

# Wasserstein Barycenters can be Computed in Polynomial Time in Fixed Dimension

**Jason M. Altschuler**

JASONALT@MIT.EDU

**Enric Boix-Adserà**

EBOIX@MIT.EDU

*Laboratory for Information and Decision Systems*

*Massachusetts Institute of Technology, Cambridge MA 02139*

**Editor:** Prateek Jain

## Abstract

Computing Wasserstein barycenters is a fundamental geometric problem with widespread applications in machine learning, statistics, and computer graphics. However, it is unknown whether Wasserstein barycenters can be computed in polynomial time, either exactly or to high precision (i.e., with  $\text{polylog}(1/\varepsilon)$  runtime dependence). This paper answers these questions in the affirmative for any fixed dimension. Our approach is to solve an exponential-size linear programming formulation by efficiently implementing the corresponding separation oracle using techniques from computational geometry.

**Keywords:** Wasserstein barycenters, free support, polynomial-time algorithm, exponential-size linear program, power diagrams

## 1. Introduction

Given discrete probability distributions  $\mu_1, \dots, \mu_k$  supported on  $\mathbb{R}^d$  and a vector  $\lambda \in \mathbb{R}^k$  of non-negative weights summing to 1, the corresponding *Wasserstein barycenters* are the probability distributions  $\nu$  minimizing

$$\operatorname{argmin}_{\nu} \sum_{i=1}^k \lambda_i \mathcal{W}(\mu_i, \nu), \quad (1)$$

where above  $\mathcal{W}(\cdot, \cdot)$  denotes the squared 2-Wasserstein distance (Agueh and Carlier, 2011). Wasserstein barycenters provide a natural extension of the notion of averaging points to the notion of averaging point clouds. Importantly, they naturally inherit the ability of optimal transportation to capture geometric properties of the data.

This desirable property has led to the widespread use of Wasserstein barycenters in many applications. Applications in statistics and machine learning include for instance the  $n$ -coupling problem (Rüschendorf and Uckelmann, 2002), constrained clustering (Cuturi and Doucet, 2014; Ho et al., 2017), fusing measurements from partial sensors (Elvander et al., 2020), and fusing measurements for scalable Bayesian learning (Srivastava et al., 2018). Applications in image processing and computer graphics include for instance texture mixing (Rabin et al., 2011) and shape interpolation (Solomon et al., 2015). For further applications, see the surveys (Peyré and Cuturi, 2017; Panaretos and Zemel, 2019).

*Open problem: computing barycenters in polynomial time.* Despite considerable algorithmic

work, it is an open problem (e.g., (Borgwardt, 2017)) whether Wasserstein barycenters between discrete distributions can be exactly computed in polynomial time in the input size. A highly related open problem is whether Wasserstein barycenters can be computed to high accuracy, i.e., whether an  $\varepsilon$ -additively approximate solution for (1) can be computed in time that is polynomial in the input size and  $\log(1/\varepsilon)$ . This paper answers these questions in the affirmative for any fixed dimension  $d$ .

Previous methods require time that depends polynomially on  $1/\varepsilon$  in order to compute  $\varepsilon$ -approximate barycenters. This means that in practice they can only solve to a few digits of precision. See the prior work section for details. In many applications, including nearly all of those mentioned above, Wasserstein barycenters are used as a subroutine in a larger pipeline to solve downstream data science tasks. Thus, high-precision algorithms are important for downstream performance and to avoid error propagation, especially in applications which require multiple barycenter computations.

The well-documented key obstacle is that *a priori*, there are  $n^k$  candidate atoms for the barycenter’s support, where  $n$  is (an upper bound on) the number of atoms in each  $\mu_i$ . While there always exists a barycenter with  $\text{poly}(n, k)$  atoms, finding which atoms those are requires pruning the  $n^k$  exponentially many candidates.

## 1.1 Prior Work

The literature on computing Wasserstein barycenters is extensive and rapidly growing.<sup>1</sup> Existing algorithms can be partitioned into two categories, depending on how they handle optimizing over the  $n^k$  exponentially many candidate atoms for the barycenter support.

*Fixed-support.* Most existing algorithms work around this exponential complexity by making a “fixed-support approximation”: they assume that the barycenter is supported on a small guessed set of points, and then only optimize over the corresponding masses. This reduces the barycenter problem to a polynomial-size LP, which can then be solved efficiently using out-of-the-box LP solvers, or alternatively using specialized methods such as entropic regularization; see, e.g., (Cuturi and Doucet, 2014; Benamou et al., 2015; Solomon et al., 2015; Carlier et al., 2015; Staib et al., 2017; Kroshnin et al., 2019; Janati et al., 2020; Lin et al., 2020) among many others. However, the key issue with fixed-support algorithms is that guessing a reasonable support set for the barycenter requires  $\varepsilon$ -covering the space. Specifically, these algorithms require fixing the support to roughly  $(R/\varepsilon)^d$  points in order to get an  $\varepsilon$ -additive approximation to the barycenter<sup>2</sup>, where  $R$  is a bound on the squared diameter of the supports of the input distributions. This results in  $\text{poly}(n, k, R/\varepsilon)$  final runtimes in constant dimension  $d$ .

---

1. We mention in passing that an orthogonal line of work aims to compute barycenters of *continuous* distributions (typically restricted to Gaussian distributions so that both  $\mu_i$  and  $\nu$  have compact representations for computational purposes). See, e.g., (Chewi et al., 2020; Álvarez-Esteban et al., 2016).

2. An alternative is to restrict to the union of the supports of  $\mu_1, \dots, \mu_k$ , which has only  $nk$  points. However, this cannot get arbitrarily close approximations (Borgwardt, 2017).

In contrast, our proposed algorithm has  $\text{poly}(n, k, \log(R/\varepsilon))$  runtime which, critically, has polylogarithmic dependence on  $R/\varepsilon$ .<sup>3</sup> In practice, this means that our algorithm can often solve up to machine precision, whereas fixed-support algorithms can only solve up to a few digits of precision—see §4 for experiments.

*Free-support.* Achieving our  $\text{poly}(n, k, \log(R/\varepsilon))$  runtime precludes making a fixed-support approximation. Such algorithms are called “free-support algorithms”. The key obstacle is that, as mentioned before, this requires optimizing over the support set of the barycenters, which *a priori* can only be restricted to  $n^k$  candidate points. All previous free-support algorithms either run in exponential time, or are heuristics without provable guarantees; see, e.g., (Cuturi and Doucet, 2014; Luise et al., 2019).

In contrast, we show how to optimize over these  $n^k$  candidate points in polynomial time by exploiting the geometric structure of their configuration.

## 1.2 Contribution

We give the first algorithm that, in any fixed dimension  $d$ , solves the Wasserstein barycenter problem exactly or to high precision in polynomial time. For simplicity of notation, throughout  $d$  is a constant; the running time for fixed  $d$  is  $(nk)^d$  times a polynomial in the input size.

**Theorem 1 (Computing high-precision barycenters)** *There is an algorithm that, given  $k$  distributions each supported on  $n$  atoms in the ball of squared radius  $R$  in  $\mathbb{R}^d$ , a weight vector  $\lambda$ , and an accuracy  $\varepsilon > 0$ , computes an  $\varepsilon$ -additively approximate Wasserstein barycenter in  $\text{poly}(n, k, \log(R/\varepsilon))$  time. Moreover, this barycenter has support size at most  $nk - k + 1$ .*

**Theorem 2 (Computing exact barycenters)** *If the weight vector and distributions are represented with  $\log U$  bits of precision, then an exact barycenter can be found in  $\text{poly}(n, k, \log U)$  time. Moreover, this barycenter has support size at most  $nk - k + 1$ .*

Our algorithm is described in §3. Briefly, the high-level idea is as follows. Our starting point is a well-known LP reformulation of the Wasserstein barycenter problem as a Multi-marginal Optimal Transport (MOT) problem, recalled in the preliminaries section. This is an exponential-size LP with  $n^k$  variables, one for each candidate atom. Nevertheless, we show that a sparse solution can be computed in polynomial time. Our approach consists of two steps. First, by leveraging tools from combinatorial optimization and exploiting the special structure of the MOT LP, we show that this problem can be solved efficiently if one can efficiently implement the separation oracle for the LP dual of MOT. Second, by leveraging tools from computational geometry such as power diagrams and the complexity of hyperplane arrangements, we show how to efficiently implement this separation oracle.

In addition to its polynomial runtime, our algorithm has two additional properties that may be useful in downstream applications. First, the outputted barycenter  $\nu$  has small support of  $O(nk)$  size, which is much smaller than the *a priori*  $n^k$  bound on the support

---

3. This means that our algorithm solves the barycenter problem exactly in polynomial time, whereas previous algorithms require *pseudo-polynomial* time. This is because solving the barycenter problem exactly requires  $\varepsilon$  to be exponentially small in the bit-complexity of the input.

size. In particular, the support size of  $\nu$  is at most the maximal sparsity of any vertex of the transportation polytope between  $\mu_1, \dots, \mu_k$ —which is at most  $nk - k + 1$ . Note that Theorem 2 is not at odds with the NP-hardness of finding the sparsest barycenter (Borgwardt and Patterson, 2019): indeed, our algorithm outputs a solution that albeit sparse is not necessarily the sparsest. Second, as a by-product, our algorithm also produces sparse solutions to the optimal transport problems  $\mathcal{W}(\mu_i, \nu)$  that are non-mass-splitting maps from  $\nu$  to  $\mu_i$ . Among other benefits, this enables easy visualization and interpretability of the results—in comparison to entropic-regularization based approaches which produce “blurry” dense maps.

Although the focus of this work is theoretical, we also provide preliminary numerical experiments in §4 demonstrating that a slight variant of our algorithm can provide high-precision solutions at previously intractable problem sizes.

Finally, in §5, we briefly mention that the techniques we develop in this paper extend to solving several related problems. In particular, this gives the first polynomial-time algorithms for computing geometric medians with respect to the 1-Wasserstein distance (a.k.a. Earth Mover’s distance) over any of the popular ground metrics  $\ell_1$ ,  $\ell_2$ , or  $\ell_\infty$ .

## 2. Preliminaries

This section is organized as follows. First, in §2.1, we establish our notation, which is mostly standard. Then in §2.2, §2.3, §2.4, respectively, we recall relevant background from the machine learning, combinatorial optimization, and computational geometry literatures—namely, background about LP formulations of Wasserstein barycenters, algorithms for solving exponential-size LP, and algorithms for manipulating power diagrams.

### 2.1 Notation

The set  $\{1, \dots, n\}$  is denoted by  $[n]$ . The  $k$ -fold tensor product space  $\mathbb{R}^n \otimes \dots \otimes \mathbb{R}^n$  is denoted by  $(\mathbb{R}^n)^{\otimes k}$ , and similarly for  $(\mathbb{R}_{\geq 0}^n)^{\otimes k}$ . For shorthand, we often denote a tuple  $(j_1, \dots, j_k) \in [n]^k$  by  $\vec{j}$ . The  $i$ -th marginal,  $i \in [k]$ , of a tensor  $P \in (\mathbb{R}^n)^{\otimes k}$  is denoted by the vector  $m_i(P) \in \mathbb{R}^n$ , and has entries  $[m_i(P)]_\ell := \sum_{\vec{j} \in [n]^k: j_i = \ell} P_{\vec{j}}$ . The transportation polytope between  $\mu_1, \dots, \mu_k$  is the set of joint distributions with one-dimensional marginal distributions  $\mu_1, \dots, \mu_k$ , and is identified with the set  $\mathcal{M}(\mu_1, \dots, \mu_k) := \{P \in (\mathbb{R}_{\geq 0}^n)^{\otimes k} : m_i(P) = \mu_i, \forall i \in [k]\}$ , where we abuse notation slightly by identifying  $\mu_i$  with its vector of probabilities (in any order). The closure of a set  $E \subset \mathbb{R}^d$  (with respect to the standard topology) is denoted by  $\bar{E}$ . Throughout, we assume without loss of generality that each  $\lambda_i$  is strictly positive, since otherwise  $\mu_i$  does not affect the barycenter (see equation (1)).

### 2.2 LP Formulation of Wasserstein Barycenters

Our starting point is the known fact (see, e.g., (Agueh and Carlier, 2011; Benamou et al., 2015; Anderes et al., 2016)) that a barycenter  $\nu$  can be found by solving the Multimarginal Optimal Transport problem

$$\min_{P \in \mathcal{M}(\mu_1, \dots, \mu_k)} \langle P, C \rangle, \tag{MOT}$$

for the cost tensor  $C \in (\mathbb{R}^n)^{\otimes k}$  with entries

$$C_{\vec{j}} = \min_{y \in \mathbb{R}^d} \sum_{i=1}^k \lambda_i \|x_{i,j_i} - y\|^2, \quad (2)$$

or equivalently,  $C_{\vec{j}} = \sum_{i=1}^k \lambda_i \|x_{i,j_i} - \sum_{\ell=1}^k \lambda_\ell x_{\ell,j_\ell}\|^2$  by optimality of  $y = \sum_{\ell=1}^k \lambda_\ell x_{\ell,j_\ell}$ . Specifically, the reduction from the Wasserstein barycenter problem to the LP (MOT) is as follows.

**Lemma 3** *If  $P \in \mathcal{M}(\mu_1, \dots, \mu_k)$  is an optimal solution to (MOT), then the pushforward of  $P$  under the map  $(X_1, \dots, X_k) \mapsto \sum_{i=1}^k \lambda_i X_i$  is an optimal barycenter  $\nu$ . Furthermore, the support size of  $\nu$  is at most the support size of  $P$ , and also the coupling  $(\sum_{i=1}^k \lambda_i X_i, X_j)$  is a non-mass-splitting map that solves the Optimal Transport problem from  $\nu$  to  $\mu_j$ .*

Notice that applying this pushforward map  $(X_1, \dots, X_k) \mapsto \sum_{i=1}^k \lambda_i X_i$  in order to compute  $\nu$  from  $P$  requires only  $O(skd)$  arithmetic operations, where  $s$  denotes the support size of  $P$ . In particular, this takes polynomial time if  $s$  is of polynomial size. Therefore it suffices to compute a sparse solution  $P$  of the LP (MOT).

Note that the solution  $P$  is guaranteed to be sparse if it is a vertex solution. Indeed, since (MOT) is a standard-form LP whose constraints have rank at most  $nk - k + 1$ , each vertex solution has at most  $nk - k + 1$  non-zero entries.

**Lemma 4** *If  $P$  is a vertex of the transportation polytope  $\mathcal{M}(\mu_1, \dots, \mu_k)$ , then  $P$  has at most  $nk - k + 1$  non-zero entries.*

An obvious obstacle for computing any solution—let alone a sparse solution—of the LP formulation (MOT) is that it has  $n^k$  exponentially many variables. An LP that will be useful to us in the sequel is its dual

$$\max_{p_1, \dots, p_k \in \mathbb{R}^n} \sum_{i=1}^k \langle p_i, \mu_i \rangle \quad \text{subject to} \quad C_{\vec{j}} - \sum_{i=1}^k [p_i]_{j_i} \geq 0, \quad \forall \vec{j} \in [n]^k. \quad (\text{MOT-D})$$

An attractive property of (MOT-D) is that it has only  $nk$  polynomially many variables. However, of course this alone is not enough to make solving (MOT-D) tractable because it has  $n^k$  exponentially many constraints. That is, dualizing has transferred the exponential complexity in the number of variables in (MOT) to the number of constraints in (MOT-D).

### 2.3 Algorithms for Exponential-Size LP

It is a classical fact (see, e.g., (Khachiyan, 1980; Bertsimas and Tsitsiklis, 1997; Grötschel et al., 2012)) that regardless of the number of constraints, an LP with polynomially many variables can be solved in polynomial time so long as there is a polynomial-time implementation of the separation oracle for its feasibility set. Here we recall the technical details of this fact.

First, we recall the definition of a separation oracle. This definition is simply an algorithmic reformulation of the Separating Hyperplane Theorem, which states that for any convex set  $\mathcal{K} \subseteq \mathbb{R}^N$  and any point  $p \in \mathbb{R}^N$ , exactly one of two alternatives must hold: either  $p \in \mathcal{K}$ , or there exists a hyperplane that separates  $p$  from  $\mathcal{K}$  (i.e., there exists a vector  $h \in \mathbb{R}^N$  and a scalar  $g \in \mathbb{R}$  such that  $\langle h, p \rangle \geq g$  and  $\langle h, x \rangle < g$  for all  $x \in \mathcal{K}$ ).

**Definition 5** A separation oracle for a convex set  $\mathcal{K}$  is an algorithm that given a point  $p$ , either outputs that  $p \in \mathcal{K}$  or outputs a hyperplane that separates  $p$  from  $\mathcal{K}$ .

Given a polynomial-time implementation of a separation oracle for a polytope, the Ellipsoid algorithm can solve an LP over that polytope in polynomial time. This result can be found in (Grötschel et al., 1981).

**Theorem 6** Let  $\log U$  be an upper bound on the number of bits needed to represent any entry in  $A \in \mathbb{R}^{M \times N}$ ,  $b \in \mathbb{R}^M$ , or  $c \in \mathbb{R}^N$ . Then the Ellipsoid algorithm finds a vertex solution to  $\operatorname{argmin}\{c^T x : x \in \mathcal{P}\}$  in  $\operatorname{poly}(N, \log U)$  time and  $\operatorname{poly}(N, \log U)$  calls to a separation oracle for the polytope  $\mathcal{P} = \{x \in \mathbb{R}^N : Ax \leq b\}$ .

## 2.4 Computational Geometry Algorithms

A key ingredient in our barycenter algorithm is power diagrams. Here we introduce these objects and some basic facts about their complexity. Although  $d$  is a fixed constant in our final results, we state the explicit dependence on the dimension  $d$  in these power diagram complexity bounds to highlight how and where our algorithm incurs exponential runtime dependence in  $d$ .

**Definition 7** The power diagram on the spheres  $S(z_1, r_1), \dots, S(z_n, r_n)$  with centers  $z_j \in \mathbb{R}^d$  and radii  $r_j \geq 0$  is the cell complex whose cells  $E_1, \dots, E_n$  are given by

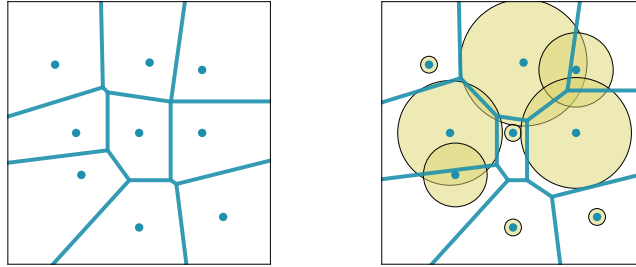
$$E_j = \{y \in \mathbb{R}^d : \|z_j - y\|^2 - r_j^2 < \|z_{j'} - y\|^2 - r_{j'}^2, \forall j' \neq j\}.$$

A power diagram “essentially” partitions  $\mathbb{R}^d$  in the sense that its cells are disjoint and their closures cover  $\mathbb{R}^d$ . See Figure 1 for an illustration. Following are two relevant classical facts. The first essentially shows that a power diagram is defined by a small hyperplane arrangement which can moreover be computed efficiently.

**Lemma 8 (Theorems 1 and 7 of (Aurenhammer, 1987), using convex hull algorithm of (Chazelle, 1993))** A power diagram on  $n$  spheres in  $\mathbb{R}^d$  has  $O(n)$  affine facets of dimension  $d-1$ . Moreover these facets can be computed in  $O((n \log n + n^{\lceil d/2 \rceil}) \cdot \operatorname{polylog} U)$  time, where  $\log U$  is the number of bits of precision.

The second is about hyperplane arrangements. In the sequel this lets us bound the complexity of the “intersection” of multiple power diagrams (defined in §3.2).

**Lemma 9 (Theorem 3.3 of (Edelsbrunner et al., 1986))** The cell complex formed by an arrangement of  $N$  hyperplanes in  $\mathbb{R}^d$ , represented up to  $\log U$  bits of precision, can be computed in  $N^d \cdot \operatorname{polylog}(N, U)$  time.



**Figure 1:** Two power diagrams on the same  $n = 9$  points with varying weights  $w$ . Left: all weights are zero (so this is a Voronoi diagram). Right: the weight of a point is indicated by the size of the ball around it. Increasing the weight of a point increases the size of its cell.

### 3. Algorithm

In this section we describe our algorithm and prove our main results (Theorems 1 and 2). We begin by overviewing the high-level approach. Let us consider the exact solver in Theorem 2, as the approximate solver in Theorem 1 is implemented by exactly solving a rounded problem (see §3.3 for details).

Recall from §2.2 that it suffices to solve the LP formulation (MOT) of the Wasserstein barycenter problem. However, solving this LP presents a computational obstacle since it has  $n^k$  decision variables. Moreover, we desire a sparse solution  $P$ —rather than a generic solution which has exponentially many non-zero entries—since a polynomial sparsity for  $P$  ensures a polynomial support size for the final barycenter  $\nu$ .

The starting point of our approach is recalling the classical fact from §2.3 that regardless of the number of constraints, an LP with polynomially many variables can be solved in polynomial time so long as the corresponding separation oracle can be implemented in polynomial time. While (MOT) is not such an LP since it has exponentially many variables, this result applies to its dual (MOT-D). That is, we can efficiently solve (MOT-D) so long as we can efficiently implement the corresponding separation oracle.

However, two key obstacles remain. First, recovering a primal solution is non-trivial in that in general a dual solution does not necessarily “help” to find a primal solution (Bertsimas and Tsitsiklis, 1997, Exercise 4.17), let alone a sparse primal solution. Second, and most importantly, this approach requires an efficient implementation of the separation oracle for (MOT-D), which does not exist for general Multimarginal Optimal Transport problems (for concrete NP-hard examples see (Altschuler and Boix-Adserà, 2020a)).

We solve these issues in two steps:

1. *Reduction to separation oracle.* We reduce solving (MOT) in polynomial time to solving the separation oracle for the dual LP (MOT-D) in polynomial time. (Further, we show how to ensure the solution is polynomially sparse.)
2. *Efficient algorithm for separation oracle.* We use tools from computational geometry to solve the separation oracle for (MOT-D) in polynomial time.

At this point, it is worth remarking what special “structure” of the exponential-size LP (MOT) we exploit in order to solve it in polynomial time. Step 1 does *not* extend to

general LP, i.e., one cannot efficiently solve an LP given only an efficient separation oracle for its dual (Grötschel et al., 2012). Instead, step 1 crucially exploits the particular “structure” of the feasibility constraints defining (MOT), details in §3.1. This extends to arbitrary Multimarginal Optimal Transport problems (i.e., arbitrary costs  $C$ ), and therefore may be of independent interest. However, step 1 is of course useless unless one can efficiently solve the separation oracle in step 2. Indeed, as mentioned above, step 2 does *not* extend to general Multimarginal Optimal Transport problems, i.e., there does not exist an efficient implementation of the the separation oracle for (MOT-D) for arbitrary costs  $C \in (\mathbb{R}^n)^{\otimes k}$ . It is here—in step 2, not step 1—that we crucially exploit the remaining “structure” in the LP reformulation (MOT) of the barycenter problem, namely the cost  $C$  defined in (2). Intuitively, this “structure” of  $C$  is geometric: the  $n^k$  entries in  $C$  correspond to  $n^k$  candidate points for the barycenter’s support, and these points must lie in certain constrained geometric configurations, details in §3.2.

Let us now elaborate on steps 1 and 2. To do this, we first recast the separation oracle for (MOT-D) in a convenient way for our algorithmic development.

**Definition 10** *Given  $p = (p_1, \dots, p_k) \in \mathbb{R}^{n \times k}$ , the oracle  $SEP$  returns a tuple  $SEP(p) \in \operatorname{argmin}_{\vec{j} \in [n]^k} C_{\vec{j}} - \sum_{i=1}^k [p_i]_{j_i}$ .*

The intuition behind  $SEP$  is that it implements<sup>4</sup> the separation oracle for (MOT-D) because given the tuple  $\vec{j} = SEP(p)$ , exactly one of the following two alternatives must hold:

- $C_{\vec{j}} - \sum_{i=1}^k [p_i]_{j_i} \geq 0$ , in which case this certifies that  $p$  is feasible for (MOT-D).
- $C_{\vec{j}} - \sum_{i=1}^k [p_i]_{j_i} < 0$ , in which case this provides a hyperplane that separates  $p$  from the feasible set of (MOT-D).

Now, in terms of this oracle  $SEP$ , steps 1 and 2 are formally summarized as follows.

**Proposition 11 (Step 1)** *Let  $C \in (\mathbb{R}^n)^{\otimes k}$  be an arbitrary cost, and let  $\log U$  be the maximum number of bits of precision in an entry of  $C$ . A vertex solution  $P^*$  for (MOT) can be found in  $\operatorname{poly}(n, k, \log U)$  time and  $\operatorname{poly}(n, k, \log U)$  calls to  $SEP$ .*

**Proposition 12 (Step 2)** *If the cost  $C$  is given by (2), the oracle  $SEP(p)$  can be implemented in  $\operatorname{poly}(n, k, \log U)$  time, where  $\log U$  is the number of bits of precision needed to represent the points  $x_{i,j} \in \mathbb{R}^d$ , weights  $\lambda_i \in \mathbb{R}_{>0}^k$  and potentials  $p \in \mathbb{R}^{n \times k}$ .*

The remainder of the section is organized as follows. In §3.1 and §3.2, we detail the algorithms in steps 1 and 2, respectively, and prove Propositions 11 and 12. Combining Propositions 11 and 12 then proves Theorem 1 aside from checking bit-complexity details, which is done formally in §3.3.

---

4. In fact, it can be shown that  $SEP$  is polynomial-time *equivalent* to the separation oracle for (MOT-D), i.e., each oracle can be implemented using polynomial many calls to the other oracle and polynomial additional computation time. However, this is not needed in the sequel.



### 3.1 Step 1: Reduction to Separation Oracle

Here we prove Proposition 11 and describe the algorithm in it. This algorithm has two steps. First, it identifies a small set  $S \subseteq [n]^k$  on which an optimal MOT solution is supported. Second, it optimizes over distributions supported on  $S$ .

**Proof of Proposition 11** First, we construct a set  $S \subseteq [n]^k$  as follows. Since  $\text{SEP}(p)$  implements a separation oracle for (MOT-D), and since (MOT-D) has  $N = nk$  variables, Theorem 6 implies that an optimal solution for (MOT-D) can be found by the Ellipsoid algorithm in  $\text{poly}(N, \log U) = \text{poly}(n, k, \log U)$  time. Let  $L$  denote the number of  $\text{SEP}$  queries made by the Ellipsoid algorithm. For  $l \in [L]$ , let  $p^{(l)} \in \mathbb{R}^{n \times k}$  be the argument of the  $l$ -th query to  $\text{SEP}$ , and let  $\vec{j}^{(l)} \in [n]^k$  be the returned tuple. Let  $S := \{\vec{j}^{(l)}\}_{\ell=1}^L$  denote the set of all returned tuples.

Next, we show how to compute an optimal vertex solution for (MOT) using  $S$ . Let  $\mathcal{M}_S(\mu_1, \dots, \mu_k) = \{P \in \mathcal{M}(\mu_1, \dots, \mu_k) : P_{\vec{j}} = 0 \ \forall \vec{j} \notin S\}$  be the set of distributions in the transportation polytope supported on  $S$ , and let (MOT $_S$ ) be (MOT) with the decision set  $\mathcal{M}(\mu_1, \dots, \mu_k)$  replaced by  $\mathcal{M}_S(\mu_1, \dots, \mu_k)$ . The key lemma is that it suffices to compute an optimal vertex solution for (MOT $_S$ ). This is proved below in Lemma 13; let us presently show how to use it complete the proof of the main proposition. Note that since the number of Ellipsoid iterations is  $\text{poly}(n, k, \log U)$  by Theorem 6, the set  $S$  has cardinality of size  $\text{poly}(n, k, \log U)$ . Thus (MOT $_S$ ) is a polynomial size LP, so we can compute a vertex solution for it in polynomial time with standard LP solvers (e.g., Theorem 6).  $\blacksquare$

We now state the key lemma used in the above proof.

**Lemma 13** *Any vertex solution  $P$  for (MOT $_S$ ) is also a vertex solution for (MOT).*

Lemma 13 follows directly from the following three observations. Below, let  $C'$  denote the tensor that agrees with  $C$  on  $S$ , and equals  $2U$  elsewhere. Also let (MOT') denote the problem (MOT) where the cost  $C$  is replaced by  $C'$ .

**Observation 14** *The optimal values of (MOT) and (MOT') are equal.*

**Proof** Since the Ellipsoid algorithm is deterministic and accesses the cost only through the  $\text{SEP}$  oracle, an inductive argument shows that all for all  $l \in [L]$ ,

$$\min_{\vec{j} \in [n]^k} C'_j - \sum_{i=1}^k [p_i^{(l)}]_{j_i} = \min_{\vec{j} \in [n]^k} C_j - \sum_{i=1}^k [p_i^{(l)}]_{j_i}.$$

That is,  $C'$  is consistent with  $C$  on the  $\text{SEP}$  queries made by the Ellipsoid algorithm. Therefore (MOT-D) has the same value with cost  $C'$  or  $C$ . We conclude by strong duality.  $\blacksquare$

**Observation 15** *The optimal values of (MOT') and (MOT $_S$ ) are equal.*

**Proof** Since  $C$  and  $C'$  agree on  $S$ , it suffices to show that every optimal solution for (MOT') is supported on  $S$ . Suppose for contradiction that there exists an optimal solution

$P$  for (MOT') that is not supported on  $S$ . Then we must have  $\langle P, C \rangle < \langle P, C' \rangle$ , since  $C'$  is strictly larger than  $C$  on  $[n]^k \setminus S$ . However, since  $P$  is feasible for (MOT), this implies that the value of (MOT) is strictly less than that of (MOT'), contradicting Observation 14. ■

**Observation 16** *Every vertex of  $\mathcal{M}_S(\mu_1, \dots, \mu_k)$  is a vertex of  $\mathcal{M}(\mu_1, \dots, \mu_k)$ .*

**Proof** Let  $P$  be a vertex of  $\mathcal{M}_S(\mu_1, \dots, \mu_k)$ , and let  $P = \lambda Q + (1 - \lambda)R$  for  $\lambda \in [0, 1]$  and  $Q, R \in \mathcal{M}(\mu_1, \dots, \mu_k)$ . It suffices to show that  $Q = R = P$ . Since  $P$  is supported on tuples in  $S$ , and since  $Q$  and  $R$  are entrywise non-negative, we have that  $Q$  and  $R$  are also supported on  $S$ . Therefore  $Q, R \in \mathcal{M}_S(\mu_1, \dots, \mu_k)$ . But since  $P$  is a vertex of  $\mathcal{M}_S(\mu_1, \dots, \mu_k)$ , this implies that  $Q = R = P$ . ■

### 3.2 Step 2: Efficient Algorithm for the Separation Oracle

Here we prove Proposition 12 and describe the algorithm in it for efficiently implementing the SEP oracle for the cost  $C$  in (2). Recall that this SEP oracle requires computing an optimal tuple for

$$\operatorname{argmin}_{\vec{j} \in [n]^k} \min_{y \in \mathbb{R}^d} g(\vec{j}, y) \quad (3)$$

where

$$g(\vec{j}, y) := \sum_{i=1}^k \lambda_i (\|x_{i,j_i} - y\|^2 - [w_i]_{j_i}),$$

and  $w_i$  denotes  $p_i/\lambda_i$ . At a high level, our approach is to swap the order of minimization, optimize over  $y \in \mathbb{R}^d$ , and then (easily) recover an optimal tuple from this optimal  $y$ . The difficulty is in the optimization over  $y \in \mathbb{R}^d$ . The key to performing this efficiently is partitioning the space  $\mathbb{R}^d$  into a “cell complex” such that (i) the optimization over  $y$  in each cell is easy, and (ii) there are only polynomially many cells. Operationally, this allows us to reduce the separation oracle optimization (3) to optimizing over only a polynomially sized set of candidate tuples in  $[n]^k$ —one for each cell—which we moreover show can be efficiently identified and enumerated.

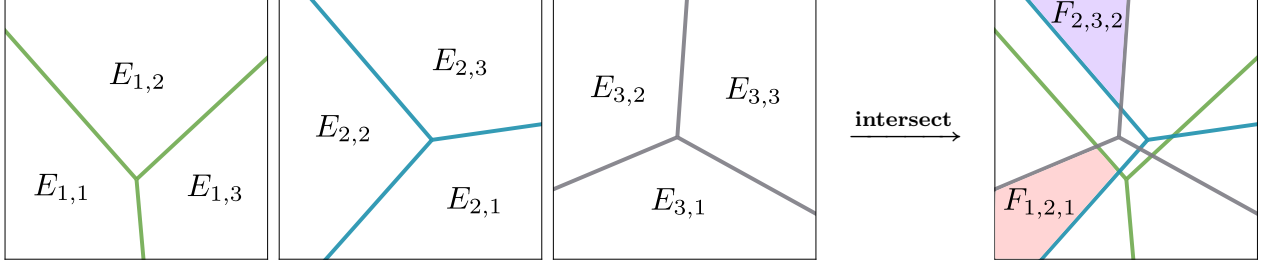
To formalize this, we make the following key definitions. Define for each  $i \in [k]$  and  $j \in [n]$  the set

$$E_{i,j} = \{y \in \mathbb{R}^d : \|x_{i,j} - y\|^2 - [w_i]_j < \|x_{i,j'} - y\|^2 - [w_i]_{j'}, \forall j' \neq j\}, \quad (4)$$

and define for each tuple  $\vec{j} \in [n]^k$  the set

$$F_{\vec{j}} = \bigcap_{i=1}^k E_{i,j_i}. \quad (5)$$

Geometrically, for each  $i \in [k]$ , the cells  $\{E_{i,j}\}_{j \in [n]}$  form a power diagram (see §2.4) on the spheres  $S(x_{i,1}, r_{i,1}), \dots, S(x_{i,n}, r_{i,n})$ , where the  $j$ -th sphere is centered at point  $x_{i,j}$  and has radius  $r_{i,j} := \sqrt{[w_i]_j - \min_{j'} [w_i]_{j'}} \geq 0$ . Each power diagram “essentially” partitions  $\mathbb{R}^d$  in



**Figure 2:** Illustrates  $k = 3$  power diagrams  $\{\{E_{i,j}\}_{j \in [n]}\}_{i \in [k]}$  each with  $n = 3$  cells, and their intersection  $\{F_{\vec{j}}\}_{\vec{j} \in [n]^k}$ . For instance, the red cell in the intersected diagram is  $F_{1,2,1} = E_{1,1} \cap E_{2,2} \cap E_{3,1}$ , and the purple cell is  $F_{2,3,2} = E_{1,2} \cap E_{2,3} \cap E_{3,2}$ . Note that the intersected diagram has only 13 non-empty cells, which is less than  $n^k = 27$  (c.f., Lemma 18).

the sense that its constituent cells are disjoint and their closures cover  $\mathbb{R}^d$ ; see Figure 1 for an illustration. The cell complex  $\{F_{\vec{j}}\}_{\vec{j} \in [n]^k}$  is the intersection of these  $k$  power diagrams and “essentially” partitions  $\mathbb{R}^d$  in the analogous way; see Figure 2 for an illustration.

The heart of our algorithm lies in the following two lemmas. The first lemma shows that the optimization (3) over the exponentially many tuples  $j \in [n]^k$  may be restricted to just those whose corresponding cell  $F_{\vec{j}}$  is non-empty, i.e., we may restrict to the tuples in

$$T := \{\vec{j} \in [n]^k : F_{\vec{j}} \neq \emptyset\}. \quad (6)$$

The second lemma shows that this candidate set  $T$  contains only polynomially many tuples and moreover can be efficiently enumerated. Briefly, the first lemma exploits the fact that the optimization over  $y \in \mathbb{R}^d$  is equivalent to optimizing over the cells in  $F_{\vec{j}}$ , and the second lemma exploits complexity bounds for the intersections of power diagrams. Together, these lemmas are sufficient to efficiently solve the separation oracle because for any fixed  $\vec{j}$ , the value  $\min_{y \in \mathbb{R}^d} g(\vec{j}, y)$  can be efficiently computed in closed-form (as shown below in (7)).

**Lemma 17** *The optimization over  $\vec{j} \in [n]^k$  in the separation oracle problem (3) can be equivalently restricted to  $\vec{j} \in T$ . That is,*

$$\min_{\vec{j} \in [n]^k} \min_{y \in \mathbb{R}^d} g(\vec{j}, y) = \min_{\vec{j} \in T} \min_{y \in \mathbb{R}^d} g(\vec{j}, y).$$

**Proof** The inequality “ $\leq$ ” is obvious; we show the other inequality “ $\geq$ ”. By swapping the order of minimization and using the fact that  $\{F_{\vec{j}}\}_{\vec{j} \in T}$  cover  $\mathbb{R}^d$  modulo closure, we have

$$\min_{\vec{\ell} \in [n]^k} \min_{y \in \mathbb{R}^d} g(\vec{\ell}, y) = \min_{y \in \mathbb{R}^d} \min_{\vec{\ell} \in [n]^k} g(\vec{\ell}, y) = \min_{\vec{j} \in T} \min_{y \in F_{\vec{j}}} \min_{\vec{\ell} \in [n]^k} g(\vec{\ell}, y).$$

We claim that the inner minimization over  $\vec{\ell}$  is explicit:  $\vec{\ell} = \vec{j}$ . Indeed, by separability of  $g$  in the coordinates of  $\vec{\ell}$  and non-negativity of  $\lambda_i$ , for each  $i \in [n]$  the optimal  $\ell_i$  is a solution of  $\operatorname{argmin}_{\ell_i \in [n]} \|x_{i,\ell_i} - y\|^2 - [w_i]_{\ell_i}$ ; and  $j_i$  is a solution of this by definition of  $E_{i,j_i}$  (see (4)) and the fact that  $\overline{E_{i,j_i}}$  contains  $y$  (by definition of  $F_{\vec{j}}$ , see (5)). Therefore

$$\min_{\vec{j} \in T} \min_{y \in F_{\vec{j}}} \min_{\vec{\ell} \in [n]^k} g(\vec{\ell}, y) = \min_{\vec{j} \in T} \min_{y \in F_{\vec{j}}} g(\vec{j}, y).$$

Now by enlarging the optimization region, we have the simple bound

$$\min_{\vec{j} \in T} \min_{y \in F_{\vec{j}}} g(\vec{j}, y) \geq \min_{\vec{j} \in T} \min_{y \in \mathbb{R}^d} g(\vec{j}, y).$$

Combining the above three displays completes the proof. ■

**Lemma 18** *For any fixed dimension  $d$ , the set  $T$  can be enumerated in  $\text{poly}(n, k, \log U)$  time.*

**Proof** By Lemma 8, the  $O(nk)$  total facets for the  $k$  power diagrams  $\{\{E_{i,j}\}_{i \in [n]}\}_{j \in [k]}$  can be computed in  $\text{poly}(n, k, \log U)$  time. For each facet, compute the  $(d-1)$ -dimensional hyperplane it lies in. The cell complex  $\mathcal{H}$  formed by these hyperplanes is a subcomplex of the cell complex formed by intersecting the power diagrams. By Lemma 9, we can enumerate the cells in  $\mathcal{H}$  in  $\text{poly}(n, k, \log U)$  time. For each cell in  $\mathcal{H}$ , the corresponding tuple  $\vec{j} \in [n]^k$  is computable in  $O(nk \cdot \text{polylog } U)$  time by computing the  $k$  coordinates of the tuple separately. Since each non-empty cell  $F_{\vec{j}}$  contains at least one cell in  $\mathcal{H}$ , this process enumerates all tuples in  $T$ . ■

We now conclude the desired efficient algorithm for the separation oracle.

**Proof of Proposition 12** By Lemma 17, it suffices to output a tuple  $\vec{j}$  minimizing  $\min_{\vec{j} \in T} \min_{y \in \mathbb{R}^d} g(\vec{j}, y)$ . Since  $\sum_{i=1}^k \lambda_i x_{i,j_i} \in \text{argmin}_{y \in \mathbb{R}^d} g(\vec{j}, y)$ , it therefore suffices to solve

$$\text{argmin}_{\vec{j} \in T} \sum_{i=1}^k \lambda_i \|x_{i,j_i}\|^2 - \left\| \sum_{i=1}^k \lambda_i x_{i,j_i} \right\|^2 - \sum_{i=1}^k \lambda_i [w_i]_{j_i}. \quad (7)$$

Perform this by enumerating the set  $T$  using the algorithm in Lemma 18. ■

### 3.3 Putting the Pieces Together

**Proof of Theorem 2** Assume each  $x_{i,j_i}$  and  $\lambda_i$  is written to  $\log U$  bits of precision. Since each entry of the cost tensor (2) requires only  $O(\log k + \log U)$  bits of precision, and since the parameter  $w \in \mathbb{R}^{n \times k}$  in each SEP query made by the algorithm in Proposition 11 requires only  $\text{poly}(n, k, \log U)$  bits of precision, it follows that the algorithm in Proposition 11 combined with the SEP oracle implementation in Proposition 12 computes a vertex solution  $P^*$  for (MOT) in  $\text{poly}(n, k, \log U)$  time. By Lemma 4,  $P^*$  has at most  $nk - k + 1$  non-zero entries. Thus we can recover from  $P^*$  an optimal barycenter  $\nu$  with support size at most  $nk - k + 1$  in time  $\text{poly}(n, k, \log U)$  by the reduction in Lemma 3. ■

**Proof of Theorem 1** By rounding both the weights  $\lambda_i$  and the coordinates of the atoms  $x_{i,j} \in \mathbb{R}^d$  to  $\text{poly}(\varepsilon/(Rkd))$  additive accuracy, it can be ensured that each of these numbers requires only  $O(\log(Rkd/\varepsilon))$  bits of precision and also that the objective function  $\nu \mapsto \sum_{i=1}^k \lambda_i \mathcal{W}(\mu_i, \nu)$  for the barycenter optimization (1) is preserved pointwise to  $\varepsilon$  additive

accuracy. This follows from a straightforward calculation and the fact (immediate from the definition of optimal transportation (Villani, 2003, §1) and an application of Hölder’s inequality) that if the squared Euclidean distance between each atom of  $\mu_i$  and each atom of  $\nu$  is preserved up to  $\varepsilon'$  additive accuracy, then the squared 2-Wasserstein distance  $\mathcal{W}(\mu_i, \nu)$  is preserved up to  $\varepsilon'$  additive accuracy. Now solve the barycenter problem for the rounded weights and atoms exactly using Theorem 2. ■

## 4. Numerical Implementation

While the focus of this paper is theoretical, here we briefly mention that a slight variant of our algorithm can provide high-precision solutions at previously intractable problem sizes. To demonstrate this, we implement our algorithm for dimension  $d = 2$  in Python. The only difference between our numerical implementation and the theoretical algorithm described above is that we use a standard cutting-plane method (see, e.g., (Bertsimas and Tsitsiklis, 1997, §6.3)) for the “outer loop” in step 1 rather than the Ellipsoid algorithm due to its good practical performance. Code and further implementation details are provided on Github.<sup>5</sup>

### 4.1 Computing Exact Solutions at Previously Intractable Scales

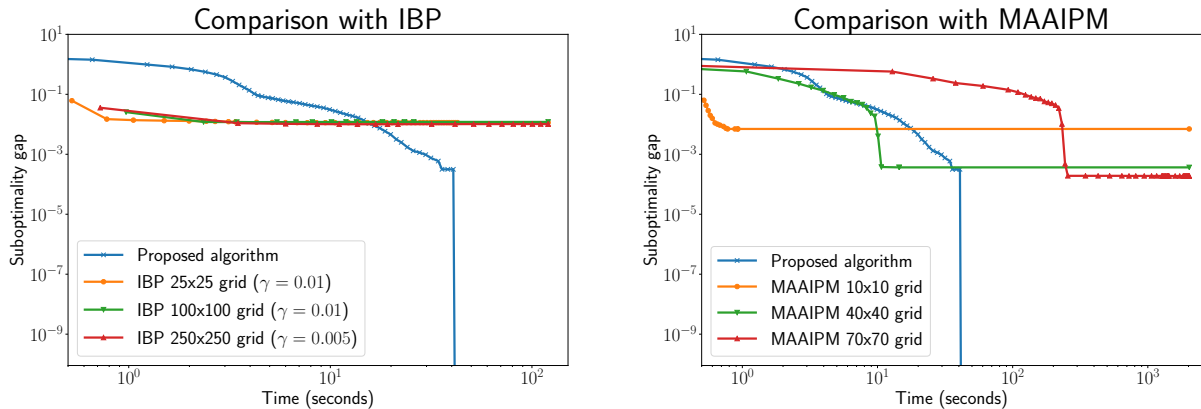
Figure 3 demonstrates that our algorithm solves the barycenter problem (1) to machine precision on an instance with  $k = 10$  uniform distributions each on  $n = 20$  points randomly drawn from  $[-1, 1]^2 \subset \mathbb{R}^2$ . In contrast, existing popular barycenter algorithms which use the fixed-support assumption can converge faster but only to lower-precision approximations. This is because the  $\Theta(1/\varepsilon^d)$  gridsize that they require for  $\varepsilon$ -additive approximation results in a large-scale LP which is prohibitive even for relatively low precision  $\varepsilon$ ; see §1.1 for details. Note also that a standard LP solver requires optimizing over  $n^k = 20^{10} \approx 10^{13}$  variables for the LP formulation (MOT) and thus is clearly infeasible at this scale.

### 4.2 Sharper Visualizations

Here we demonstrate that the high-precision solutions computed by our algorithm yield significantly sharper visualizations than the low-precision solutions that were previously computable. Specifically, here we compare our barycenter algorithm against state-of-the-art methods on a standard benchmark dataset of images of nested ellipses (Cuturi and Doucet, 2014; Janati et al., 2020). This dataset consists of  $k = 10$  images, each of size  $60 \times 60$ . Five of these images are shown in Figure 4. Figure 5 contains a visual comparison of the exact barycenter computed by our algorithm and the approximate barycenters produced by the most competitive algorithms tested in the recent paper (Janati et al., 2020). All are fixed-support algorithms except for the algorithm of (Luise et al., 2019).

Of the compared algorithms, MAAIPM gives the most accurate barycenter approximation. It uses a  $60 \times 60$  grid fixed-support assumption. Although MAAIPM solves this fixed-support problem exactly, the support of an optimal barycenter does not lie on a  $60 \times 60$  grid, and thus MAAIPM only computes an approximate barycenter. A natural approach is

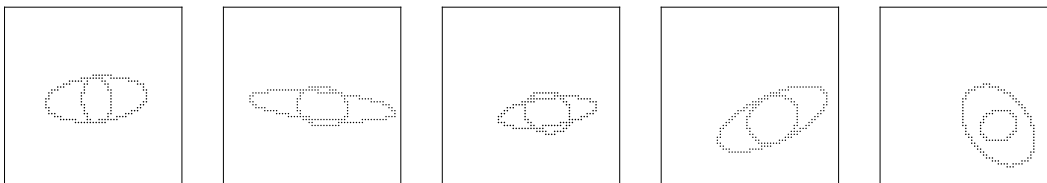
5. [https://github.com/eboix/high\\_precision\\_barycenters](https://github.com/eboix/high_precision_barycenters)



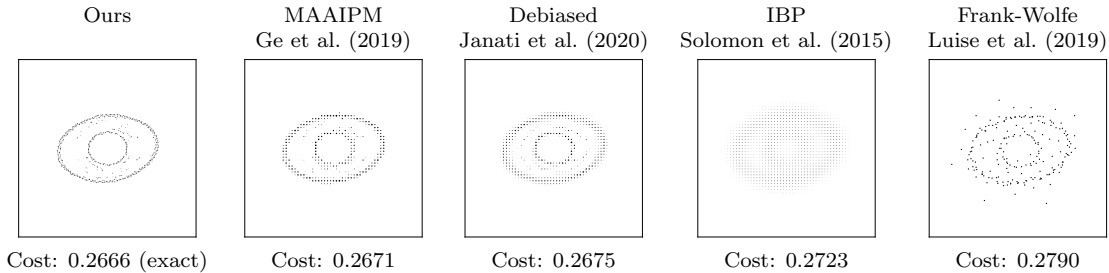
(a) Comparison with the Iterated Bregman Projection (IBP) algorithm of (Solomon et al., 2015) using their implementation <https://github.com/gpeyre/2015-SIGGRAPH-convolutional-ot>.

(b) Comparison with the Matrix-based Adaptive Alternating Interior-Point Method (MAAIPM) of (Ge et al., 2019) using their implementation <https://gitlab.com/ZXiong/wasserstein-barycenter>.

**Figure 3:** Comparison with state-of-the-art algorithms. The  $y$ -axis is the suboptimality for the barycenter optimization (1); note that while standard LP solvers cannot be run at this scale, our algorithm yields an exact solution (certified by our separation oracle) which enables plotting this suboptimality. Both compared algorithms require a fixed-support assumption and are run on uniform grids of increasing sizes. IBP has an additional parameter: the entropic regularization  $\gamma$ , which significantly impacts the algorithm’s accuracy and numerical stability, see (Solomon et al., 2015; Peyré and Cuturi, 2017). We provide a generous comparison here for IBP by (i) fine-tuning  $\gamma$  for it (we binary search for the most accurate  $\gamma$ ; note that their code does not always converge for  $\gamma$  small due to numerical instability); and (ii) exactly computing the Wasserstein distances  $\mathcal{W}(\mu_i, \nu)$  to IBP’s current barycenter  $\nu$  in the barycenter objective (1) using (Dezsó et al., 2011), which is more accurate than IBP’s approximation (this is slow for large grids but is not counted in IBP’s timing). Our algorithm finds an exact barycenter after  $\sim 50$  seconds. All experiments are run on a standard 2014 Lenovo Yoga 720-13IKB laptop.



**Figure 4:** Five sample images from the nested ellipses dataset in (Janati et al., 2020).



**Figure 5:** Comparison of barycenter algorithms on a standard benchmark dataset of ellipse images. Each barycenter atom is plotted as a disk with area proportional to its probability mass. All compared methods are run with the code, parameter choices, and dataset of (Janati et al., 2020).

to run MAAIPM on a finer grid discretization, i.e., finer than  $60 \times 60$ . However, this does not work, since MAAIPM does not scale to much larger grid sizes (see also Figure 3b).

The other two fixed-support algorithms are based on entropic regularization: debiased Sinkhorn barycenters (Janati et al., 2020) and IBP (Solomon et al., 2015). These use entropic regularization parameter  $\gamma = 0.002$  and the same  $60 \times 60$  fixed-support approximation as MAAIPM. Again, these methods produce suboptimal barycenters. While these methods scale to larger grid sizes than MAAIPM, this results in qualitatively similar and blurry visualizations as in this  $60 \times 60$  case due to the entropic regularization.

The final compared algorithm is the free-support algorithm of (Luiße et al., 2019), which is based on the Frank-Wolfe algorithm. Although this method does not make a fixed-support assumption, it still returns an approximate solution due to the approximate nature of the Frank-Wolfe algorithm.

## 5. Discussion

Wasserstein barycenters are used in many applications despite the fact that fundamental questions about their computational complexity are open—in particular, it was previously unknown whether barycenters can be computed in polynomial time. This paper addresses this issue by giving the first algorithm that, in any fixed dimension  $d$ , solves the barycenter problem exactly or to high precision in polynomial time.

Now, while our result answers the polynomial-time computability of barycenters from a *theoretical* perspective, from a *practical* perspective it is still a hard and interesting problem to compute high-precision barycenters for large-scale inputs. Indeed, our current implementation is not efficient beyond moderate-scale inputs; and while existing algorithms such as IBP scale to larger inputs, they have limited accuracy. Moreover, all existing algorithms pay for the curse of dimensionality in one way or another. We emphasize that our implementation does not contain further optimizations or heuristics; it is an interesting direction for future work to investigate potential such options including pruning cutting planes, warm starts, and specially tailored algorithms for the power diagram intersections in §3.2 (e.g., in  $\mathbb{R}^2$  or  $\mathbb{R}^3$ , settings which commonly arise in image processing and computer graphics applications).

We remark that another research direction suggested by the results of this paper is the possibility of solving “structured” Multimarginal Optimal Transport problems in  $\text{poly}(n, k)$

time despite the fact that they have  $n^k$  exponentially many variables. Such problems arise in a variety of applications throughout data science and applied mathematics. However, the context of the application results in the cost tensor  $C$  having different types of “structure”, which in turn necessitates developing different techniques for efficiently solving the corresponding separation oracle (see the discussion in §3). We pursue this direction and develop such techniques in upcoming work (Altschuler and Boix-Adserà, 2020b).

### 5.1 Extension to Wasserstein Geometric Median

We conclude the paper by briefly mentioning that the techniques we develop in this paper extend to solving several related problems. In particular, this gives the first polynomial-time algorithms for computing the *Wasserstein geometric median*

$$\inf_{\nu} \sum_{i=1}^k \lambda_i \rho(\mu_i, \nu) \quad (8)$$

of probability measures  $\mu_1, \dots, \mu_k$ , where  $\rho$  is the 1-Wasserstein distance  $\rho$  (a.k.a., Earth Mover’s Distance) over any of the popular ground metrics  $\ell_1$ ,  $\ell_2$ , or  $\ell_\infty$ . See, e.g., (Bacák, 2014) and the references within for background on this problem and its applications.

For brevity, we just state this result for exact computation (i.e., the analog of Theorem 2). The high-precision analog of Theorem 1 also holds analogously.

**Theorem 19** *Consider the space  $\mathbb{R}^d$  endowed with any of the ground metrics  $\ell_1$ ,  $\ell_2$ , or  $\ell_\infty$ . There is an algorithm that, given  $k$  distributions each supported on  $n$  atoms in  $\mathbb{R}^d$  and a weight vector  $\lambda$ , computes an exact Wasserstein geometric median in  $\text{poly}(n, k, \log U)$  time, where  $\log U$  is the bits of precision in the input. Moreover, this solution has support size at most  $nk - k + 1$ .*

Since the proof of this extension is nearly identical, we sketch only the differences. First, observe that this geometric median problem (8) is identical to the Wasserstein barycenter problem (1), except that it depends on the 1-Wasserstein distance  $\rho$  rather than the squared 2-Wasserstein distance  $\mathcal{W}$ . It can be shown that the geometric median problem admits an analogous LP formulation as a Multimarginal Optimal Transport problem, just as in (MOT), except that here the cost tensor  $C \in (\mathbb{R}^n)^{\otimes k}$  has entries

$$C_j = \min_{y \in \mathbb{R}^d} \sum_{i=1}^k \lambda_i c(x_{i,j_i}, y) \quad (9)$$

where  $c(\cdot, \cdot)$  denotes the relevant ground metric. This is identical to the cost (2) for barycenters, except that the squared Euclidean distance is replaced by the ground metric  $c(\cdot, \cdot)$ .

Now, since step 1 of our algorithm—i.e., reducing finding a vertex solution of (MOT) to solving the separation oracle for (MOT-D)—holds for any Multimarginal Optimal Transport problem (see the discussion in §3), it remains only to adapt step 2 of our algorithm. That is, it suffices to show that SEP can be efficiently implemented for the cost  $C$  in (9).

This extension requires only one small change: replace power diagrams with the analogous partitions of space that are defined with  $\|x - y\|^2$  replaced by  $c(x, y)$ . These diagrams are sometimes called “additively-weighted Voronoi diagrams with metric  $c$ ”; for shorthand, we just call them “ $c$ -diagrams” here.



**Definition 20 (*c*-diagram)** *The  $c$ -diagram for points  $z_1, \dots, z_n \in \mathbb{R}^d$  and radii  $r_1, \dots, r_n \geq 0$  is the cell complex whose cells  $E_1, \dots, E_n$  are given by*

$$E_j = \{y \in \mathbb{R}^d : c(z_j, y) - r_j^2 < c(z_{j'}, y) - r_{j'}^2, \forall j' \neq j\}.$$

Specifically, adapt the definition (4) of the sets  $E_{i,j}$  in this way to  $\{y \in \mathbb{R}^d : c(x_{i,j}, y) - [w_i]_j < c(x_{i,j'}, y) - [w_i]_{j'}, \forall j' \neq j\}$ . It is straightforward to check that our algorithm for the SEP oracle in §3.2 and its proof then extend unchanged so long as the ground metric  $c(\cdot, \cdot)$  satisfies the following basic properties, the first three of which are somewhat trivial but needed for rigor:

- (i) There is a polynomial-time algorithm for evaluating  $c(x, y)$  given points  $x$  and  $y$ .
- (ii) There is a polynomial-time algorithm for evaluating the geometric median  $\min_{y \in \mathbb{R}^d} \sum_{i=1}^k \lambda_i c(x_i, y)$  given points  $x_1, \dots, x_k$  and weights  $\lambda_1, \dots, \lambda_k$ .
- (iii) The closure of the cells in any  $c$ -diagram covers  $\mathbb{R}^d$ .
- (iv) The intersection of any  $k$   $c$ -diagrams on  $n$  points has  $\text{poly}(n, k)$  many non-empty subsets that can be enumerated in polynomial time.

The first three properties are trivially satisfied by all of the ground metrics  $c(x, y) = \|x - y\|_1$ ,  $\|x - y\|_2$ , and  $\|x - y\|_\infty$ . Therefore in order to prove Theorem 19, it remains only to verify property (iv). For  $\ell_1$  and  $\ell_\infty$ ,  $c$ -diagrams are cell complexes with polynomially many affine facets which can moreover be computed in polynomial time (Klein, 1989), and thus the claim follows by Lemma 9. For  $\ell_2$ , the corresponding  $c$ -diagram is an additively-weighted Voronoi diagram, for which the desired complexity bounds are also known (see section 6.4 of (Aurenhammer, 1987)).

## Acknowledgments

We thank Sinho Chewi, Jonathan Niles-Weed, and Pablo Parrilo for helpful conversations. Work partially supported by NSF Graduate Research Fellowship 1122374, a TwoSigma PhD Fellowship, and a Siebel PhD Fellowship.

## References

- Martial Agueh and Guillaume Carlier. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.
- Jason M. Altschuler and Enric Boix-Adserà. Hardness results for Multimarginal Optimal Transport problems. *arXiv:2012.05398*, 2020a.
- Jason M. Altschuler and Enric Boix-Adserà. Polynomial-time algorithms for Multimarginal Optimal Transport problems with decomposable structure. *arXiv:2008.03006*, 2020b.
- Pedro C Álvarez-Esteban, E Del Barrio, JA Cuesta-Albertos, and C Matrán. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2): 744–762, 2016.

- Ethan Anderes, Steffen Borgwardt, and Jacob Miller. Discrete Wasserstein barycenters: Optimal transport for discrete data. *Mathematical Methods of Operations Research*, 84(2):389–409, 2016.
- Franz Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM Journal on Computing*, 16(1):78–96, 1987.
- Miroslav Bacák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, 2014.
- Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.
- Steffen Borgwardt. Strongly polynomial 2-approximations of discrete Wasserstein barycenters. *arXiv:1704.05491*, 2017.
- Steffen Borgwardt and Stephan Patterson. On the computational complexity of finding a sparse Wasserstein barycenter. *arXiv:1910.07568*, 2019.
- Guillaume Carlier, Adam Oberman, and Edouard Oudet. Numerical methods for matching for teams and Wasserstein barycenters. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1621–1642, 2015.
- Bernard Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete & Computational Geometry*, 10(4):377–409, 1993.
- Sinho Chewi, Tyler Maunu, Philippe Rigollet, and Austin J Stromme. Gradient descent algorithms for Bures-Wasserstein barycenters. In *Conference on Learning Theory*, pages 1276–1304, 2020.
- Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, pages 685–693, 2014.
- Balázs Dezső, Alpár Jüttner, and Péter Kovács. LEMON—an open source C++ graph template library. *Electronic Notes in Theoretical Computer Science*, 264(5):23–45, 2011.
- Herbert Edelsbrunner, Joseph O’Rourke, and Raimund Seidel. Constructing arrangements of lines and hyperplanes with applications. *SIAM Journal on Computing*, 15(2):341–363, 1986.
- Filip Elvander, Isabel Haasler, Andreas Jakobsson, and Johan Karlsson. Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion. *Signal Processing*, 171:107474, 2020.
- Dongdong Ge, Haoyue Wang, Zikai Xiong, and Yinyu Ye. Interior-point methods strike back: Solving the Wasserstein barycenter problem. In *Advances in Neural Information Processing Systems*, pages 6891–6902, 2019.
- Martin Grötschel, László Lovász, and Alexander Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.
- Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

- Nhat Ho, Xuan Long Nguyen, Mikhail Yurochkin, Hung Hai Bui, Viet Huynh, and Dinh Phung. Multilevel clustering via Wasserstein means. In *International Conference on Machine Learning*, pages 1501–1509, 2017.
- Hicham Janati, Marco Cuturi, and Alexandre Gramfort. Debiased Sinkhorn barycenters. *International Conference on Machine Learning*, 2020.
- Leonid G Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.
- Rolf Klein. *Concrete and abstract Voronoi diagrams*, volume 400. Springer Science & Business Media, 1989.
- Alexey Kroshnin, Darina Dvinskikh, Pavel Dvurechensky, Alexander Gasnikov, Nazarii Tupitsa, and Cesar Uribe. On the complexity of approximating Wasserstein barycenter. *arXiv:1901.08686*, 2019.
- Tianyi Lin, Nhat Ho, Xi Chen, Marco Cuturi, and Michael Jordan. Fixed-support Wasserstein barycenters: Computational hardness and fast algorithm. *Advances in Neural Information Processing Systems*, 33, 2020.
- Giulia Luise, Saverio Salzo, Massimiliano Pontil, and Carlo Ciliberto. Sinkhorn barycenters with free support via Frank-Wolfe algorithm. In *Advances in Neural Information Processing Systems*, pages 9322–9333, 2019.
- Victor M Panaretos and Yoav Zemel. Statistical aspects of Wasserstein distances. *Annual review of statistics and its application*, 6:405–431, 2019.
- Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 2017.
- Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.
- Ludger Rüschendorf and Ludger Uckelmann. On the  $n$ -coupling problem. *Journal of Multivariate Analysis*, 81(2):242–258, 2002.
- Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics*, 34(4):1–11, 2015.
- Sanvesh Srivastava, Cheng Li, and David B Dunson. Scalable Bayes via barycenter in Wasserstein space. *The Journal of Machine Learning Research*, 19(1):312–346, 2018.
- Matthew Staib, Sebastian Claiici, Justin M Solomon, and Stefanie Jegelka. Parallel streaming Wasserstein barycenters. In *Advances in Neural Information Processing Systems*, pages 2647–2658, 2017.
- Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Society, 2003.