int_0^1 \mathbf{v}(t, \mathbf{x}) \rho(t, \mathbf{x}) dt \right\|_{\mathrm{tr}} d\mathbf{x} = \int_{\mathbb{R}^n} \left\|\mathbf{m}(\mathbf{x})\right\|_{\mathrm{tr}} d\mathbf{x}.$$

By integrating over the time variable in the constraint, we observe that

$$\begin{aligned} &\left\{ \int_0^1 \int_{\mathbb{R}^n} \left\|\mathbf{v}(t, \mathbf{x})\right\|_{\mathrm{tr}} \rho(t, \mathbf{x}) d\mathbf{x} dt \ : \ \partial_t \rho(t, \mathbf{x}) + \nabla \cdot \big(\rho(t, \mathbf{x}) \mathbf{v}(t, \mathbf{x})\big) = 0, \ \rho(0, \mathbf{x}) = \rho^0(\mathbf{x}), \ \rho(1, \mathbf{x}) = \rho^1(\mathbf{x}) \right\} \\ \geq &\left\{ \int_{\mathbb{R}^n} \left\|\mathbf{m}(\mathbf{x})\right\|_{\mathrm{tr}} d\mathbf{x} \ : \ \rho_1(\mathbf{x}) - \rho_0(\mathbf{x}) + \nabla \cdot \mathbf{m}(\mathbf{x}) = 0 \right\}. \end{aligned}$$

Finally, choosing $\rho(t, \mathbf{x}) = (1 - t)\rho^0(\mathbf{x}) + t\rho^1(\mathbf{x})$, the above variational problem attains the minimizer. □

Concretely, $\mathbf{m}(\mathbf{x})$ is the flux vector field that assigns a vector to each point in the measure and determines how much of the mass (measure) should be moved, and in which direction.

The reformulation of the tropical Wasserstein-1 distance given in Proposition 16 has enormous computational benefits, compared to that given in Definition 8 (Li et al., 2018). Notably, the size of the optimization variable is much smaller in solving a discrete approximation; additionally, the structure of the formulation given in Proposition 16 borrows from L_1-type minimization problems, which are well-studied and for which there exist fast and simple algorithms (see references in Li et al. (2018)). We will reap these benefits in formulating explicit algorithms to compute the tropical Wasserstein-p distances for $p = 1, 2$, as discussed further on in Section 4.

Geodesics on the Tropical Projective Torus. Geodesics on the tropical projective torus are not unique (Lin et al., 2017; Monod et al., 2018). In particular, between any two given points in $\mathbb{R}^{n+1}/\mathbb{R}\mathbf{1}$, there are infinitely-many geodesics. The following result gives the explicit connection between geodesics on the tropical projective torus and the minimizer of the tropical Wasserstein-1 distance.

Proposition 17 (Minimizer of the Tropical Wasserstein-1 distance)**.** *The minimizer of the tropical Wasserstein-1 distance is given by the following pair:*

$$\begin{cases} \nabla_{\mathbf{m}} \left\|\mathbf{m}(\mathbf{x})\right\|_{\mathrm{tr}} = \nabla \Phi(\mathbf{x}) \quad \textit{if } \mathbf{m}(\mathbf{x}) > 0, \\ \rho^1(\mathbf{x}) - \rho^0(\mathbf{x}) + \nabla \cdot \mathbf{m}(\mathbf{x}) = 0. \end{cases} \tag{13}$$

Proof. The minimizer of tropical Wasserstein-1 distance may be derived as follows. Define a Lagrange multiplier $\Phi\colon \mathbb{R}^n \to \mathbb{R}$ for the equality constraint of (12), and consider the saddle point problem

$$L(\mathbf{m}, \Phi) = \int_{\mathbb{R}^n} \|\mathbf{m}(\mathbf{x})\|_{\mathrm{tr}} d\mathbf{x} + \int_{\mathbb{R}^n} \Phi(\mathbf{x})(\nabla \cdot \mathbf{m}(\mathbf{x}) + \rho^1(\mathbf{x}) - \rho^0(\mathbf{x}))d\mathbf{x}.$$

Notice that L is a convex in $\mathbf{m}$ and concave in Φ. Thus, the saddle point $(\mathbf{m}, \Phi)$ satisfies $\delta_{\mathbf{m}} L(\mathbf{m}, \Phi) = 0$, $\delta_{\Phi} L(\mathbf{m}, \Phi) = 0$. This corresponds to the equation pair (13). □

Remark 18. We notice that the first equation in (13) represents the *tropical Eikonal equation*

$$\zeta\big(\nabla\Phi(\mathbf{x})\big) = 1.$$

The tropical Eikonal equation describes the movement of each particle according to the infinitely-many geodesics under the tropical metric between ρ^0 to ρ^1. This behavior will be explored and demonstrated numerically in experiments further on in Section 5.

Proposition 19. *The set of all infinitely-many tropical geodesics is contained in a classical convex polytope.*

Proof. For any point c on a tropical geodesic connecting $a, b \in \mathbb{R}^n/\mathbb{R}\mathbf{1}$, by the definition of geodesics, we have

$$d_{\mathrm{tr}}(c, a) + d_{\mathrm{tr}}(c, b) = d_{\mathrm{tr}}(a, b).$$

So c belongs to a tropical ellipse with foci a, b. By Proposition 26 of Lin and Yoshida (2018), this set of points c on tropical geodesics is a classical convex polytope. □

3.2 The Tropical Wasserstein-2 Distance

We now consider the case where $p = 2$. Here we denote (9) by using the notation $\eta(\mathbf{b})$.

Proposition 20 (Minimizer of the Tropical Wasserstein-2 Distance)**.** *The minimizer of the tropical Wasserstein-2 distance* $\big(\mathbf{v}(t, \mathbf{x}), \rho(t, \mathbf{x})\big)$ *satisfies*

$$\mathbf{v}(t, \mathbf{x}) = \eta\big(\nabla\Phi(t, \mathbf{x})\big),$$

where $\eta\colon \mathbb{R}^n \to \mathbb{R}^n$ *is given by*

$$\eta(t, \mathbf{x})_i = \begin{cases} \dfrac{\nabla_{x_i}\Phi(t, \mathbf{x})}{|\nabla_{x_i}\Phi(t, \mathbf{x})|} \cdot \zeta\big(\nabla\Phi(t, \mathbf{x})\big) & \text{for } i \in S; \\ 0 & \text{for } i \notin S, \end{cases}$$

where S *is as in (8). Also,*

$$\begin{cases} \partial_t \rho(t, \mathbf{x}) + \nabla \cdot \big(\rho(t, \mathbf{x})\eta(\nabla\Phi(t, \mathbf{x}))\big) = 0 \\ \partial_t \Phi(t, \mathbf{x}) + \dfrac{1}{2}\zeta\big(\nabla\Phi(t, \mathbf{x})\big)^2 \le 0 \\ \rho(0, \mathbf{x}) = \rho^0(\mathbf{x}), \quad \rho(1, \mathbf{x}) = \rho^1(\mathbf{x}). \end{cases} \tag{14}$$

In particular, if $\rho(t, \mathbf{x}) > 0$*, then*

$$\partial_t \Phi(t, \mathbf{x}) + \frac{1}{2}\zeta\big(\nabla\Phi(t, \mathbf{x})\big)^2 = 0.$$

Proof. The minimizer path for the tropical Wasserstein-2 distance is derived as follows. For $p = 2$, denote $\mathbf{m}(t, \mathbf{x}) := \rho(t, \mathbf{x})v(t, \mathbf{x})$ where

$$F(\mathbf{m}, \rho) = \begin{cases} \dfrac{\|\mathbf{m}\|_{\mathrm{tr}}^2}{2\rho} & \text{if } \rho > 0; \\ 0 & \text{if } \rho = 0,\ \mathbf{m} = 0; \\ +\infty & \text{otherwise.} \end{cases}$$

Then the variational problem (11) can be reformulated as

$$\frac{1}{2}W_2^{\mathrm{tr}}(\rho_0,\rho_1)^2 = \inf_{\mathbf{m},\rho}\Big\{\int_0^1\int_{\mathbb{R}^n} F\big(\mathbf{m}(t,\mathbf{x}),\rho(t,\mathbf{x})\big)d\mathbf{x}dt: \\ \partial_t\rho(t,\mathbf{x}) + \nabla\cdot\big(\mathbf{m}(t,\mathbf{x})\big) = 0, \\ \rho(0,\mathbf{x}) = \rho_0(\mathbf{x}),\ \rho(1,\mathbf{x}) = \rho_1(\mathbf{x})\Big\}. \tag{15}$$

Notice that variational problem (15) is convex in $(\mathbf{m},\mu)$. Again, we denote the Lagrange multiplier $\Phi\colon [0,1]\times\mathbb{R}^n\to\mathbb{R}$, then we can reformulate (15) into a saddle point problem.

$$L(\mathbf{m},\rho,\Phi) = \int F(\mathbf{m},\rho) + \Phi(t,\mathbf{x})\Big(\partial_t\rho(t,\mathbf{x}) + \nabla\cdot\mathbf{m}(t,\mathbf{x})\Big)d\mathbf{x}.$$

Thus the saddle point $(\mathbf{m},\rho,\Phi)$ satisfies the system $\delta_{\mathbf{m}}L = 0$, $\delta_\rho L \geq 0$, $\delta_\Phi L = 0$, i.e.,

$$\begin{cases} \dfrac{\nabla_{\mathbf{m}}\|\mathbf{m}\|_{\mathrm{tr}}^2}{\rho} = \nabla\Phi \\ -\dfrac{\|\mathbf{m}\|_{\mathrm{tr}}^2}{2\rho} - \partial_t\Phi \geq 0. \end{cases}$$

Following Proposition 14, we obtain the minimizer of the system (14). □

4 Algorithms: Solving the Optimal Transport Problem

In this section, we design algorithms for solving the optimal transport problems that give rise to the tropical Wasserstein-p distances and geodesics. Our approach is mainly based on the G-Prox primal-dual hybrid gradient (G-Prox PDHG) algorithm (Jacobs et al., 2018), which is a modified version of Chambolle–Pock primal-dual algorithms (Chambolle and Pock, 2011; Pock and Chambolle, 2011).

We now provide a brief overview of the algorithm; see Jacobs et al. (2018); Chambolle and Pock (2011); Pock and Chambolle (2011) for further details. The classical primal-dual hybrid gradient algorithms convert the following minimization problem

$$\min_X f(KX) + g(X)$$

into the following saddle point problem

$$\min_X\max_Y\Big\{L(X,Y) = \langle KX,Y\rangle + g(X) - f^*(Y)\Big\},$$

where f and g are convex functions with respect to a variable X, f^* is a convex dual function of F, and K is a continuous linear operator. For each iteration, the algorithm performs gradient descent on the primal variable X and gradient ascent on the dual variable Y as follows:

$$\begin{cases} X^{k+1} = & \arg\min_X L(X,Y^k) + \frac{1}{2h}\|X - X^k\|^2; \\ Y^{k+1} = & \arg\max_Y L(2X^{k+1} - X^k, Y) - \frac{1}{2\tau}\|Y - Y^k\|^2, \end{cases} \tag{16}$$

where suitable norms need to be considered in the update.

For the tropical Wasserstein-1 and Wasserstein-2 distances, we apply the algorithm in (16) to (12) and (15) by setting $Y = \Phi$ and specifying

$$W_1^{\mathrm{tr}}:,\quad X = \mathbf{m},\qquad KX = \nabla\cdot\mathbf{m},\qquad g(X) = \|\mathbf{m}\|_{\mathrm{tr}},\qquad f(X) = \begin{cases} 0 & \text{if } X + \rho^1 - \rho^0 = 0, \\ \infty & \text{otherwise;} \end{cases}$$

$$W_2^{\mathrm{tr}}:,\quad X = (\mathbf{m},\rho),\quad KX = \partial_t\rho + \nabla\cdot\mathbf{m},\quad g(X) = F(\mathbf{m},\rho),\quad f(X) = \begin{cases} 0 & \text{if } X = 0, \\ \infty & \text{otherwise.} \end{cases}$$

In this paper, we use a version of the G-Prox PDHG algorithm that applies the H^1 norm in the dual variable Y update and uses the L^2 norm in the primal variable X update. This choice of norms gives us more stable and faster convergence of the algorithm than the standard PDHG algorithm (Chambolle and Pock, 2011).

4.1 Computing the Tropical Wasserstein-1 Distances

In this subsection, we consider $p = 1$. We first present the spatial discretization to compute the general Wasserstein-1 distance. Consider a uniform lattice graph $G = (V, E)$ with spacing $\Delta\mathbf{x}$ to discretize the spatial domain, where V is the vertex set

$$V = \{1, 2, \cdots, N\},$$

and E is the edge set. Here $\mathbf{i} = (i_1, \cdots, i_d) \in V$ represents a point in $\mathbb{R}^d$. Consider a discrete probability set supported on all vertices:

$$\mathcal{P}(G) = \{(\pi_{\mathbf{i}})_{\mathbf{i}=1}^N \in \mathbb{R}^N \mid \sum_{\mathbf{i}=1}^N \pi_{\mathbf{i}} = 1,\ \pi_{\mathbf{i}} \geq 0,\ \mathbf{i} \in V\},$$

where $\pi_{\mathbf{i}}$ here represents a probability at node i, i.e., $\pi_{\mathbf{i}} = \int_{C_{\mathbf{i}}} \rho(\mathbf{x})d\mathbf{x}$, $C_{\mathbf{i}}$ is a cube centered at $\mathbf{i}$ with length $\Delta\mathbf{x}$. So $\rho^0(\mathbf{x})$, $\rho^1(\mathbf{x})$ is approximated by $\pi^0 = (\pi^0_{\mathbf{i}})_{\mathbf{i}=1}^N$ and $\pi^1 = (\pi^1_{\mathbf{i}})_{\mathbf{i}=1}^N$.

We use two steps to compute the Wasserstein-1 distance on $\mathcal{P}(G)$. We first define a flux on a lattice. Denote the flux matrix $\mathbf{m} = (\mathbf{m}_{\mathbf{i}+\frac{1}{2}})_{\mathbf{i}=1}^N \in \mathbb{R}^{N\times d}$, where each component $\mathbf{m}_{\mathbf{i}+\frac{1}{2}}$ is a row vector in $\mathbb{R}^d$, i.e.,

$$\mathbf{m}_{\mathbf{i}+\frac{1}{2}} = \Big(\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v}\Big)_{v=1}^d = \left(\int_{C_{\mathbf{i}+\frac{1}{2}e_v}} m^v(\mathbf{x})d\mathbf{x}\right)_{v=1}^d,$$

where $e_v = (0, \cdots, \Delta\mathbf{x}, \cdots, 0)^\intercal$, $\Delta\mathbf{x}$ is at the vth column. In other words, if we denote $\mathbf{i} = (i_1, \cdots, i_d) \in \mathbb{R}^d$ and $\mathbf{m}(\mathbf{x}) = (\mathbf{m}^1(\mathbf{x}), \cdots, \mathbf{m}^d(\mathbf{x}))$, then

$$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v} \approx \mathbf{m}^v\Big(i_1, \cdots, i_{v-1}, i_v + \frac{1}{2}\Delta\mathbf{x}, i_{v+1}, \cdots, i_d\Big)\Delta\mathbf{x}^d.$$

We consider a zero flux condition. Specifically, if a point $\mathbf{i} + \frac{1}{2}e_v$ is outside the domain of interest Ω, we let $\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v} = 0$. Based on such a flux $\mathbf{m}$, we define a discrete divergence operator $\mathrm{div}_G(\mathbf{m}) := (\mathrm{div}_G(\mathbf{m}_{\mathbf{i}}))_{\mathbf{i}=1}^N$, where

$$\mathrm{div}_G(\mathbf{m}_{\mathbf{i}}) := \frac{1}{\Delta\mathbf{x}}\sum_{v=1}^d (\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v} - \mathbf{m}_{\mathbf{i}-\frac{1}{2}e_v}).$$

We next introduce the discrete cost functional

$$\|\mathbf{m}\| := \sum_{\mathbf{i}=1}^N \|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}\|_2 = \sum_{\mathbf{i}=1}^N \sqrt{\sum_{v=1}^d |\mathbf{m}_{\mathbf{i}+\frac{e_v}{2}}|^2}.$$

This gives rise to the following optimization problem in the tropical setting

$$\begin{aligned} \underset{\mathbf{m}}{\text{minimize}} \quad & \|\mathbf{m}\|_{\mathrm{tr}} = \sum_{\mathbf{i}=1}^N \sqrt{\sum_{v=1}^d \|\mathbf{m}_{\mathbf{i}+\frac{e_v}{2}}\|_{\mathrm{tr}}^2} \\ \text{subject to} \quad & \frac{1}{\Delta\mathbf{x}}\sum_{v=1}^d (\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v} - \mathbf{m}_{\mathbf{i}-\frac{1}{2}e_v}) + \pi^1_{\mathbf{i}} - \pi^0_{\mathbf{i}} = 0, \quad \mathbf{i} = 1, \cdots, N;\ v = 1, \cdots, d. \end{aligned} \tag{17}$$

We solve (17) by looking at its saddle point structure. Denote $\Phi = (\Phi_{\mathbf{i}})_{\mathbf{i}=1}^N$ as the Lagrange multiplier of (17), thus we have

$$\min_{\mathbf{m}} \max_{\Phi} \quad L(\mathbf{m}, \Phi) := \min_{\mathbf{m}} \max_{\Phi} \quad \|\mathbf{m}\|_{\mathrm{tr}} + \Phi^\intercal(\mathrm{div}_G(\mathbf{m}) + \pi^1 - \pi^0). \tag{18}$$

Saddle point problems such as (18) are well studied by the first-order primal-dual hybrid gradient (PDHG) algorithm. By implementing the G-Prox PDHG algorithm, the iteration steps are as follows:

$$\begin{cases} \mathbf{m}^{k+1} = & \arg\min_{\mathbf{m}} L(\mathbf{m}, \Phi^k) + \frac{1}{2h}\|\mathbf{m} - \mathbf{m}^k\|_{L^2}^2, \\ \Phi^{k+1} = & \arg\max_{\Phi} L(2\mathbf{m}^{k+1} - \mathbf{m}^k, \Phi) - \frac{1}{2\tau}\|\Phi - \Phi^k\|_{H^1}^2, \end{cases} \tag{19}$$

where h, τ are two small step sizes, $\|\mathbf{m}-\mathbf{m}^k\|_{L^2}^2 = \sum_{\mathbf{i}=1}^{N}\sum_{v=1}^{d}\left(\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v} - \mathbf{m}^k_{\mathbf{i}+\frac{1}{2}e_v}\right)^2\Delta\mathbf{x}$ and $\|\Phi-\Phi^k\|_{H^1}^2 = \sum_{\mathbf{i}=1}^{N}\left(\nabla_G\Phi_{\mathbf{i}} - \nabla_G\Phi^k_{\mathbf{i}}\right)^2\Delta\mathbf{x}$. These steps are alternating a gradient ascent in the dual variable Φ and a gradient descent in the primal variable $\mathbf{m}$.

It turns out that iteration (19) can be solved by simple explicit formulae. Since the unknown variables $\mathbf{m}$, Φ are component-wise separable in this problem, each of its components $\mathbf{m}_{\mathbf{i}+\frac{1}{2}}$, $\Phi_{\mathbf{i}}$ can be independently obtained by solving (19). First, notice that

$$\arg\min_{\mathbf{m}} L(\mathbf{m},\Phi^k)+\frac{1}{2h}\|\mathbf{m}-\mathbf{m}^k\|_{L^2}^2 = \arg\min_{\mathbf{m}_{\mathbf{i}+\frac{1}{2}}}\sum_{\mathbf{i}=1}^{N}\left(\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}\|_{\mathrm{tr}} - \left(\nabla_G\Phi^k_{\mathbf{i}+\frac{1}{2}}\right)^{\intercal}\mathbf{m}_{\mathbf{i}+\frac{1}{2}} + \frac{1}{2h}\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}-\mathbf{m}^k_{\mathbf{i}+\frac{1}{2}}\|_{L^2}^2\right),$$

where $\nabla_G\Phi^k_{\mathbf{i}+\frac{1}{2}} := \frac{1}{\Delta\mathbf{x}}(\Phi^k_{\mathbf{i}+e_v}-\Phi^k_{\mathbf{i}})_{v=1}^d$. The first iteration in (19) has an explicit solution, which is:

$$\mathbf{m}^{k+1}_{i+\frac{1}{2}} = \mathrm{shrink}_{\mathrm{tr}}(\mathbf{m}^k_{\mathbf{i}+\frac{1}{2}} + h\nabla_G\Phi^k_{\mathbf{i}+\frac{1}{2}}, h),$$

where the shrink operator is a projection operation to the unit ball with norm $\|\cdot\|_{\mathrm{tr}}$; its exact formulation is given further on in Proposition 22.

Second, consider

$$\arg\max_{\Phi} L(2\mathbf{m}^{k+1}-\mathbf{m}^k,\Phi)-\frac{1}{2\tau}\|\Phi-\Phi^k\|_2^2 = \arg\max_{\Phi}\sum_{\mathbf{i}=1}^{N}\max_{\Phi_{\mathbf{i}}}\left\{\Phi_{\mathbf{i}}\left[\mathrm{div}_G(2\mathbf{m}^{k+1}_{\mathbf{i}}-\mathbf{m}^k_i)+\pi^1_{\mathbf{i}}-\pi^0_{\mathbf{i}}\right]-\frac{1}{2\tau}\|\Phi_{\mathbf{i}}-\Phi^k_{\mathbf{i}}\|_{H^1}^2\right\}.$$

Thus the second iteration in (19) becomes

$$\Phi^{k+1}_{\mathbf{i}} = \Phi^k_{\mathbf{i}} + \tau(-\Delta_G)^{-1}\left(\mathrm{div}_G(2\mathbf{m}^{k+1}_{\mathbf{i}}-\mathbf{m}^k_{\mathbf{i}})+\pi^1_{\mathbf{i}}-\pi^0_{\mathbf{i}}\right).$$

where $\Delta_G = \mathrm{div}_G\cdot\nabla_G$ is the discrete Laplacian operator.

We are now ready to state our algorithm.

G-Prox Primal-Dual Method for Computing the Tropical Wasserstein-1 Distance

Input: Discrete probabilities p^0, p^1;
Initial guess of $\mathbf{m}^0$, step size h, τ, tolerance ϵ.

Output: $\mathbf{m}$ and $W_1^{\mathrm{tr}}(\rho^0,\rho^1)$.

---

1. **while** the relative error of $\|\mathbf{m}\|_{\mathrm{tr}} > \epsilon$
2. $\quad\mathbf{m}^{k+1}_{\mathbf{i}+\frac{1}{2}} = \mathrm{shrink}_{\mathrm{tr}}(\mathbf{m}^k_{\mathbf{i}+\frac{1}{2}} + h\nabla_G\Phi^k_{\mathbf{i}+\frac{1}{2}}, h)$;
3. $\quad\Phi^{k+1}_{\mathbf{i}} = \Phi^k_{\mathbf{i}} + \tau(-\Delta_G)^{-1}\left(\mathrm{div}_G(2\mathbf{m}^{k+1}_{\mathbf{i}}-\mathbf{m}^k_{\mathbf{i}})+\pi^1_{\mathbf{i}}-\pi^0_{\mathbf{i}}\right)$;
4. **end**

Remark 21. The relative error at iteration k is given by $\dfrac{|\|\mathbf{m}^k\|_{\mathrm{tr}} - \|\mathbf{m}^{k-1}\|_{\mathrm{tr}}|}{\|\mathbf{m}^{k-1}\|_{\mathrm{tr}}}$.

In the algorithm, we require the shrink operator with respect to the tropical metric, $\mathrm{shrink}_{\mathrm{tr}}$, which is given in the following result.

Proposition 22. *Let $h>0$ and $b_1\geq b_2\geq\ldots\geq b_k\geq 0 > b_{k+1}\geq\cdots\geq b_n$. We denote*

$$u_i = b_i\ \forall\ 1\leq i\leq k,\quad u_{k+1}=0$$

and

$$v_i = -b_{n+1-i}\ \forall\ 1\leq i\leq n-k,\quad v_{n-k+1}=0.$$

Suppose

$$j_1 = \begin{cases} \max\left\{1 \le j \le k+1 \mid \sum_{i=1}^{j}(u_i - u_j) < 1\right\}, & \text{if } k \ge 1, \\ 0, & \text{if } k = 0, \end{cases} \quad \text{and} \quad l_1 = \max(j_1, k);$$

and

$$j_2 = \begin{cases} \max\left\{1 \le j \le n-k+1 \mid \sum_{i=1}^{j}(v_i - v_j) < 1\right\}, & \text{if } k \le n-1, \\ 0, & \text{if } k = n, \end{cases} \quad \text{and} \quad l_2 = \max(j_2, n-k).$$

We let

$$t_1 = \begin{cases} \frac{\left(\sum_{i=1}^{j_1} u_i\right) - 1}{j_1}, & \text{if } 1 \le j_1 \le k; \\ 0, & \text{otherwise.} \end{cases}$$

and

$$t_2 = \begin{cases} \frac{\left(\sum_{i=1}^{j_2} v_i\right) - 1}{j_2}, & \text{if } 1 \le j_2 \le n-k; \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$shrink_{tr}(\mathbf{b}, h) := argmin_{\mathbf{a}\in\mathbb{R}^n}\left[\frac{\sum_{i=1}^{n} a_i^2}{2h} + \|\mathbf{a}\|_{\text{tr}} - \sum_{i=1}^{n} b_i \cdot a_i\right] \tag{20}$$

is the following unique point $\mathbf{x} \in \mathbb{R}^n$*, where*

$$x_i = \begin{cases} h \cdot t_1, & \text{if } i \le l_1; \\ h \cdot b_i, & \text{if } l_1 < i < n+1-l_2; \\ -h \cdot t_2, & \text{if } i \ge n+1-l_2. \end{cases}$$

Proof. Note that by definition of t_1, t_2, they are bounded by all of u_i with $i \le j_1$ and all of v_i with $i \le j_2$, respectively. In addition, we have

$$\sum_{i=1}^{l_1}(u_i - t_1) \le 1 \tag{21}$$

and

$$\sum_{i=1}^{l_2}(v_i - t_2) \le 1. \tag{22}$$

Now we claim that

$$\|\mathbf{a}\|_{\text{tr}} \ge \sum_{i=1}^{l_1}(u_i - t_1)\cdot a_i - \sum_{i=1}^{l_2}(v_i - t_2)\cdot a_{n+1-i}. \tag{23}$$

In fact, (21) implies that

$$\sum_{i=1}^{l_1}(u_i - t_1)\cdot a_i \le \max_{1\le i\le j_1} a_i.$$

We have that (22) implies that

$$\sum_{i=1}^{l_2}(v_i - t_2)\cdot a_{n+1-i} \ge \left[\sum_{i=1}^{l_2}(v_i - t_2)\right]\cdot \min_{1\le i\le j_2} a_{n+1-i} \ge \min(0, \min_{1\le i\le j_2} a_{n+1-i}).$$

Hence, the right-hand side of (23)

$$\sum_{i=1}^{l_1}(u_i-t_1)\cdot a_i - \sum_{i=1}^{l_2}(v_i-t_2)\cdot a_{n+1-i} \le \max_{1\le i\le j_1} a_i - \min(0, \min_{1\le i\le j_2} a_{n+1-i}) = \max_{1\le i_1\le j_1, 1\le i_2\le j_2}(a_{i_1}, a_{i_1}-a_{i_2}) \le \|a\|_{\mathrm{tr}}.$$

So our claim is proved.

Since $h>0$ is a constant, we can multiply the objective function in (20) by $2h$. Now, this new function is greater than or equal to

$$\begin{aligned}&\sum_{i=1}^{n} a_i^2 + 2h\left(\sum_{i=1}^{l_1}(u_i-t_1)\cdot a_i - \sum_{i=1}^{l_2}(v_i-t_2)\cdot a_{n+1-i}\right) - 2h\sum_{i=1}^{n} b_i\cdot a_i\\ =&\sum_{i=1}^{n} a_i^2 - \sum_{i=1}^{l_1} 2ht_1\cdot a_i + \sum_{i=1}^{l_2} 2ht_2\cdot a_{n+1-i} - 2h\sum_{i=l_1+1}^{n-l_2} b_i\cdot a_i.\end{aligned}$$

The global minimum of the last quadratic polynomial is attained exactly at the point $\mathbf{x}$ in Proposition 22, so we have a lower bound for the new objective function, which is its value when $\mathbf{a}=\mathbf{x}$. Finally, we note that the equality of (23) is attained at $\mathbf{x}$, so this value is actually attained by $\mathbf{a}=\mathbf{x}$. □

Example 23. When $n=2$, given $(b_1,b_2)\in\mathbb{R}^2$, suppose $x_1=f_1(b_1,b_2)$ and $x_2=f_2(b_1,b_2)$, then the shrink operator is given as follows.

$\mathbf{b}$	k	j_1	j_2	l_1	l_2	t_1	t_2	x_1	x_2
$b_1\ge b_2+1, b_2\ge 0$	2	1	0	1	0	b_1-1	0	$h(b_1-1)$	b_2
$b_1<b_2+1, b_1\ge 1-b_2, b_1\ge b_2$	2	2	0	2	0	$\frac{b_1+b_2-1}{2}$	$\frac{b_1+b_2-1}{2}$	$h\frac{b_1+b_2-1}{2}$	$h\frac{b_1+b_2-1}{2}$
$b_1<1-b_2, b_1\ge b_2\ge 0$	2	3	0	2	0	0	0	0	0
$1>b_1\ge 0, 0\ge b_2>-1$	1	2	2	1	1	0	0	0	0
$1>b_1\ge 0, b_2\le -1$	1	2	1	1	1	0	$-b_2-1$	0	$h(b_2+1)$
$b_1\ge 1, 0\ge b_2>-1$	1	1	2	1	1	b_1-1	0	$h(b_1-1)$	0
$b_1\ge 1, b_2\le -1$	1	1	1	1	1	b_1-1	$-b_2-1$	$h(b_1-1)$	$h(b_2+1)$
$0\ge b_1\ge b_2$	0	1		1		0		$-f_2(-b_2,-b_1)$	$-f_1(-b_2,-b_1)$
$b_1<b_2$								$f_2(b_2,b_1)$	$f_1(b_2,b_1)$

Figure 3: The operator $\mathrm{shrink}_{\mathrm{tr}}$ when $n=2$

Remark 24. Proposition 22 provides an algorithm to compute the shrink. Suppose we have $h>0$ and $\mathbf{a}_0,\mathbf{b}\in\mathbb{R}^n$ and we would like find

$$\mathrm{shrink}_{\mathrm{tr}}(\mathbf{a}_0+h\mathbf{b},h) = \mathrm{argmin}_{\mathbf{a}\in\mathbb{R}^n}\frac{|\mathbf{a}-\mathbf{a}_0|_2^2}{2h} + \|\mathbf{a}\|_{\mathrm{tr}} - \sum_{i=1}^{n} b_i\cdot a_i.$$

Note that

$$|\mathbf{a}-\mathbf{a}_0|_2^2 = \sum_{i=1}^{n}(a_i-a_{0i})^2 = \sum_{i=1}^{n} a_i^2 - \sum_{i=1}^{n} 2a_{0i}\cdot a_i + \text{ constant}.$$

Then we let $\mathbf{b}'=\mathbf{b}+\frac{\mathbf{a}_0}{h}$, the optimization problem becomes the one in Proposition 22 for $\mathbf{b}'$ and h after sorting the coordinates of $\mathbf{b}'$.

4.2 Computing the Tropical Wasserstein-2 Distances

We now present an algorithm to compute the tropical Wasserstein-2 distance in the tropical projective torus $\mathbb{R}^3/\mathbb{R}\mathbf{1}$ identified with $\mathbb{R}^2$. Consider the same uniform lattice graph on a domain $\Omega\subset\mathbb{R}^2$ as in the case for

the tropical Wasserstein-1 distance. Define the following matrices

$$\boldsymbol{\rho} = \left(\boldsymbol{\rho}_i^n\right)_{\mathbf{i},n=1}^{N_x,N_t}$$
$$\mathbf{m} = \left(\mathbf{m}_{i+\frac{1}{2}e_v}^n\right)_{v,\mathbf{i},n=1}^{d,N_x,N_t}$$

where the time interval is discretized uniformly with N_t points, and N_x is the number of vertices from a uniform lattice graph. Here we assume Neumann boundary conditions for $\boldsymbol{\rho}$: $\dfrac{\partial \rho}{\partial \hat{\mathbf{n}}} = 0$ on $\partial\Omega$, where $\hat{\mathbf{n}}$ is a outward normal vector. Given initial densities ρ_0 and ρ_1, the boundary conditions for ρ at $t = 0$ and $t = 1$ are

$$\left(\boldsymbol{\rho}_{\mathbf{i}}^1\right)_{\mathbf{i}=1}^{N_x} = \rho_0 \quad \text{and} \quad \left(\boldsymbol{\rho}_{\mathbf{i}}^{N_t}\right)_{\mathbf{i}=1}^{N_x} = \rho_1.$$

Define $\Delta t := \frac{1}{N_t}$. We can reformulate the minimization problem (15) into a discretization formulation.

$$\begin{aligned} \underset{\mathbf{m}}{\text{minimize}} \quad & \sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x} \frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\text{tr}}^2}{2\boldsymbol{\rho}_{\mathbf{i}}^n}\Delta\mathbf{x}\Delta t \\ \text{subject to} \quad & \partial_t \boldsymbol{\rho}_{\mathbf{i}}^n + \text{div}_G(\mathbf{m}_{\mathbf{i}}^n) = 0, \quad \mathbf{i} = 1,\cdots,N_x;\ n = 1,\cdots,N_t \\ & \left(\boldsymbol{\rho}_{\mathbf{i}}^1\right)_{\mathbf{i}=1}^{N_x} = \rho_0, \\ & \left(\boldsymbol{\rho}_{\mathbf{i}}^{N_t}\right)_{\mathbf{i}=1}^{N_x}. \end{aligned} \tag{24}$$

where

$$\partial_t \boldsymbol{\rho}_{\mathbf{i}}^n = \begin{cases} \frac{1}{\Delta t}(\boldsymbol{\rho}_{\mathbf{i}}^{n+1} - \boldsymbol{\rho}_{\mathbf{i}}^n) & \text{for } n = 1 \\ \frac{1}{2\Delta t}(\boldsymbol{\rho}_{\mathbf{i}}^{n+1} - \boldsymbol{\rho}_{\mathbf{i}}^{n-1}) & \text{for } n = 2,\cdots,N_t - 1 \\ \frac{1}{\Delta t}(\boldsymbol{\rho}_{\mathbf{i}}^n - \boldsymbol{\rho}_{\mathbf{i}}^{n-1}) & \text{for } n = N_t \end{cases}$$

and

$$\text{div}_G(\mathbf{m}_{\mathbf{i}}^n) = \frac{1}{\Delta x}\sum_{v=1}^{2}(\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v}^n - \mathbf{m}_{\mathbf{i}-\frac{1}{2}e_v}^n) \qquad \text{for } n = 1,\cdots,N_t$$

In $\mathbb{R}^2$, using (3), we can calculate the tropical norm of the flux function $\mathbf{m}$ by considering the six different cases based on $\{\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_v}\}_{v=1}^2$. The tropical norm of $\mathbf{m}$ is given as follows:

$\mathbf{m}_{\mathbf{i}+\frac{1}{2}}$	$\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}\|_{\text{tr}}$
$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1} > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2} > 0$	$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}$
$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2} > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1} > 0$	$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}$
$0 > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2} > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}$	$-\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}$
$0 > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1} > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}$	$-\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}$
$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1} > 0 > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}$	$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1} - \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}$
$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2} > 0 > \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}$	$\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2} - \mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}$

Figure 4: Tropical norm when $n = 2$

Denote $\Phi = (\Phi_i^n \mathbf{i}_{\mathbf{i}=1}^{N_x}{}_{n=1}^{N_t}$ as the Lagrange multiplier which satisfies the Neumann boundary condition on the boundary of the domain. The minimization problem (24) can be reformulated as a saddle point problem.

$$\min_{\mathbf{m},\boldsymbol{\rho}}\max_{\Phi} \quad L(\mathbf{m},\boldsymbol{\rho},\Phi) := \min_{\mathbf{m},\boldsymbol{\rho}}\max_{\Phi} \quad \sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x} \frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\text{tr}}^2}{2\boldsymbol{\rho}_{\mathbf{i}}^n} + \Phi_{\mathbf{i}}^n\left(\partial_t \rho_{\mathbf{i}}^n + \text{div}_G\left(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\right)\right). \tag{25}$$

Again, we implement G-Prox PDHG to solve the problem as follows:

$$\begin{cases} \boldsymbol{\rho}^{k+1} = \arg\min_{\boldsymbol{\rho}} \quad L(\mathbf{m}^k,\boldsymbol{\rho},\Phi^k) + \frac{1}{2\tau}\|\boldsymbol{\rho} - \boldsymbol{\rho}^k\|_{L^2(\Omega\times[0,1])}^2, \\ m^{k+1} = \arg\min_{\mathbf{m}} \quad L(\mathbf{m},\boldsymbol{\rho}^{k+1},\Phi^k) + \frac{1}{2\tau}\|\mathbf{m} - \mathbf{m}^k\|_{L^2(\Omega\times[0,1])}^2, \\ \Phi^{k+1} = \arg\max_{\Phi} \quad L(2\mathbf{m}^{k+1} - \mathbf{m}^k, 2\boldsymbol{\rho}^{k+1} - \boldsymbol{\rho}^k,\Phi) - \frac{1}{2h}\|\Phi - \Phi^k\|_{H^1(\Omega\times[0,1])}^2, \end{cases} \tag{26}$$

where h, τ are two small step sizes and

$$\|\boldsymbol{\rho}-\boldsymbol{\rho}^k\|_{L^2}^2=\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\left(\boldsymbol{\rho}_{\mathbf{i}}^n-(\boldsymbol{\rho}_{\mathbf{i}}^n)^k\right)^2\Delta\mathbf{x}\Delta t$$
$$\|\Phi-\Phi^k\|_{H^1}^2=\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\left((\partial_t\Phi_{\mathbf{i}}^n-\partial_t(\Phi_{\mathbf{i}}^n)^k)^2+\|\nabla_G\Phi_{\mathbf{i}}^n-\nabla_G(\Phi_{\mathbf{i}}^n)^k\|^2\right)\Delta\mathbf{x}\Delta t.$$

From (26), each component $\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n$, $\boldsymbol{\rho}_{\mathbf{i}}^n$, and $\Phi_{\mathbf{i}}^n$ can be obtained. From the first iteration,

$$\begin{aligned}\boldsymbol{\rho}^{k+1}=&\arg\min_{\boldsymbol{\rho}}\quad L(\mathbf{m}^k,\boldsymbol{\rho},\Phi^k)+\frac{1}{2\tau}\|\boldsymbol{\rho}-\boldsymbol{\rho}^k\|_{L^2}^2\\ =&\arg\min_{\boldsymbol{\rho}}\quad\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\frac{\|(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^k\|_{\mathrm{tr}}^2}{2\boldsymbol{\rho}_{\mathbf{i}}^n}+(\Phi_{\mathbf{i}}^n)^k\partial_t\boldsymbol{\rho}_{\mathbf{i}}^n+\frac{1}{2\tau}\|\boldsymbol{\rho}-\boldsymbol{\rho}^k\|_{L^2}^2\end{aligned}$$

We calculate the minimizer by differentiating the equation with respect to $\boldsymbol{\rho}_{\mathbf{i}}^n$. The minimizer $\boldsymbol{\rho}^{k+1}$ is a positive root of the following cubic polynomial:

$$-\frac{\|(\mathbf{m}_{\mathbf{i}}^n)^k\|_{\mathrm{tr}}^2}{2((\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1})^2}-\partial_t(\Phi_{\mathbf{i}}^n)^k+\frac{1}{\tau}\left((\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}-(\boldsymbol{\rho}_{\mathbf{i}}^n)^k\right)=0.$$

Thus, we can calculate the root by using a cubic solver.

$$(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}=\mathrm{root}^+\left(-(\boldsymbol{\rho}_{\mathbf{i}}^n)^k-\tau\partial_t(\Phi_{\mathbf{i}}^n)^k,0,-\frac{\tau}{2}\|(\mathbf{m}_{\mathbf{i}}^n)^k\|_{\mathrm{tr}}^2\right),$$

where $\mathrm{root}^+(a,b,c)$ is a solution for a cubic polynomial $x^3+ax^2+bx+c=0$.

We can reformulate the second iteration as follows:

$$\begin{aligned}\mathbf{m}^{k+1}&=\arg\min_{\mathbf{m}}\quad L(\mathbf{m},\boldsymbol{\rho}^{k+1},\Phi^k)+\frac{1}{2\tau}\|\mathbf{m}-\mathbf{m}^k\|_{L^2}^2\\ &=\arg\min_{\mathbf{m}}\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\mathrm{tr}}^2}{2(\rho_{\mathbf{i}}^n)^{k+1}}+\Phi_{\mathbf{i}}^n\mathrm{div}_G(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)+\frac{1}{2\tau}\|\mathbf{m}-\mathbf{m}^k\|_{L^2}^2\\ &=\arg\min_{\mathbf{m}}\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\mathrm{tr}}^2}{2(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}}-\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\nabla_G\Phi_{\mathbf{i}}^n+\frac{1}{2\tau}\|\mathbf{m}-\mathbf{m}^k\|_{L^2}^2\\ &=\arg\min_{\mathbf{m}}\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x}\frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\mathrm{tr}}^2}{2(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}}+\frac{1}{2\tau}\|\mathbf{m}-\mathbf{m}^k-\tau\nabla_G\Phi\|_{L^2}^2.\end{aligned}$$

Differentiating the equation with respect to $\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n$, we obtain the following expression:

$$\frac{\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\mathrm{tr}}\nabla_G\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\|_{\mathrm{tr}}}{(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}}+\frac{1}{\tau}(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n-(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^k-\tau\nabla_G\Phi_{\mathbf{i}}^n)=0.$$

Solving this expression gives an explicit solution for $(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^{k+1}$:

$$(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^{k+1}=\boldsymbol{F}\left((\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^k+\tau\nabla_G(\Phi_{\mathbf{i}}^n)^k,\,\tau/(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}\right).\tag{27}$$

Let $\mu=\tau/(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}$ and $c=(c_1,c_2)$ be

$$\begin{aligned}c_1&=(\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_1}^n)^k+\tau\nabla_{x_1}(\Phi_{\mathbf{i}+\frac{1}{2}e_1}^n)^k\\ c_2&=(\mathbf{m}_{\mathbf{i}+\frac{1}{2}e_2}^n)^k+\tau\nabla_{x_2}(\Phi_{\mathbf{i}+\frac{1}{2}e_2}^n)^k.\end{aligned}$$

The function $\boldsymbol{F}(c,\mu)$ is then given as follows:

c_1, c_2, μ	$\boldsymbol{F}(c,\mu)$
$c_2 > (1+\mu)c_1 > 0$ or $c_2 < (1+\mu)c_1 < 0$	$\left(c_1, \frac{c_2}{1+\mu}\right)$
$c_1 > (1+\mu)c_2 > 0$ or $c_1 < (1+\mu)c_2 < 0$	$\left(\frac{c_1}{1+\mu}, c_2\right)$
$-\frac{\mu}{1+\mu}c_1 > c_2 > -\frac{1+\mu}{\mu}c_1$ or $-\frac{\mu}{1+\mu}c_1 < c_2 < -\frac{1+\mu}{\mu}c_1$	$\left(\frac{(1+\mu)c_1+\mu c_2}{1+2\mu}, \frac{(1+\mu)c_2+\mu c_1}{1+2\mu}\right)$
$-\frac{\mu}{1+\mu}c_1 > c_2 > 0$ or $-\frac{\mu}{1+\mu}c_1 < c_2 < 0$	$\left(\frac{c_1}{1+\mu}, 0\right)$
$c_2 > -\frac{1+\mu}{\mu}c_1 > 0$ or $c_2 < -\frac{1+\mu}{\mu}c_1 < 0$	$\left(0, \frac{c_2}{1+\mu}\right)$
$(1+\mu)c_1 > c_2 > \frac{1}{1+\mu}c_1$ or $(1+\mu)c_1 < c_2 < \frac{1}{1+\mu}c_1$	$\left(\frac{c_1+c_2}{2+\mu}, \frac{c_1+c_2}{2+\mu}\right)$

Figure 5: The definition of $\boldsymbol{F}(c,\mu)$

Similarly, we get an explicit formula of Φ^{k+1} from the third iteration.

$$(\Phi_{\mathbf{i}}^n)^{k+1} = (\Phi_{\mathbf{i}}^n)^k + h(-\Delta_{t,G})^{-1}\Big(\partial_t(2(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1} - (\boldsymbol{\rho}_{\mathbf{i}}^n)^k) + \mathrm{div}_{t,G}\Big(2\Big(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\Big)^{k+1} - \Big(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\Big)^k\Big)\Big)$$

for $\mathbf{i} = 1, \cdots N_x$ and $n = 1, \cdots, N_t$. Here, $\Delta_{t,G} = \partial_{tt} + \Delta_G$ is the discrete Laplacian operator over time and space.

Now, define $E^k := \sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x} \frac{\big\|\big(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\big)^k\big\|_{\mathrm{tr}}^2}{2(\boldsymbol{\rho}_{\mathbf{i}}^n)^k}$. Then the relative error at iteration k is calculated as $\frac{|E^k - E^{k-1}|}{|E^{k-1}|}$.

We are now ready to present our algorithm to compute the tropical Wasserstein-2 metric.

G-Prox Primal-Dual Method for the Tropical Wasserstein-2 Distance

Input: Discrete probabilities ρ^0, ρ^1;
Initial guess of $\boldsymbol{\rho}$, $\mathbf{m}$, Φ, step size τ, h,tolerance ϵ

Output: $\mathbf{m}$ and $W_2^{\mathrm{tr}}(\rho^0, \rho^1)$.

---

1. **while** the relative error of $\sum_{n=1}^{N_t}\sum_{\mathbf{i}=1}^{N_x} \frac{\big\|\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n\big\|_{\mathrm{tr}}^2}{2\boldsymbol{\rho}_{\mathbf{i}}^n} > \epsilon$
2. $\quad(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1} = \mathrm{root}^+\Big(-(\boldsymbol{\rho}_{\mathbf{i}}^n)^k - \tau\partial_t(\Phi_{\mathbf{i}}^n)^k,\, 0,\, -\frac{\tau}{2}\|(\mathbf{m}_{\mathbf{i}}^n)^k\|_{\mathrm{tr}}^2\Big)$;
3. $\quad(\mathbf{m}_{\mathbf{i}+\frac{1}{2}}^n)^{k+1} = \boldsymbol{F}\big((\mathbf{m}_{i+\frac{1}{2}}^n)^k + \tau\nabla_G(\Phi_{\mathbf{i}}^n)^k,\, \tau/(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1}\big)$;
4. $\quad(\Phi_{\mathbf{i}}^n)^{k+1} = (\Phi_{\mathbf{i}}^n)^k + h(-\Delta_{t,G})^{-1}\big(\partial_t\big(2(\boldsymbol{\rho}_{\mathbf{i}}^n)^{k+1} - (\boldsymbol{\rho}_i^n)^k\big) + \mathrm{div}_G\big(2(\mathbf{m}_{\mathbf{i}}^n)^{k+1} - (\mathbf{m}_{\mathbf{i}}^n)^k\big)\big)$;
5. **end**

4.3 Convergence

Our proposed primal-dual algorithms for the tropical Wasserstein-1 and tropical Wasserstein-2 distances converge to their respective minimizers as given by Propositions 17 and 20.

Theorem 25. *(i) Consider the G-Prox PDHG algorithm for computing the tropical Wasserstein-1 distance. Let*

$$\sqrt{\tau\mu}\|(-\Delta_G)^{-\frac{1}{2}}\mathrm{div}_G\|_2 < 1.$$

Then $(\mathbf{m}^k, \Phi^k)$ defined by (19) converges weakly to $(\mathbf{m}^*, \Phi^*)$.

(ii) Consider the G-Prox PDHG algorithm for computing the tropical Wasserstein-2 distance. Let

$$\sqrt{\tau\mu}\|(-\Delta_{t,G})^{-\frac{1}{2}}\mathrm{div}_{t,G}\|_2 < 1.$$

Then $(\mathbf{m}^k, \boldsymbol{\rho}^k, \Phi^k)$ *defined by* (26) *converges weakly to* $(\mathbf{m}^*, \boldsymbol{\rho}^*, \Phi^*)$.

Proof. The proof follows that of Theorem 1 in Pock and Chambolle (2011). We justify the conditions in Pock and Chambolle (2011). In the case of (i), we write the Lagrangian L as

$$L(\mathbf{m}, \Phi) = g(\mathbf{m}) + \Phi^\intercal K\mathbf{m} - f^*(\Phi),$$

where $g(\mathbf{m}) = \|\mathbf{m}\|_{\mathrm{tr}}$, $K = \mathrm{div}_G$, and $f^*(\Phi) = \sum_{\mathbf{i}} \Phi_{\mathbf{i}}(\pi^0_{\mathbf{i}} - \pi^1_{\mathbf{j}})$. Observe that g, f^* are convex functions and K is a linear operator. Then there exists a saddle point $(\mathbf{m}^*, \Phi^*)$. Notice that the preconditioning norm for Φ is $\Sigma := \sigma(-\Delta_G)^{-1}$ and the preconditioning norm for $\mathbf{m}$ is $T := \tau\mathbb{I}$. Thus, the algorithm converges when $\|\Sigma^{\frac{1}{2}} K T^{\frac{1}{2}}\|_2^2 < 1$. This is our condition $\sqrt{\tau\mu}\|(-\Delta_G)^{-\frac{1}{2}}\mathrm{div}_G\|_2 < 1$, which finishes the proof. A similar argument holds for (ii). □

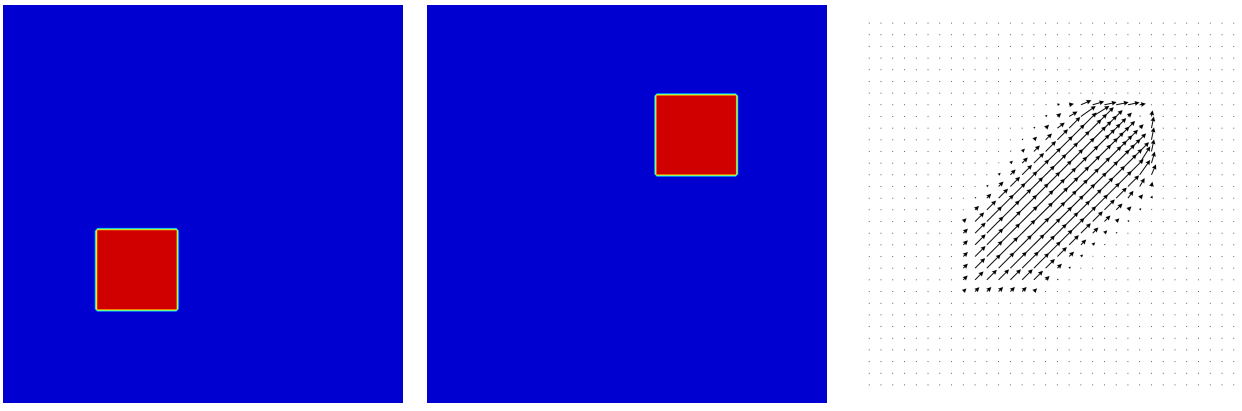
5 Numerical Experiments

In this section, we present the results of numerical experiments solving the tropical optimal transport problem for three different sets of initial densities using our proposed G-Prox primal-dual methods for L^1 and L^2. In particular, we give the minimizers of L^1 and L^2 tropical optimal transport problems from each experiment.

Experiment 1 We consider a two-dimensional problem on $\Omega = [0,1] \times [0,1]$. The initial densities ρ_0 and ρ_1 are same sizes of squares centered at $(\frac{1}{3}, \frac{1}{3})$ and $(\frac{2}{3}, \frac{2}{3})$, respectively. In this experiment, the parameters are

$$N_x = 128 \times 128,\ N_t = 15$$

Figure 6 shows the minimizer $m(x)$ of the tropical Wasserstein-1 distance and Figure 7 shows the minimizer $\rho(t,x)$ of the tropical Wasserstein-2 distance.



(a) ρ_0 (b) ρ_1 (c) $\mathbf{m}$

Figure 6: Experiment 1: L^1 tropical optimal transport. (a) and (b) show the initial densities ρ_0 and ρ_1, while (c) shows the geodesics of the L^1 tropical optimal transport between ρ_0 and ρ_1.

Experiment 2 Similar to Experiment 1, we consider a two dimensional problem on $\Omega = [0,1] \times [0,1]$. The initial densities ρ_0 and ρ_1 are same sizes of squares centered at $(\frac{1}{3}, \frac{2}{3})$ and $(\frac{2}{3}, \frac{1}{3})$ respectively. The same parameters are set as in Experiment 1. Together with Experiment 1, Experiment 2 shows that the minimizers of tropical optimal transport show different geodesics depending on the positions of initial densities. See Figure 8 for L^1 result and Figure 9 for L^2 result.

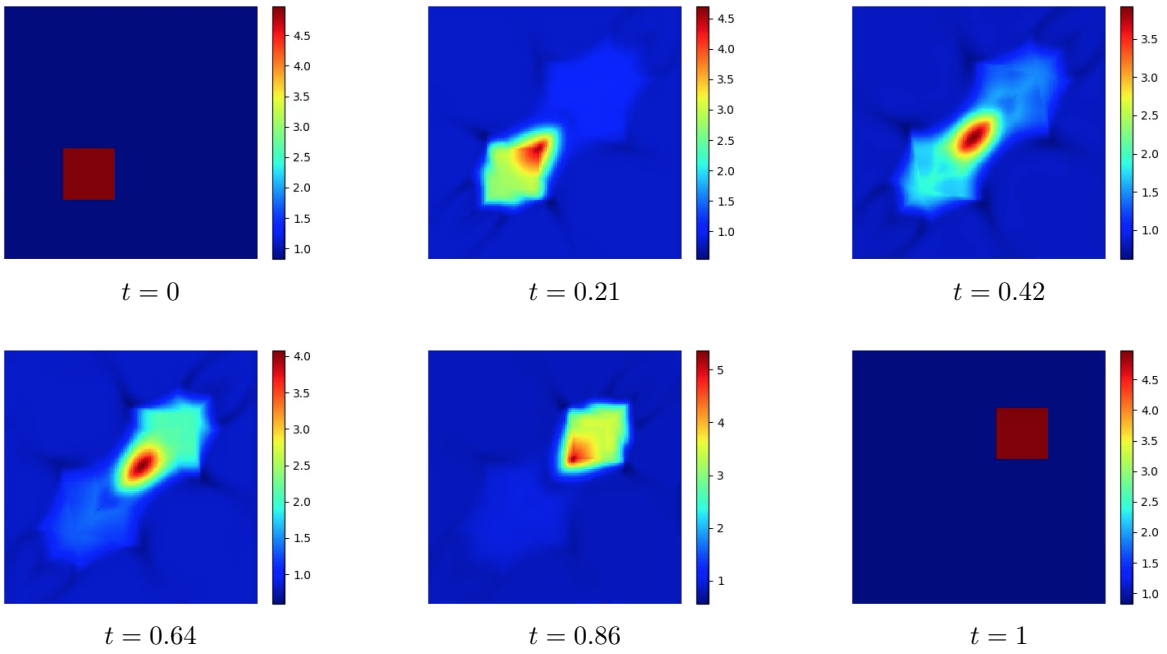


Figure 7: Experiment 1: L^2 tropical optimal transport. The six figures show the geodesics of L^2 tropical optimal transport from $t = 0$ to $t = 1$. The initial densities are same as in Figure 6.

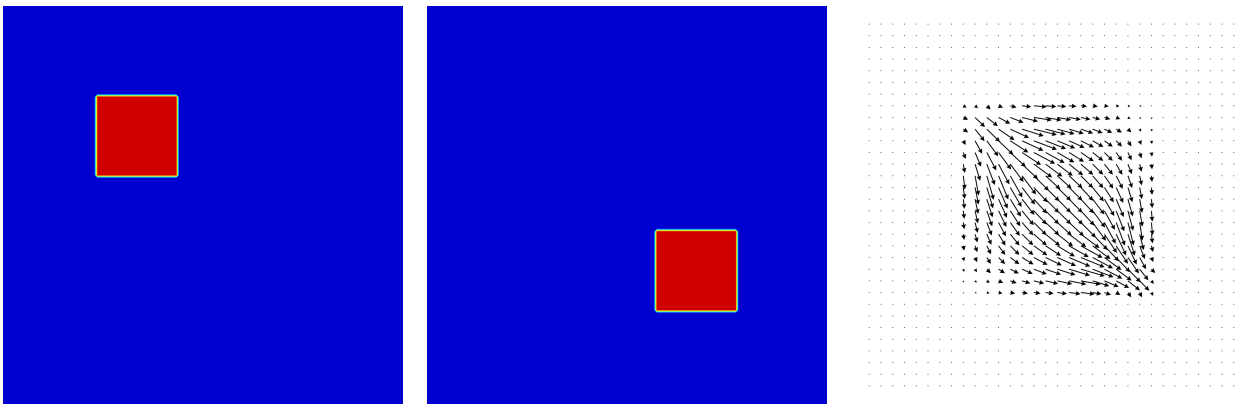
Experiment 3 We again consider a two dimensional problem on $\Omega = [0, 1] \times [0, 1]$. The initial density ρ_0 at time 0 is a square centered at $(0.5, 0.5)$ with width 0.2. The initial density ρ_1 at time 1 is four squares of the same size centered at $(0.2, 0.2)$, $(0.2, 0.8)$, $(0.8, 0.2)$ and $(0.8, 0.8)$ with width 0.1. The same parameters are set as in Experiment 1. See Figure 10 for the L^1 result and Figure 11 for the L^2 result; notice that the geodesics of minimizers from both results depend on the direction that the densities travel. We see that Experiment 3 coincides with Experiments 1 and 2.

Software Software to implement the numerical experiments presented in this paper is publicly available and located on the TropicalOT GitHub repository at `https://github.com/antheamonod/TropicalOT`.

6 Discussion

In this paper, we connect optimal transport theory—specifically, dynamic optimal transport—with tropical geometry. In particular, we explicitly formulate geodesics for the tropical Wasserstein-p distances over the tropical ambient space of phylogenetic trees. We also construct and implement primal-dual algorithms to compute tropical Wasserstein-1 and 2 geodesics on this space. These results provide a framework to identifying all infinitely-many geodesic paths between points in the tropical projective torus, including between phylogenetic trees, which leads to a better understanding of paths within phylogenetic tree space, following the work of Monod et al. (2018). In addition, the Wasserstein-2 distance possesses an important structure for statistical inference, since it provides the form for Fréchet means on the tropical projective torus, as well as a general inner product structure.

Our research lays the foundation for further connections between optimal transport and the tropical geometry of phylogenetic tree space, and provides powerful tools to study important aspects such as geometry and statistics on the tropical projective torus. From the perspective of optimal transport, we observe that the combinatorial structure of the tropical metric poses several interesting challenges in optimal transport. For example, the partial differential equations derived in Section 3 are defined in a piecewise manner: in two-dimensional sample space, there are six corresponding equations for geodesics in optimal transport. In



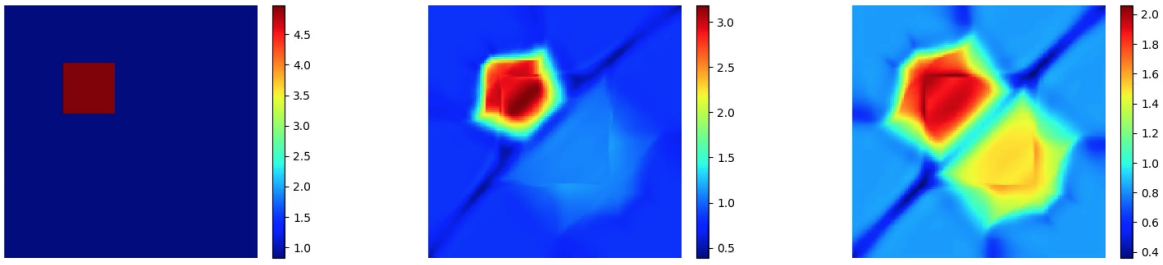
(a) ρ_0 (b) ρ_1 (c) $\mathbf{m}$

Figure 8: Experiment 2: L^1 tropical optimal transportation. (a) and (b) show the initial densities ρ_0 and ρ_1. (c) shows the geodesics of the L^1 tropical optimal transportation between ρ_0 and ρ_1.

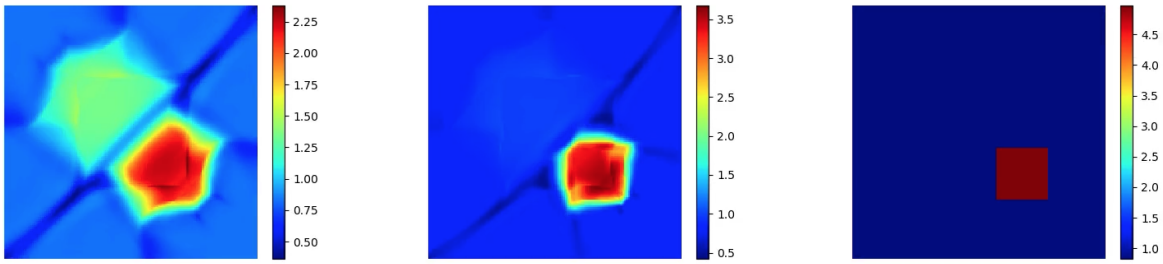
the general case, there are interesting regularity issues to be studied further. From the perspective of the tropical geometry of phylogenetic tree space, the theory of optimal transport and studying associated density manifolds provide a natural base to constructing heat equations with respect to the tropical metric. This provides an important potential to defining non-uniform probability distributions on the ambient space of the tropical geometric phylogenetic tree space: classically, the solution to the heat equation gives rise to the Gaussian distribution, thus, a solution to the tropical heat equation is a candidate for a tropical Gaussian distribution on the tropical projective torus (Tran, 2018; Maazouz and Tran, 2019). The dynamic setting of optimal transport with the tropical ground metric introduced in this paper also provides a foundation to studying the displacement convexity and Ricci curvature tensor on the tropical projective torus. In forthcoming work, we further study such questions by applying the relevant work of Li (2018, 2019), which also studies geometric and probabilistic questions in the context of optimal transport theory.

Acknowledgments

The authors wish to thank Marzieh Eidi, Théo Lacombe, Victor Panaretos, Ronen Talmon, and Yoav Zemel for helpful discussions. W.L. and W.L. are supported by grant AFOSR MURI FA9550-18-1-0502. A.M. wishes to acknowledge the Max Planck Institute for Mathematics in the Sciences for hosting her visit in Leipzig in July 2018, which inspired this work.

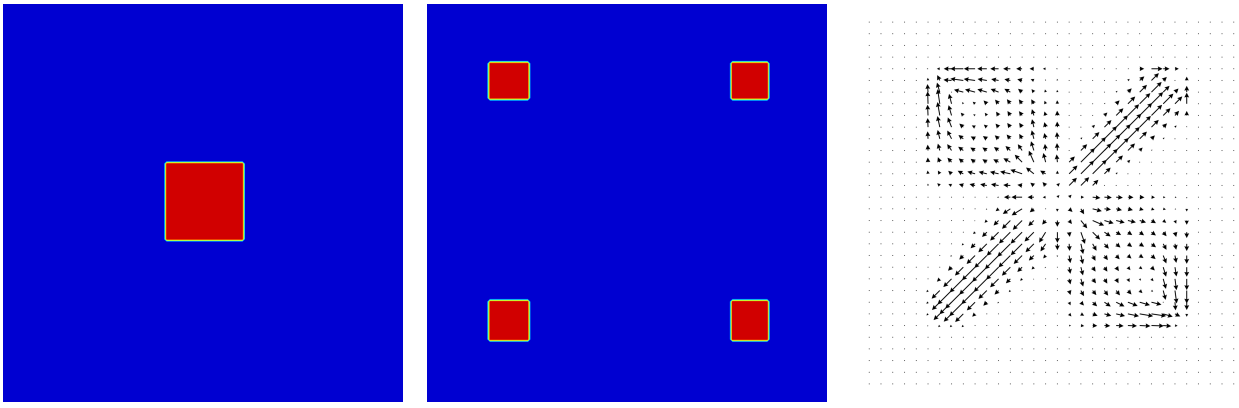


$t = 0$ $t = 0.21$ $t = 0.42$



$t = 0.64$ $t = 0.86$ $t = 1$

Figure 9: Experiment 2: L^2 tropical optimal transport. The figures show the geodesics of L^2 tropical optimal transport between two initial densities from $t = 0$ to $t = 1$. The initial densities are same as in Figure 8.



(a) ρ_0 (b) ρ_1 (c) $\mathbf{m}$

Figure 10: Experiment 3: L^1 tropical optimal transport. (a) and (b) show the initial densities ρ_0 and ρ_1, while (c) shows the geodesics of the L^1 tropical optimal transport between ρ_0 and ρ_1. This experiment shows similar patterns of geodesics from Experiment 1 and Experiment 2.

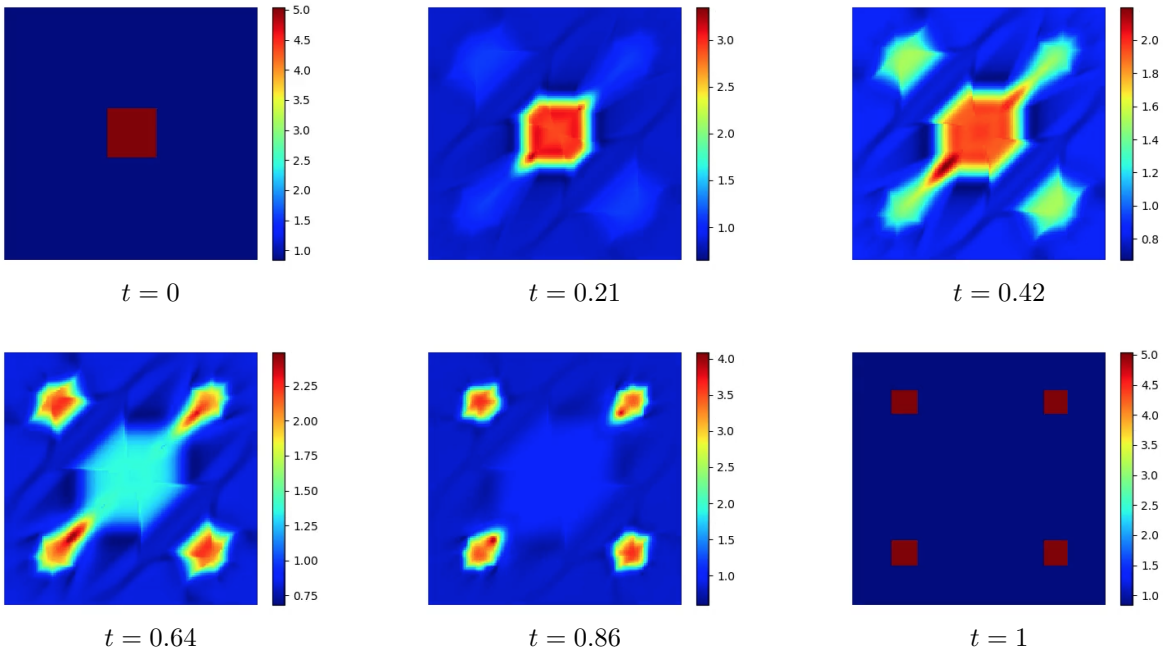


Figure 11: Experiment 3: L^2 tropical optimal transport. The six figures show the geodesics of L^2 tropical optimal transportation from $t = 0$ to $t = 1$. The initial densities are same as in Figure 10.

References

Akian, M., S. Gaubert, V. Niţică, and I. Singer (2011). Best Approximation in Max-plus Semimodules. *Linear Algebra and its Applications 435*(12), 3261–3296.

Ambrosio, L. and N. Gigli (2013). A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pp. 1–155. Springer.

Ambrosio, L., N. Gigli, and G. Savaré (2008). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

Benamou, J.-D. and Y. Brenier (2000). A Computational Fluid Mechanics Solution to the Monge-Kantorovich Mass Transfer Problem. *Numerische Mathematik 84*(3), 375–393.

Benamou, J.-D., G. Carlier, and R. Hatchi (2016). A numerical solution to Monge's problem with a Finsler distance as cost. *<hal-01261094>*.

Buneman, P. (1974). A Note on the Metric Properties of Trees. *Journal of Combinatorial Theory, Series B 17*(1), 48–50.

Çelik, T. Ö., A. Jamneshan, Montúfar, B. Sturmfels, and L. Venturello (2019). Optimal Transport to a Variety. *arXiv:1909.11716*.

Chambolle, A. and T. Pock (2011). A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision 40*(1), 120–145.

Cohen, G., S. Gaubert, and J.-P. Quadrat (2004). Duality and Separation Theorems in Idempotent Semimodules. *Linear Algebra and its Applications 379*, 395–422. Special Issue on the Tenth ILAS Conference (Auburn, 2002).

Evans, S. N. and F. A. Matsen (2012). The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(3), 569–592.

Jacobs, M., F. Léger, W. Li, and S. Osher (2018). Solving Large-Scale Optimization Problems with a Convergence Rate Independent of Grid Size. *arXiv:1805.09453 [math]*.

Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, Volume 37, pp. 199–201.

Kloeckner, B. R. (2015). A geometric study of Wasserstein spaces: Ultrametrics. *Mathematika 61*(1), 162–178.

Lafferty, J. D. (1988). The Density Manifold and Configuration Space Quantization. *Transactions of the American Mathematical Society 305*(2), 699–741.

Li, W. (2018). Geometry of Probability Simplex via Optimal Transport. *arXiv:1803.06360 [math]*.

Li, W. (2019). Diffusion Hypercontractivity via Generalized Density Manifold. *arXiv:1907.12546 [cs, math]*.

Li, W., E. K. Ryu, S. Osher, W. Yin, and W. Gangbo (2018). A Parallel Method for Earth Mover's Distance. *Journal of Scientific Computing 75*(1), 182–197.

Lin, B., B. Sturmfels, X. Tang, and R. Yoshida (2017). Convexity in Tree Spaces. *SIAM Journal on Discrete Mathematics 31*(3), 2015–2038.

Lin, B. and R. Yoshida (2018). Tropical Fermat–Weber Points. *SIAM Journal on Discrete Mathematics 32*(2), 1229–1245.

Maazouz, Y. E. and N. M. Tran (2019). Statistics of Gaussians on local fields and their tropicalizations. *arXiv preprint arXiv:1909.00559*.

Maclagan, D. and B. Sturmfels (2015). *Introduction to Tropical Geometry (Graduate Studies in Mathematics)*. American Mathematical Society.

Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie royale des sciences de Paris*.

Monod, A., B. Lin, Q. Kang, and R. Yoshida (2018). Tropical Geometry of Phylogenetic Tree Space: A Statistical Perspective. *arXiv:1805.12400*.

Otto, F. (2001). The Geometry of Dissipative Evolution Equations: The Porous Medium Equation. *Communications in Partial Differential Equations 26*(1-2), 101–174.

Otto, F. and C. Villani (2000). Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *Journal of Functional Analysis 173*(2), 361 – 400.

Panaretos, V. M. and Y. Zemel (2019). Statistical Aspects of Wasserstein Distances. *Annual Review of Statistics and Its Application 6*(1), 405–431.

Pock, T. and A. Chambolle (2011). Diagonal Preconditioning for First Order Primal-Dual Algorithms in Convex Optimization. In *2011 International Conference on Computer Vision*, pp. 1762–1769.

Sommerfeld, M. and A. Munk (2018). Inference for empirical Wasserstein distances on finite spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 219–238.

Speyer, D. and B. Sturmfels (2004). The Tropical Grassmannian. *Advances in Geometry 4*(3).

Tran, N. M. (2018). Tropical Gaussians: A Brief Survey. *arXiv:1808.10843*.

Villani, C. (2003). *Topics in Optimal Transportation*. Number 58. American Mathematical Soc.

Villani, C. (2008). *Optimal Transport: Old and New*, Volume 338. Springer Science & Business Media.

Wasserman, L. (2019, April). Lecture notes on Statistical Methods for Machine Learning.