Closing the Gap: Efficient Algorithms for Discrete Wasserstein Barycenters

Jiaqi Wang

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA, jwang3737@gatech.edu

Weijun Xie

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA, wxie@gatech.edu

Abstract

The Wasserstein barycenter problem seeks a probability measure that minimizes the weighted average of the Wasserstein distances to a given collection of probability measures. We study the discrete setting, where each measure has finite support — a regime that frequently arises in machine learning and operations research. The discrete Wasserstein barycenter problem is known to be NP-hard, which motivates us to study approximation algorithms with provable guarantees. The best-known algorithm to date achieves an approximation ratio of two. We close this gap by developing a polynomial-time approximation scheme (PTAS) for the discrete Wasserstein barycenter problem that generalizes and improves upon the 2-approximation method. In addition, for the special case of equally weighted measures, we obtain a strictly tighter approximation guarantee. Numerical experiments show that the proposed algorithms are computationally efficient and produce near-optimal barycenter solutions.

1 Introduction

A Wasserstein barycenter, also known as the "center of mass" in Wasserstein space, is the probability measure that minimizes the average optimal-transport cost using the type-2 Wasserstein metric for a given collection of measures. The type-2 Wasserstein distance between $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^d)$ is

$$W_2(\mathbb{P}, \mathbb{Q}) := \left(\inf_{\Pi \in \mathcal{M}(\mathbb{P},\mathbb{Q})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \, d\Pi(\boldsymbol{x}, \boldsymbol{y}) \right)^{1/2},$$

where $\mathcal{M}(\mathbb{P}, \mathbb{Q})$ denotes the set of joint distributions with marginals $\mathbb{P}$ and $\mathbb{Q}$.

Let $\{\mathbb{P}_i\}_{i\in[k]}$ be discrete marginal probability measures with $[k] := \{1, \ldots, k\}$, each supported at most on n points, denoted by $\{\Xi_i\}_{i\in[k]} \subseteq \mathbb{R}^d$. For notational convenience, for each discrete probability measure $\mathbb{P}_i$ with $i \in [k]$, we let $\hat{\boldsymbol{x}}_{ij}$ denote the j-th support point of $\mathbb{P}_i$ with mass $\hat{p}_{ij}$ for any $j \in [|\Xi_i|]$. Given weights $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k) \in \Delta_k := \{\boldsymbol{\lambda} \in \mathbb{R}^k_+ : \sum_{i\in[k]} \lambda_i = 1\}$, the discrete Wasserstein barycenter can be formulated as

$$v^* = \inf_{\mathbb{P}\in\mathcal{P}(\mathbb{R}^d)} \sum_{i\in[k]} \lambda_i \, W_2^2 (\mathbb{P}, \mathbb{P}_i), \tag{WBCenter}$$

where $\mathcal{P}(\mathbb{R}^d)$ is the set of Borel probability measures on $\mathbb{R}^d$ with finite second moment. The Wasserstein barycenter problems are fundamental and have been widely applied to clustering (Ye et al. 2017), regression (Bonneel et al. 2016), dictionary learning (Schmitz et al. 2018), texture mixing (Bonneel et al. 2015), image morphing (Simon & Aberdam 2020), medical imaging (Janati et al. 2020*a*), facial detection (Yan et al. 2021), time series modeling (Cheng et al. 2021), and distributionally robust optimization (Lau & Liu 2022).

1.1 Literature review

The notion of the Wasserstein barycenter was introduced in Agueh & Carlier (2011), which also connected it to the multi-marginal optimal transport (MOT) problem. The Wasserstein barycenter problem has also been extended to different notions (see, e.g., Bigot et al. 2012, Huang et al. 2021, Uribe et al. 2018). For discrete probability measures with finite support, Anderes et al. (2016) established foundational structural properties—including discrete support, sparsity, and a non–mass-splitting property of the optimal barycenter—and proposed a linear programming (LP) formulation via MOT. However, the resulting LP formulation has $O(n^k)$ variables and constraints, which are exponential in the number of marginals k and become computationally expensive as n and k grow. For the *sparse* Wasserstein barycenter problem (i.e., when the optimal barycenter has sparse support), Borgwardt & Patterson (2021) proved NP-hardness even in dimension $d = 2$ with only $k = 3$ measures. For the general barycenter problem, Altschuler & Boix-Adsera (2022) showed that solving it to optimality is intractable unless P = NP. These hardness results motivate the development of efficient approximation algorithms. However, existing work is limited. Borgwardt (2022) proposed a 2-approximation that takes the union of the input supports as a candidate support, and Lindheim (2023) gave another 2-approximation based on a sequence of two-marginal optimal transport plans. In this paper, we improve the factor-2 guarantee. In particular, we develop a polynomial-time approximation scheme (PTAS) for the Wasserstein barycenter problem. That is, for any $\alpha \in (0, 1]$, our algorithm computes a $(1 + \alpha)$-approximate Wasserstein barycenter in time polynomial in $(nk)^{1/\alpha}$ and d.

Despite NP-hardness, there is a line of work on computing Wasserstein barycenters to optimality. In fixed dimension d, Altschuler & Boix-Adsera (2021) showed that one can compute an exact type-p Wasserstein barycenter for $p \in \{1, 2\}$ in time polynomial in n, k, and $\log U$ (where U is the upper bound of the input bit-length), or an approximate barycenter with additive error $\epsilon > 0$ in time polynomial in n, k, and $\log(1/\epsilon)$, by solving an exponential-size LP via an efficient separation oracle based on power diagrams. Beyond the fixed-dimension method, Borgwardt & Patterson (2020) proposed strengthened LP formulations that exploit the non-mass-splitting structure of optimal barycenters. Their models first encoded that each barycenter support point aggregates mass from a single tuple of input support points, then aggregated transport variables at the level of such tuples, therefore shrinking the variable set relative to the naïve MOT LP formulation, and finally incorporated preprocessing/pruning rules and structural bounds on the barycenter support size, which together reduce problem size substantially in practice. Building on these structural ideas, Borgwardt & Patterson (2022) developed a column-generation framework whose restricted master problem uses compact LP, while the pricing subproblem searches for new barycenter atoms (i.e., support tuples) with a negative reduced cost. They design efficient pricing routines with bounding and heuristics, stabilization, and warm starts to generate exact solutions. However, none of these approaches yields a polynomial-time algorithm in the general input parameters, and they remain challenging to apply at scale when n, k, and d are large.

A related line of work considers a simpler variant in which the candidate support of the barycenter is fixed *a priori*, reducing the problem to a linear program. This fixed-support formulation can still be computationally challenging when the given support is large. To address this, Cuturi & Doucet (2014) introduced entropic regularization and developed subgradient-based schemes; Benamou et al. (2015) proposed a tractable, parallelizable iterative Bregman projection (Sinkhorn) method that achieves scalability at the cost of an entropic bias; and Janati et al. (2020*b*) provided a debiased variant that removes this regularization bias while preserving convergence rates comparable to the Sinkhorn approach. Other directions include an interior-point method (Ge et al. 2019), a fast deterministic Bregman-projection algorithm with improved complexity guarantees (Lin et al. 2020), and a symmetric Gauss–Seidel-based ADMM from a dual perspective (Yang et al. 2021) for the given-support barycenter LP formulation. However, none of these methods can be applied to the Wasserstein barycenter problems studied in this paper.

For the general Wasserstein barycenter problem, Claici et al. (2018) proposed a stochastic scheme that updates the barycenter support by sampling subgradients from the input measures. Luise et al. (2019) developed a Frank–Wolfe optimization strategy for estimating barycenters with respect to the Sinkhorn divergence and established convergence guarantees. Lin & Ruszczyński (2025) formulated the free-support barycenter problem as an integer program and designed a federated algorithm via a dual subgradient method. Kroshnin et al. (2019) analyzed the complexity of iterative Bregman projections and accelerated gradient methods for the entropically regularized barycenter objective. Despite these advances, none of the above methods yield a polynomial-time algorithm in the general input parameters (n, k, d) (and, when applicable, $(nk)^{1/\alpha}$ and d).

1.2 Summary of contributions

In this paper, we develop a unified framework of approximation algorithms for the discrete Wasserstein barycenter problem (WBCenter). Our approach is based on systematic candidate-support reduction and yields provable guarantees that improve upon the best-known results in the literature. The main contributions are summarized below.

(i) We establish the first polynomial-time approximation scheme (PTAS) for the Wasserstein barycenter problem. In particular, we propose a family of algorithms parameterized by an integer t that achieve a $(1 + \alpha)$-approximation guarantee with running time polynomial in $(nk)^{1/\alpha}$ and d.

(ii) We present both randomized (sampling-based) and deterministic variants, with the same approximation guarantee. In the special case of equally weighted input measures, we further refine the construction to obtain an improved approximation ratio.

(iii) We evaluate the proposed algorithms on both synthetic and real datasets, including instances with many input measures and instances with a large support size. The results demonstrate that our methods are computationally practical and produce high-quality barycenters in challenging regimes.

The remainder of the paper is organized as follows. In Section 2, we present preliminary results that form the basis of our approximation framework. In Section 3, we develop a subset sampling algorithm and its deterministic counterpart, and analyze their computational complexity and approximation guarantees. In Section 4, we show that, under uniform weights, the approximation ratio can be further improved via a modified sampling procedure and its deterministic counterpart. Section 5 reports numerical results, and Section 6 concludes the paper.

Notation. Let $[k] := \{1, \ldots, k\}$, $\|\cdot\|_2$ denote the ℓ_2-norm, $\delta_{(\cdot)}$ denote the Dirac measure, $\mathbb{I}_{(\cdot)}$ denote the indicator function, $(\mathbb{R}^n_+)^{\otimes k}$ denote the k-fold tensor product of the nonnegative orthant $\mathbb{R}^d_+$, $\mathbb{P}(S)$ denote a probability measure with support S, $O(\cdot)$ denote the standard big-O notation for computational complexity, $\tilde{O}(\cdot)$ denote the soft-O notation for computational complexity, which suppresses logarithmic and polylogarithmic factors in input size and accuracy.

2 Preliminary results and approximation scheme

In this section, we establish the theoretical foundation for our approximation framework, which generalizes the approach of Borgwardt (2022). We first recall the equivalence between the discrete Wasserstein barycenter problem and the multi-marginal optimal transport (MOT) problem, which implies that an optimal barycenter may require an exponentially large support set S^*. To address this intractability, we introduce a restricted MOT formulation in which the barycenter support is constrained to be a smaller candidate set S, and we show that this restricted problem admits a separation oracle with complexity polynomial in n, k, d, and $|S|$, making it tractable whenever $|S|$ is polynomial in the input size. We further show how to recover an explicit barycenter measure supported on S from the optimal solution of the restricted MOT problem, and we prove that if set S appropriately approximates the optimal support S^*, then the resulting restricted barycenter achieves an objective value within a provable multiplicative factor of the true optimum.

2.1 Optimal Support

A key ingredient of our results is the well-known equivalence between the Wasserstein barycenter problem and the multi-marginal optimal transport (MOT) problem. It has been shown (see Proposition 2.1 in Altschuler & Boix-Adsera (2022) and Section 6 in Carlier & Ekeland (2010)) that the Wasserstein barycenter problem (WBCenter) is equivalent to the following MOT formulation:

$$\begin{aligned} v^* = \min_{\mathbf{\Pi} \in (\mathbb{R}^n_+)^{\otimes k}} & \langle \boldsymbol{C}, \mathbf{\Pi} \rangle, \\ \text{s.t.} & \sum_{j_1 \in [|\Xi_1|]} \cdots \sum_{j_{i-1} \in [|\Xi_{i-1}|]} \sum_{j_{i+1} \in [|\Xi_{i+1}|]} \cdots \sum_{j_k \in [|\Xi_k|]} \Pi_{j_1,\ldots,j_k} = \hat{p}_{ij_i}, \forall i \in [k], j_i \in [|\Xi_i|], \end{aligned} \tag{1}$$

where $\mathbf{\Pi}$ is a k-way coupling tensor and the transportation cost is given by

$$C_{\vec{j}} = \min_{\boldsymbol{w} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\hat{\boldsymbol{x}}_{i,j_i} - \boldsymbol{w}\|_2^2, \qquad \vec{j} = (j_1, \ldots, j_{i-1}, j_i, j_{i+1}, \ldots, j_k).$$

In other words, each entry $\Pi_{j_1,\ldots,j_k}$ represents the joint mass assigned to the tuple of support points $(\hat{\boldsymbol{x}}_{1,j_1}, \ldots, \hat{\boldsymbol{x}}_{k,j_k})$, and $C_{\vec{j}}$ is the optimal cost of fitting a single barycenter location $\boldsymbol{w}$ to that tuple under the weighted squared Euclidean loss.

From the MOT formulation, we observe that the support of the optimal barycenter is finite and consists only of points which attain the minimum cost in the definition of $\boldsymbol{C}$.

Lemma 1. The support S^* of the optimal barycenter of problem (WBCenter) is given by

$$S^* = \bigcup_{\boldsymbol{x}_i \in \Xi_i, i \in [k]} \left\{ \arg\min_{\boldsymbol{w} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\boldsymbol{x}_i - \boldsymbol{w}\|_2^2 \right\}.$$

This characterization appears in the remark following Proposition 2.1 of Altschuler & Boix-Adsera (2022) and in Section 6 of Carlier & Ekeland (2010), and is consistent with known finiteness results for discrete barycenters of (WBCenter) (see, e.g., Anderes et al. 2016). In particular, the optimal barycenter can be chosen to have finite support, and all its support points arise as minimizers of a weighted least-squares fit to some k-tuple of input support points.

By Lemma 1, we have $|S^*| = O(n^k)$, which is exponential in k. Using this notation, we may equivalently rewrite the transportation cost in the MOT formulation as

$$C_{\vec{j}} = \min_{\boldsymbol{w} \in S^*} \sum_{i \in [k]} \lambda_i \|\hat{\boldsymbol{x}}_{i,j_i} - \boldsymbol{w}\|_2^2,$$

for each index tuple $\vec{j} = (j_1, \ldots, j_k)$.

Lemma 1 also shows that, although an optimal barycenter exists and is finitely supported, its support S^* may be exponentially large. This motivates the construction of a reduced candidate support in our approximation framework.

2.2 The dual representation of MOT formulation and candidate support reduction

Let $\{\gamma_{ij}\}_{i \in [k], j \in [|\Xi_i|]}$ be the dual variables associated with the marginal constraints in the MOT formulation (1). By LP strong duality, the optimal value v^* of the Wasserstein barycenter problem admits the dual representation:

$$v^* = \max_{\boldsymbol{\gamma}} \left\{ \sum_{i \in [k]} \sum_{j \in [|\Xi_i|]} \hat{p}_{ij} \gamma_{ij} : C_{\vec{j}} - \sum_{i \in [k]} \gamma_{ij_i} \geq 0, \forall \vec{j} \in \otimes_{i \in [k]} [|\Xi_i|], \vec{j} = (j_1, \ldots, j_k) \right\}. \tag{2}$$

This dual problem involves at most nk variables γ_{ij}, which are polynomial in input size. However, enforcing feasibility in the dual is still challenging, because each dual constraint depends on the transportation cost $C_{\vec{j}}$ and the optimal support set S^* may contain exponentially many candidates, i.e., $|S^*|= O(n^k)$. Thus, even though the dual has only $O(nk)$ variables, it is not directly tractable: (i) there are exponentially many dual constraints (one per k-tuple $\vec{j}$), and (ii) evaluating each left-hand side requires minimizing over S^*, which itself is exponentially large.

This motivates our approach. Rather than working with the full optimal support S^*, we seek a reduced candidate support $S \subseteq \mathbb{R}^d$ of manageable size such that replacing S^* by S perturbs the cost $C_{\vec{j}}$ by only a controlled multiplicative factor. This support reduction is the core mechanism that enables our polynomial-time approximation scheme.

For notational convenience, given a candidate support $S = \{\boldsymbol{w}_\ell\}_{\ell \in [|S|]} \subseteq \mathbb{R}^d$ for a barycenter $\mathbb{P}(S)$, we define its transportation cost as

$$C_{\vec{j}}(S) = \min_{\boldsymbol{w} \in S} \sum_{i \in [k]} \lambda_i \|\hat{\boldsymbol{x}}_{i,j_i} - \boldsymbol{w}\|_2^2,$$

for each multi-index $\vec{j} = (j_1, \dots, j_k)$.

We then define the corresponding restricted dual objective value

$$v(S) = \max_{\boldsymbol{\gamma}} \left\{ \sum_{i \in [k]} \sum_{j \in [|\Xi_i|]} \hat{p}_{ij}\gamma_{ij} : C_{\vec{j}}(S) - \sum_{i \in [k]} \gamma_{ij_i} \geq 0, \forall \vec{j} \in \otimes_{i \in [k]}[|\Xi_i|], \vec{j} = (j_1, \dots, j_k) \right\}. \tag{3}$$

In other words, $v(S)$ is the optimal dual objective value when we restrict all barycenter support to be S instead of the full optimal support.

When the candidate support size $|S|$ is polynomial in n and k, the restricted MOT dual (3) can be solved in polynomial time, since its separation oracle admits a closed-form evaluation..

Lemma 2. Given a candidate dual vector $\boldsymbol{\gamma}$ for (3), feasibility of $\boldsymbol{\gamma}$ can be checked in time $O(kn|S|d)$. Therefore, the restricted MOT dual (3) can be solved in time $\tilde{O}(n^3k^3d|S|)$.

Proof. We observe that the fact that $\boldsymbol{\gamma}$ is feasible to (3) is equivalent to

$$\min_{\vec{j} \in \otimes_{i \in [k]}[|\Xi_i|]} C_{\vec{j}}(S) - \sum_{i \in [k]} \gamma_{ij_i} \geq 0,$$

which is equivalent to the equivalent two-stage minimization problem:

$$\min_{\ell \in [|S|]} \min_{\vec{j} \in \otimes_{i \in [k]}[|\Xi_i|]} \sum_{i \in [k]} \lambda_i \|\boldsymbol{w}_\ell - \hat{\boldsymbol{x}}_{ij_i}\|_2^2 - \sum_{i \in [k]} \gamma_{ij_i} \geq 0.$$

Note that the inner minimization problem can be decomposed for each $i \in [k]$ when $\ell \in [|S|]$ is fixed. Let us define the cost $c_{ij_i\ell} = \lambda_i \|\boldsymbol{w}_\ell - \hat{\boldsymbol{x}}_{ij_i}\|_2^2 - \gamma_{ij_i}$. Then for each $i \in [k]$, let $j_i^*(\ell) \in \arg\min_{j_i} c_{ij_i\ell}$. Next, we solve the outer minimization by picking the ℓ^* such that $\ell^* \in \arg\min_{\ell \in [|S|]} \sum_{i \in [k]} c_{ij_i^*(\ell)\ell}$.

Computing each cost coefficient $c_{ij_i\ell}$ requires $O(d)$ time. Therefore, computing all such coefficients costs $O(kn|S|d)$ in total, solving the inner minimization requires $O(kn|S|)$, and solving the outer minimization takes $O(k|S|)$. In total, the complexity of the separation oracle is $O(kn|S|d)$.

By the complexity result of the ellipsoid method with a separation oracle (Grötschel et al. 2012), the overall complexity for solving the restricted MOT dual (3) is $\tilde{O}((nk)^2 kn|S|d) = \tilde{O}(k^3 n^3 |S|d)$. $\square$

We also show that a feasible barycenter of a given support can be explicitly constructed from the MOT formulation (1).

Lemma 3. Fix a candidate support $S \subseteq \mathbb{R}^d$. Let $\mathbf{\Pi}^*(S)$ be an optimal primal solution of the MOT formulation (1) in which the cost $C_{\vec{j}}$ is restricted to $C_{\vec{j}}(S)$. For each $\vec{j} \in \otimes_{i\in[k]}[|\Xi_i|]$ such that $\Pi^*_{\vec{j}}(S) > 0$, choose

$$\boldsymbol{w}_{\vec{j}} \in \arg\min_{\boldsymbol{w}\in S} \sum_{i\in[k]} \lambda_i \|\hat{\boldsymbol{x}}_{i,j_i} - \boldsymbol{w}\|_2^2, \qquad m_{\boldsymbol{w}} = \sum_{\vec{j}\in\otimes_{i\in[k]}[|\Xi_i|]} \Pi^*_{\vec{j}}(S)\mathbb{I}_{\{\boldsymbol{w}_{\vec{j}}=\boldsymbol{w}\}}, \forall \boldsymbol{w} \in S,$$

where $\mathbb{I}_{(\cdot)}$ is the indicator function. Then the measure $\mathbb{P}(S) = \sum_{\boldsymbol{w}\in S} m_{\boldsymbol{w}} \delta_{\tilde{\boldsymbol{w}}=\boldsymbol{w}}$ is a feasible barycenter supported on S, where $\delta(\cdot)$ is the Dirac measure and the random variable $\tilde{\boldsymbol{w}}$ follows the distribution of $\mathbb{P}(S)$.

The MOT primal solution $\mathbf{\Pi}^*(S)$ tells how often each tuple of support points $(\hat{\boldsymbol{x}}_{1,j_1}, \ldots, \hat{\boldsymbol{x}}_{k,j_k})$ is jointly selected. For each such tuple, we attach the "best-fitting" barycenter location in S, i.e., the minimizer $\boldsymbol{w}_{\vec{j}}$. Aggregating the mass $\Pi^*_{\vec{j}}(S)$ onto these chosen points yields a discrete measure $\mathbb{P}(S)$ on S. This measure is therefore a feasible barycenter with support contained in S.

We conclude this section by stating an approximation guarantee that quantifies the effect of reducing the full optimal support S^* to a smaller candidate set $\tilde{S}$.

Theorem 1. Let $\tilde{S} \subseteq \mathbb{R}^d$ be a (possibly random) candidate support set. Suppose there exists $\alpha \geq 0$ such that, for any choice of points $\boldsymbol{x}_i \in \Xi_i$ for all $i \in [k]$, we have

$$\mathbb{E}_{\tilde{S}}\left[\min_{\boldsymbol{w}\in\tilde{S}} \sum_{i\in[k]} \lambda_i \|\boldsymbol{w} - \boldsymbol{x}_i\|_2^2\right] \leq (1+\alpha) \min_{\boldsymbol{w}\in\mathbb{R}^d} \sum_{i\in[k]} \lambda_i \|\boldsymbol{w} - \boldsymbol{x}_i\|_2^2. \tag{4}$$

Then the expected objective value of the restricted barycenter problem satisfies $\mathbb{E}_{\tilde{S}}[v(\tilde{S})] \leq (1+\alpha)v(S^*)$.

Proof. We first observe that in the MOT formulation (1), replacing transportation cost $\boldsymbol{C}$ by the approximate one $\boldsymbol{C}(\tilde{S})$, the value of $v(\tilde{S})$ can be obtained by solving the following restricted MOT problem:

$$\begin{aligned} v(\tilde{S}) = & \min_{\mathbf{\Pi}\in(\mathbb{R}^n_+)^{\otimes k}} \langle \boldsymbol{C}(\tilde{S}), \mathbf{\Pi}\rangle, \\ \text{s.t. } & \sum_{j_1\in[|\Xi_1|]} \cdots \sum_{j_{i-1}\in[|\Xi_{i-1}|]} \sum_{j_{i+1}\in[|\Xi_{i+1}|]} \cdots \sum_{j_k\in[|\Xi_k|]} \Pi_{j_1,\ldots,j_{i-1},j_i,j_{i+1},\ldots,j_k} = \hat{p}_{ij_i}, \forall i \in [k], j_i \in [|\Xi_i|]. \end{aligned} \tag{5}$$

Suppose that $\mathbf{\Pi}^*$ is an optimal solution of the MOT formulation (1), which is also feasible to the restricted MOT formulation (5). Thus,

$$v(\tilde{S}) \leq \langle \boldsymbol{C}(\tilde{S}), \boldsymbol{\Pi}^* \rangle = \sum_{\vec{j} \in \otimes_{i\in[k]}[|\Xi_i|]} C_{\vec{j}}(\tilde{S}) \Pi^*_{\vec{j}},$$

where $\vec{j} = (j_1, \ldots, j_k)$. Taking expectation on both sides of the inequality, we have

$$\mathbb{E}_{\tilde{S}}\left[v(\tilde{S})\right] \leq \sum_{\vec{j} \in \otimes_{i\in[k]}[|\Xi_i|]} \mathbb{E}_{\tilde{S}}\left[C_{\vec{j}}(\tilde{S})\Pi^*_{\vec{j}}\right] \leq (1+\alpha) \sum_{\vec{j} \in \otimes_{i\in[k]}[|\Xi_i|]} \mathbb{E}_{\tilde{S}}\left[C_{\vec{j}}(S^*)\Pi^*_{\vec{j}}\right] := (1+\alpha)v(S^*),$$

where the second inequality is due to the condition (4) and the fact that $\Pi^*_{\vec{j}} \geq 0$ for any $\vec{j}$. $\square$

Remark. If the set $\tilde{S}$ is deterministic, then the same guarantee holds directly under condition (4) (i.e., the expectation over $\tilde{S}$ is not needed).

By Theorem 1, if every support point of the optimal barycenter has a "good enough" representative in the reduced candidate set, then solving the barycenter problem over that reduced set yields an objective value within a factor $(1+\alpha)$ of the optimal one. In the next section, we therefore construct such a reduced support $\tilde{S}$ that satisfies (4), either deterministically or in expectation.

3 Approximation algorithms for the Wasserstein barycenter problem with general weights

By Lemma 1, the optimal barycenter of the Wasserstein barycenter problem admits a finite support $S^* \subseteq \mathbb{R}^d$ of the form

$$S^* := \left\{ \sum_{i\in[k]} \lambda_i \boldsymbol{x}_i : \boldsymbol{x}_i \in \Xi_i, \forall i \in [k] \right\}.$$

The cardinality of this set satisfies $|S^*| = \prod_{i\in[k]} |\Xi_i| = O(n^k)$, which is exponential in the number of measures k. In this section, we develop two approximation schemes for the case of general (i.e., not necessarily uniform) weights $\boldsymbol{\lambda} \in \mathbb{R}^k_+$: a randomized subset sampling algorithm and a deterministic subset enumeration algorithm. We show that both methods achieve the same approximation guarantee.

3.1 Sampling algorithm with repetition

We begin by drawing t indices $T_1, \ldots, T_t \in [k]$ independently according to the weight vector $\boldsymbol{\lambda}$; i.e.,

$$\text{Prob}\left[T_j = i\right] = \lambda_i, \forall i \in [k], j \in [t].$$

Let $\mathcal{T}$ denote the resulting multiset of selected indices. Based on $\mathcal{T}$, we construct the candidate support set

$$S_1^{\mathcal{T}} = \left\{ \frac{1}{t} \sum_{i \in [t]} \boldsymbol{x}_{T_i} : \boldsymbol{x}_{T_i} \in \Xi_{T_i}, \forall i \in [t] \right\}.$$

Note that each element of $S_1^{\mathcal{T}}$ is obtained by averaging t support points, where at each position $i \in [t]$ we select one atom from the support of the probability measure indexed by T_i. The cardinality of $S_1^{\mathcal{T}}$ satisfies $|S_1^{\mathcal{T}}| = O(n^t)$, which is polynomial in n for any fixed t. We then solve the restricted MOT dual (3) using $S_1^{\mathcal{T}}$ as the candidate support, and obtain an approximation barycenter $\mathbb{P}(S_1^{\mathcal{T}})$ based on Lemma 3. The complete procedure is summarized in Algorithm 1.

Algorithm 1 Multi-subset sampling algorithm for Wasserstein barycenter problem

---

1: **Input**: Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$, respectively; a constant $t \in [k]$.
2: Uniformly sample a size-t multi-subset $\mathcal{T} = \{T_i\}_{i \in [k]}$ from subset $[k]$.
3: $S_1^{\mathcal{T}} := \left\{ \frac{1}{t} \sum_{i \in [t]} \boldsymbol{x}_{T_i} : \boldsymbol{x}_{T_i} \in \Xi_{T_i} \right\}$.
4: Solve the restricted MOT dual (3) and obtain its corresponding primal optimal solution $\boldsymbol{\Pi}^*(S_1^{\mathcal{T}})$.
5: **Output**: Obtain an approximation barycenter $\mathbb{P}(S_1^{\mathcal{T}})$ based on Lemma 3.

The approximation guarantee is formally stated in the following theorem.

Theorem 2. Let $t \in [k]$ be a fixed constant, and let $S_1^{\mathcal{T}}$ be the candidate support returned by Algorithm 1. Then

$$\mathbb{E}_{\mathcal{T}} \left[v \left(S_1^{\mathcal{T}} \right) \right] \leq \left(1 + \frac{1}{t} \right) v^*.$$

Proof. By Theorem 1, it is sufficient to show that for any fixed realization $\boldsymbol{x}_i \in \Xi_i$, $i \in [k]$ and $\boldsymbol{c} := \sum_{i \in [k]} \lambda_i \boldsymbol{x}_i$, we have

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\boldsymbol{s} \in S_1^{\mathcal{T}}} \sum_{i \in [k]} \lambda_i \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2 \right] \leq \left(1 + \frac{1}{t} \right) \sum_{i \in [k]} \lambda_i \|\boldsymbol{c} - \boldsymbol{x}_i\|_2^2. \tag{6a}$$

Since we consider a particular realization from each marginal distribution, for each random sample T_j, we also consider the same realization; i.e., we let $\boldsymbol{x}_{T_j} = \sum_{i \in [k]} \boldsymbol{x}_i \mathbb{I}_{\{T_j = i\}}$.

Note that $\boldsymbol{c} \in \arg\min_{\boldsymbol{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2$. Therefore, to approximate the solution of $\min_{\boldsymbol{s} \in S_1^{\mathcal{T}}} \sum_{i \in [k]} \lambda_i \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2$, we consider $\boldsymbol{s}_c \in \arg\min_{\boldsymbol{s} \in S_1^{\mathcal{T}}} \|\boldsymbol{s} - \boldsymbol{c}\|_2^2$; i.e., the point in $S_1^{\mathcal{T}}$ that is closest to $\boldsymbol{c}$. We have

$$\sum_{i \in [k]} \lambda_i \|\boldsymbol{s}_c - \boldsymbol{x}_i\|_2^2 = \boldsymbol{s}_c^\top \boldsymbol{s}_c - 2\boldsymbol{c}^\top \boldsymbol{s}_c + \sum_{i \in [k]} \lambda_i \boldsymbol{x}_i^\top \boldsymbol{x}_i = \sum_{i \in [k]} \lambda_i (\|\boldsymbol{s}_c - \boldsymbol{c}\|_2^2 + \|\boldsymbol{c} - \boldsymbol{x}_i\|_2^2), \tag{6b}$$

where the equalities follow by direct expansion and $\sum_{i \in [k]} \lambda_i = 1$.

Since $\boldsymbol{s}_c \in S_1^{\mathcal{T}}$, by (6b), the left-hand side of (6a) can be upper bounded as

$$\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_1^{\mathcal{T}}}\sum_{i\in[k]}\lambda_i\|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2\right] \le \sum_{i\in[k]}\lambda_i\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2+\mathbb{E}_{\mathcal{T}}\left[\|\boldsymbol{s}_c-\boldsymbol{c}\|_2^2\right]. \tag{6c}$$

It remains to bound $\mathbb{E}_{\mathcal{T}}\left[\|\boldsymbol{s}_c-\boldsymbol{c}\|_2^2\right]$. Since $\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t} \in S_1^{\mathcal{T}}$, we have

$$\mathbb{E}_{\mathcal{T}}\left[\|\boldsymbol{s}_c-\boldsymbol{c}\|_2^2\right] = \mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_1^{\mathcal{T}}}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right] \le \mathbb{E}_{\mathcal{T}}\left[\left\|\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t}-\boldsymbol{c}\right\|_2^2\right]. \tag{6d}$$

Since $\mathcal{T}=\{T_i\}_{i\in[k]}$ and T_i's are i.i.d., we have

$$\mathbb{E}\left[\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t}\right] = \frac{\sum_{j\in[t]}\sum_{i\in[k]}\mathrm{Prob}\left[T_j=i\right]\boldsymbol{x}_i}{t} = \frac{t\sum_{i\in[k]}\lambda_i\boldsymbol{x}_i}{t} = \sum_{i\in[k]}\lambda_i\boldsymbol{x}_i = \boldsymbol{c}.$$

Therefore, we have

$$\mathbb{E}_{\mathcal{T}}\left[\|\boldsymbol{s}_c-\boldsymbol{c}\|_2^2\right] \le \mathbb{E}_{\mathcal{T}}\left[\left\|\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t}-\boldsymbol{c}\right\|_2^2\right] = \frac{1}{t}\mathbb{E}_{T_1}\left[\|\boldsymbol{x}_{T_1}-\boldsymbol{c}\|_2^2\right] = \frac{1}{t}\sum_{i\in[k]}\lambda_i\|\boldsymbol{x}_i-\boldsymbol{c}\|_2^2. \tag{6e}$$

Plugging the upper bound of (6e) into (6c), we obtain that

$$\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_1^{\mathcal{T}}}\sum_{i\in[k]}\lambda_i\|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2\right] \le \left(1+\frac{1}{t}\right)\sum_{i\in[k]}\lambda_i\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2.$$

This completes the proof. □

Remark. To obtain a $(1+\alpha)$-approximation in expectation, it suffices to choose $t=\lceil 1/\alpha\rceil$, so that $\frac{1}{t}\le\alpha$. Since $|S_1^{\mathcal{T}}|=O(n^t)$, Lemma 2 implies that solving the restricted problem over $S_1^{\mathcal{T}}$ has complexity $\tilde{O}(n^{3+\frac{1}{\alpha}}k^3d)$, which is polynomial in $n^{1/\alpha}$, k, and d.

3.2 Deterministic counterpart based on multi-subset enumeration

A deterministic counterpart to Algorithm 1 can be obtained by enumerating all possible sampling outcomes of the random multiset $\mathcal{T}$ and aggregating all resulting averages into a single reduced candidate support. Specifically, we define

$$S_1^t := \left\{\frac{1}{t}\sum_{i\in[t]}\boldsymbol{x}_{T_i} : \boldsymbol{x}_{T_i}\in\Xi_{T_i}, T_i\in[k], i\in[t]\right\}.$$

We then solve the restricted MOT dual (3) using S_1^t as the candidate support, and obtain an approximation barycenter $\mathbb{P}(S_1^t)$ based on Lemma 3. The full deterministic procedure is summarized in Algorithm 2.

Algorithm 2 Multi-subset enumeration algorithm for Wasserstein barycenter problem

1: **Input**: Probability measures $\{\mathbb{P}_i\}_{i\in[k]}$ with supports $\{\Xi_i\}_{i\in[k]}$, respectively; a constant $t \in [k]$.
2: $S_1^t := \left\{ \frac{1}{t} \sum_{i\in[t]} \boldsymbol{x}_{T_i} : \boldsymbol{x}_{T_i} \in \Xi_{T_i}, T_i \in [k], i \in [t] \right\}$.
3: Solve the restricted MOT dual (3) and obtain its corresponding primal optimal solution $\boldsymbol{\Pi}^*(S_1^t)$.
4: **Output**: Obtain an approximation barycenter $\mathbb{P}(S_1^t)$ based on Lemma 3.

Since S_1^t contains the support points generated by all possible realizations of the random construction $S_1^{\mathcal{T}}$, we have

$$\min_{\boldsymbol{s}\in S_1^t} \|\boldsymbol{s}-\boldsymbol{c}\|_2^2 \le \mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_1^{\mathcal{T}}} \|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right] \le \mathbb{E}_{\mathcal{T}}\left[\left\|\frac{\sum_{i\in\mathcal{T}} \boldsymbol{x}_i}{t} - \boldsymbol{c}\right\|_2^2\right],$$

for any reference point $\boldsymbol{c}$. Using this observation and following the same argument as in the proof of Theorem 2, we obtain the following deterministic guarantee.

Theorem 3. Let $t \in [k]$ be a fixed constant, and let S_1^t be the candidate support returned by Algorithm 2. Then

$$v(S_1^t) \le \left(1 + \frac{1}{t}\right) v^*,$$

and the support size satisfies $|S_1^t| = O((nk)^t)$.

Proof. According to the proof of Theorem 2, it is sufficient to show that

$$\min_{\boldsymbol{s}\in S_2^t} \sum_{i\in[k]} \lambda_i \|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2 \le \mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_1^{\mathcal{T}}} \sum_{i\in[k]} \lambda_i \|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2\right],$$

holds for any $\boldsymbol{x}_i \in \Xi_i$ and $i \in [k]$. This is indeed true since by definition, we have $S_1^{\hat{T}} \subseteq S_1^t$ for any realization $\hat{T}$ of random set $\mathcal{T}$. The approximation ratio then follows from Theorem 2.

Finally, note that there are $\binom{k+t-1}{t} n^t$ points in the set S_1^t by stars and bars method. Therefore, the size of S_1^t is $O(nk)^t$. □

Remark. When $t = 1$, Algorithm 2 enumerates all individual data points and thus reduces to the best-known 2-approximation method of Borgwardt (2022); our analysis recovers the same approximation guarantee for this special case. Since $|S_1^t| = O((nk)^t)$, achieving a $(1+\alpha)$ approximation requires choosing $t = \lceil 1/\alpha \rceil$, which leads to an overall complexity of $\tilde{O}\left(n^{3+\frac{1}{\alpha}} k^{3+\frac{1}{\alpha}} d\right)$.

4 Approximation algorithms for the Wasserstein barycenter problem with equal weights

In this section, we focus on the special case of (WBCenter) in which the input probability measures are equally weighted, i.e., $\lambda_i = \frac{1}{k}$ for each $i \in [k]$. In this setting, we show that the approximation ratio can be further improved by using a sampling procedure without repetition, leading to a tighter guarantee than that in the general-weight case.

4.1 Subset sampling algorithm without repetition

We uniformly sample a subset $\mathcal{T} \subseteq [k]$ of size t, without repetition. Based on this subset, we define the candidate support

$$S_2^{\mathcal{T}} := \left\{ \frac{1}{t} \sum_{i \in \mathcal{T}} \boldsymbol{x}_i : \boldsymbol{x}_i \in \Xi_i \right\}.$$

We then solve the restricted MOT dual (3) with $S_2^{\mathcal{T}}$ as the candidate support and recover an approximate barycenter $\mathbb{P}(S_2^{\mathcal{T}})$ using Lemma 3. A formal description of this procedure is given in Algorithm 3.

Algorithm 3 Subset sampling algorithm for Wasserstein barycenter problem

1: **Input**: Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$, respectively; a constant $t \in [k]$.
2: Uniformly sample a size-t subset $\mathcal{T} \subseteq [k]$.
3: $S_2^{\mathcal{T}} := \left\{ \frac{1}{t} \sum_{i \in \mathcal{T}} \boldsymbol{x}_i : \boldsymbol{x}_i \in \Xi_i \right\}$.
4: Solve the restricted MOT dual (3) and obtain its corresponding primal optimal solution $\boldsymbol{\Pi}^*(S_2^{\mathcal{T}})$
5: **Output**: Obtain an approximation barycenter $\mathbb{P}(S_2^{\mathcal{T}})$ based on Lemma 3.

The approximation guarantee is formally stated in the following theorem.

Theorem 4. Let $t \in [k]$ be a fixed constant, and let $S_2^{\mathcal{T}}$ be the candidate support returned by Algorithm 3. Then

$$\mathbb{E}_{\mathcal{T}} \left[v(S_2^{\mathcal{T}}) \right] \leq \left(1 + \frac{k-t}{t(k-1)} \right) v^*.$$

Proof. By Theorem 1, we only need to show that for any fixed realization $\boldsymbol{x}_i \in \Xi_i$, $i \in [k]$ and $\boldsymbol{c} := \frac{1}{k} \sum_{i \in [k]} \boldsymbol{x}_i$, we must have

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\boldsymbol{s} \in S_2^{\mathcal{T}}} \frac{1}{k} \sum_{i \in [k]} \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2 \right] \leq \left(1 + \frac{k-t}{t(k-1)} \right) \frac{1}{k} \sum_{i \in [k]} \|\boldsymbol{c} - \boldsymbol{x}_i\|_2^2. \tag{7a}$$

Note that $\boldsymbol{c} \in \arg\min_{\boldsymbol{s} \in \mathbb{R}^d} \frac{1}{k} \sum_{i \in [k]} \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2$. Therefore, to approximate the solution of $\min_{\boldsymbol{s} \in S_2^{\mathcal{T}}} \frac{1}{k} \sum_{i \in [k]} \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2$, we consider $\boldsymbol{s}_c \in \arg\min_{\boldsymbol{s} \in S_2^{\mathcal{T}}} \|\boldsymbol{s} - \boldsymbol{c}\|_2^2$, which is the closest point in $S_2^{\mathcal{T}}$ to $\boldsymbol{c}$. Similar to the proof of Theorem 2, we have

$$\frac{1}{k} \sum_{i \in [k]} \|\boldsymbol{s_c} - \boldsymbol{x}_i\|_2^2 = \frac{1}{k} \sum_{i \in [k]} (\|\boldsymbol{s_c} - \boldsymbol{c}\|_2^2 + \|\boldsymbol{c} - \boldsymbol{x}_i\|_2^2).$$

Thus, the left-hand side can be upper bounded by

$$\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}}\frac{1}{k}\sum_{i\in[k]}\|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2\right]\le\frac{1}{k}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2+\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right]. \tag{7b}$$

It remains to bound $\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right]$. For notational convenience, we let $\hat{\boldsymbol{\mu}}_{\mathcal{T}}=\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t}$. Since $\hat{\boldsymbol{\mu}}_{\mathcal{T}}\in S_2^{\mathcal{T}}$, we have

$$\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right]\le\mathbb{E}_{\mathcal{T}}\left[\|\hat{\boldsymbol{\mu}}_{\mathcal{T}}-\boldsymbol{c}\|_2^2\right]. \tag{7c}$$

Note that

$$\begin{aligned}\mathbb{E}_{\mathcal{T}}\left[\hat{\boldsymbol{\mu}}_{\mathcal{T}}\right]&=\sum_{\binom{\hat{T}\in[k]}{t}}\mathrm{Prob}[\mathcal{T}=\hat{T}]\frac{\sum_{i\in\hat{T}}\boldsymbol{x}_i}{t}=\frac{1}{t\binom{k}{t}}\sum_{\binom{\hat{T}\in[k]}{t}}\sum_{i\in\hat{T}}\boldsymbol{x}_i\\&=\frac{1}{t\binom{k}{t}}\sum_{i\in[k]}\boldsymbol{x}_i\binom{k-1}{t-1}=\frac{1}{k}\sum_{i\in[k]}\boldsymbol{x}_i=\boldsymbol{c}.\end{aligned}$$

In addition, we also have

$$\mathrm{Cov}\left[\hat{\boldsymbol{\mu}}_{\mathcal{T}}\right]=\frac{1}{t^2}\sum_{i,j\in\mathcal{T}}(\boldsymbol{x}_i-\boldsymbol{c})(\boldsymbol{x}_j-\boldsymbol{c})^\top\mathrm{Prob}\left[i,j\in\mathcal{T}\right],$$

where

$$\mathrm{Prob}[i,j\in\mathcal{T}]=\begin{cases}\frac{t}{k}, & \text{if } i=j,\\ \frac{\binom{k-2}{t-2}}{\binom{k}{t}}=\frac{t(t-1)}{k(k-1)}, & \text{if } i\ne j,\end{cases}$$

by direct computation.

Therefore, we further have

$$\mathrm{Cov}\left[\hat{\boldsymbol{\mu}}_{\mathcal{T}}\right]=\frac{1}{t^2}\sum_{i\in[k]}(\boldsymbol{x}_i-\boldsymbol{c})(\boldsymbol{x}_i-\boldsymbol{c})^\top\frac{t}{k}+\frac{1}{t^2}\sum_{i\ne j\in[k]}(\boldsymbol{x}_i-\boldsymbol{c})(\boldsymbol{x}_j-\boldsymbol{c})^\top\frac{t(t-1)}{k(k-1)},$$

which implies that

$$\begin{aligned}\mathrm{Var}\left[\hat{\boldsymbol{\mu}}_{\mathcal{T}}\right]=&\mathbb{E}_{\mathcal{T}}\left[\|\hat{\boldsymbol{\mu}}_{\mathcal{T}}-\boldsymbol{c}\|_2^2\right]\\=&\frac{1}{tk}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2+\frac{t-1}{tk(k-1)}\sum_{i\ne j\in[k]}(\boldsymbol{x}_i-\boldsymbol{c})^\top(\boldsymbol{x}_j-\boldsymbol{c})\\=&\frac{1}{tk}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2-\frac{t-1}{tk(k-1)}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2\end{aligned}$$

$$=\frac{k-t}{t(k-1)}\frac{1}{k}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2 \tag{7d}$$

where the third equality is due to $\sum_{i\in[k]}(\boldsymbol{x}_i-\boldsymbol{c})=0$.

Plugging (7d) to (7c), we have

$$\mathbb{E}_{\mathcal{T}}\left[\frac{1}{k}\sum_{i\in[k]}\|\boldsymbol{s}-\boldsymbol{x}_i\|_2^2\right]\leq\frac{k-t}{tk(k-1)}\sum_{j\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_j\|_2^2+\frac{1}{k}\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2=\left(1+\frac{k-t}{t(k-1)}\right)\sum_{i\in[k]}\|\boldsymbol{c}-\boldsymbol{x}_i\|_2^2.$$

□

Remark. To obtain a $(1+\alpha)$-approximation guarantee, it suffices to choose t such that $\frac{k-t}{t(k-1)}\leq\alpha$; equivalently, $t\geq\frac{k}{\alpha k-\alpha+1}$. Since $|S_2^{\mathcal{T}}|=O(n^t)$, solving the restricted problem over $S_2^{\mathcal{T}}$ requires time $\tilde{O}\left(n^{3+\frac{k}{\alpha k-\alpha+1}}k^3d\right)$ to achieve a $(1+\alpha)$-approximation.

4.2 Deterministic counterpart based on subset enumeration

Similar to Algorithm 2, we obtain a deterministic counterpart to Algorithm 3 by enumerating all possible choices of the subset $\mathcal{T}\subseteq[k]$ of size t (without repetition), and aggregating all corresponding averaged points into a single reduced candidate support. Specifically, we define

$$S_2^t:=\left\{\frac{1}{t}\sum_{i\in\mathcal{T}}\boldsymbol{x}_i:\boldsymbol{x}_i\in\Xi_i,\mathcal{T}\subseteq[k],|\mathcal{T}|=t\right\}$$

We then solve the restricted MOT dual (3) using S_2^t as the candidate support and recover an approximate barycenter $\mathbb{P}(S_2^t)$ via Lemma 3. The resulting deterministic procedure is summarized in Algorithm 4.

Algorithm 4 Subset enumeration algorithm for Wasserstein barycenter problem

1: **Input**: Probability measures $\{\mathbb{P}_i\}_{i\in[k]}$ with supports $\{\Xi_i\}_{i\in[k]}$, respectively; a constant $t\in[k]$.
2: $S_2^t:=\left\{\frac{1}{t}\sum_{i\in\mathcal{T}}\boldsymbol{x}_i:\boldsymbol{x}_i\in\Xi_i,\mathcal{T}\subseteq[k],|\mathcal{T}|=t\right\}$.
3: Solve the restricted MOT dual (3) and obtain its corresponding primal optimal solution $\boldsymbol{\Pi}^*(S_2^t)$.
4: **Output**: Obtain an approximation barycenter $\mathbb{P}(S_2^t)$ based on Lemma 3.

Since S_2^t contains the support points generated by all possible realizations of the random construction $S_2^{\mathcal{T}}$, we have, for any reference point $\boldsymbol{c}$,

$$\min_{\boldsymbol{s}\in S_2^t}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\leq\mathbb{E}_{\mathcal{T}}\left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2\right]\leq\mathbb{E}_{\mathcal{T}}\left[\left\|\frac{\sum_{i\in\mathcal{T}}\boldsymbol{x}_i}{t}-\boldsymbol{c}\right\|_2^2\right].$$

Using this observation and following the same argument as in the proof of Theorem 4, we obtain the corresponding deterministic approximation guarantee.

Theorem 5. Let $t\in[k]$ be a fixed constant, and let S_2^t be the candidate support returned by

Algorithm 3. Then

$$\min_{\mathcal{T}:|\mathcal{T}|=t} v(S_2^t) \leq \left(1 + \frac{k-t}{t(k-1)}\right) v^*,$$

and the support size satisfies $|S_2^t| = O((nk)^t)$

Proof. According to the proof of Theorem 4, it is sufficient to show that

$$\min_{\boldsymbol{s}\in S_2^t} \sum_{i\in[k]} \lambda_i \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2 \leq \mathbb{E}_{\mathcal{T}} \left[\min_{\boldsymbol{s}\in S_2^{\mathcal{T}}} \sum_{i\in[k]} \lambda_i \|\boldsymbol{s} - \boldsymbol{x}_i\|_2^2 \right],$$

holds for any $\boldsymbol{x}_i \in \Xi_i$ and any $i \in [k]$. This is indeed true since by definition, we have $S_2^{\hat{T}} \subseteq S_2^t$ for any realization $\hat{T}$ of random set $\mathcal{T}$. The approximation ratio then follows from Theorem 4.

Finally, note that there are $\binom{k}{t} n^t$ points in the set S_2^t. Therefore, the size of S_2^t is $O((nk)^t)$. □

Remark. Note that

$$1 + \frac{k-t}{t(k-1)} \leq 1 + \frac{1}{t},$$

which implies that, for $t > 1$, the approximation ratio achieved by Algorithm 4 (which applies to the equal-weight Wasserstein barycenter problem) is strictly better than that of Algorithm 2 (which applies to the general-weight Wasserstein barycenter problem).

5 Numerical experiments

In this section, we conduct numerical experiments to evaluate the computational efficiency and solution quality of the proposed algorithms. Our main goals are:

(i) To empirically validate the theoretical $(1 + \alpha)$-approximation guarantee and the tunable accuracy-complexity trade-off;

(ii) To demonstrate that our algorithms scale to large instances that are computationally intractable for exact extensive LP formulations; and

(iii) To compare our results with the state-of-the-art 2-approximation algorithm (Borgwardt 2022), which is equivalent to Algorithm 4 with $t = 1$.

All of our algorithms were implemented in Python using Gurobi 11.0.3 as the LP solver and executed on a MacBook Pro with an Apple M4 Pro Chip and 48 GB RAM.

5.1 Synthetic nested ellipse dataset

The synthetic dataset used in this subsection is a standard benchmark from the literature (Cuturi & Doucet 2014, Janati et al. 2020*b*, Altschuler & Boix-Adsera 2021), for which the optimal Wasserstein barycenter is known (Altschuler & Boix-Adsera 2021). It consists of ten probability measures

$\{\mathbb{P}_i\}_{i\in[10]}$, each supported on a 60×60 grid and corresponding to a nested ellipse (see Figure 1). Each measure has between 139 and 192 support points, all with equal mass. We emphasize that the exact barycenter may not be , in general, supported on the grid. Solving the full MOT formulation (1) over S^* would require on the order of $192^{10} \approx 10^{22}$ basic arithmetic operations, which is computationally intractable.

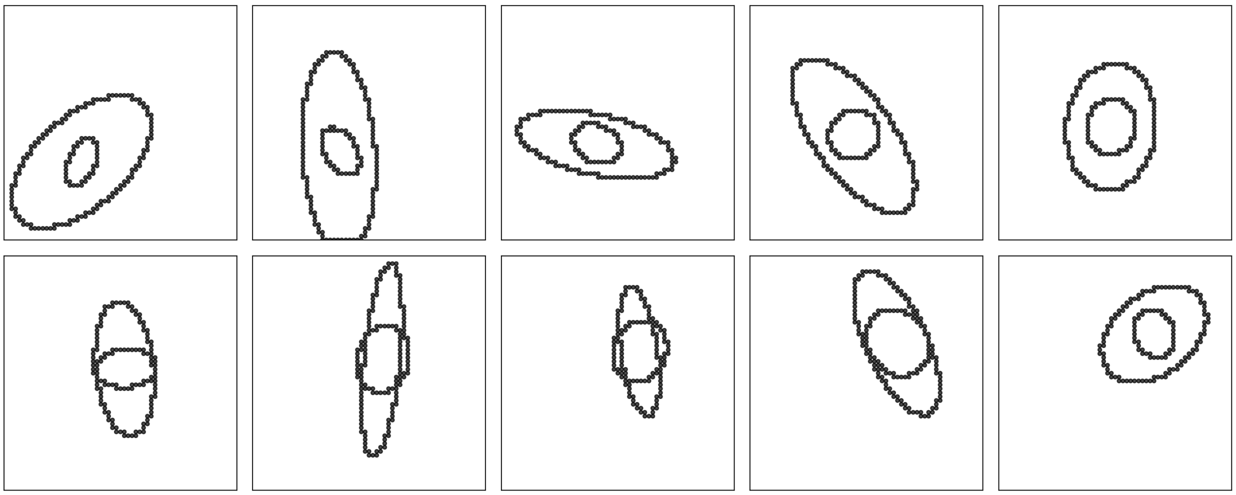


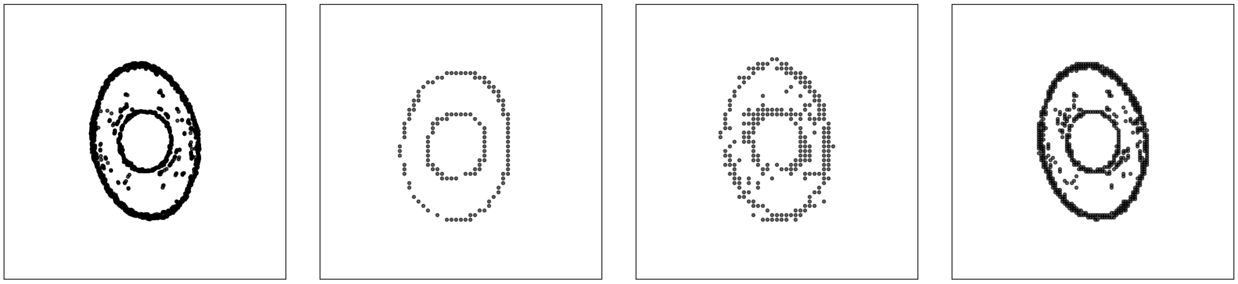
Figure 1: Ten images from the nested ellipses dataset.

To ensure higher accuracy, we adopt the subset enumeration method (i.e., Algorithm 4) rather than the subset sampling method (i.e., Algorithm 3), even though the latter is more efficient and enjoys the same theoretical approximation ratio. The key reason is that sampling operates over the set of measures, so regardless of how many samples are drawn, the number of distinct candidate supports is finite. As illustrated in Figure 2, even the best result among the ten possible samples for $t = 1$ is still substantially worse than the result obtained by deterministic enumeration. Consequently, in all subsequent experiments, we apply our deterministic subset enumeration algorithm with $t = 1$ and $t = 2$.

Additionally, as illustrated in Figure 2, all methods produce solutions that are substantially better than their theoretical worst-case guarantees. We also observe that even small differences in objective values between $t = 1$ and $t = 2$ can correspond to visibly inferior barycenter quality. While the $t = 1$ setting offers faster computation, the $t = 2$ setting consistently yields barycenters of much higher visual fidelity.

5.2 MNIST Dataset

To assess scalability with respect to the number of measures k, we conduct experiments on the MNIST dataset (LeCun et al. 2002). For each of three digit classes, we randomly select $k = 50$ empirical measures, each supported on a 28×28 grid ($d = 2$) with (potentially) different mass distributions. The total number of support points per measure ranges from 165 to 263. Solving the full MOT formulation in this setting would require on the order of $263^{50} \approx 10^{120}$ basic arithmetic operations, which is computationally intractable.



(a) Exact ellipse (Altschuler & Boix-Adsera 2021), objective=0.02666, time=8865.03s.

(b) 10 runs of Algorithm 3 with $t = 1$, objective=0.02709, time=17.46s.

(c) Algorithm 4 with $t = 1$. objective=0.02673, time=21.08s.

(d) Algorithm 4 with $t = 2$. objective=0.02667, time=212.44s.

Figure 2: Barycenters produced by different algorithms for nested ellipse.

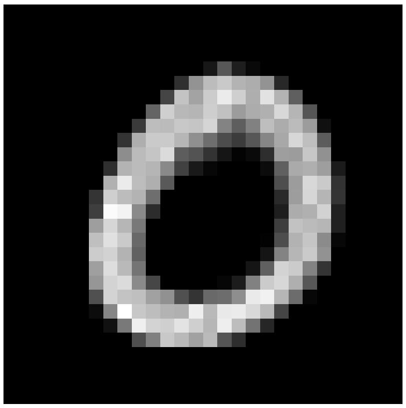
We evaluate the performance of Algorithm 2 for $t = 1$ and $t = 2$, and report both the objective values and the running times for all digit classes. All experiments were completed in a reasonable time (i.e., less than an hour). Even $t = 1$ produces reasonable barycenters, and $t = 2$ consistently improves the objective values. The visualizations in Figure 3 further show that the resulting barycenters successfully capture the characteristic shapes of the digits.

5.3 Sign language dataset

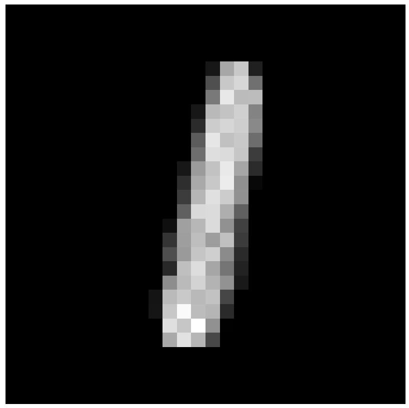
The final experiment evaluates the performance of our algorithms on measures with large supports. We represent each RGB image of size $64 \times 64 \times 3$ as a 5-dimensional empirical measure: the first two coordinates encode the spatial location (x, y), and the remaining three coordinates encode the color channels. For each of three distinct hand gestures, we randomly select three images as input measures.

Since the resulting LP is extremely large, we combine our subset-enumeration method (Algorithm 4) with column generation (Borgwardt & Patterson 2022) and early termination. To interpolate between the computational profiles of $t = 1$ and $t = 2$, we construct a hybrid candidate support as follows: we begin with the active atoms of the $t = 1$ barycenter, then augment this set by adding, for each input measure, the five nearest support points to each active atom, and finally include all pairwise averages of these points. This hybrid candidate expansion balances between accuracy and running time, which is referred as the "Hybrid Algorithm 4 with $t = 2$."

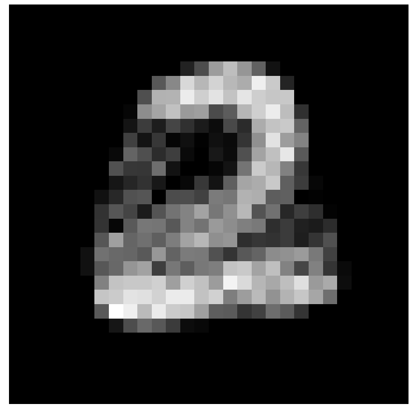
Representative results are shown in Figure 4–6. Across all sign gestures, the barycenters computed with $t = 2$ achieve approximately 3–6% lower objective values than those computed with $t = 1$, while remaining visually better gestures. For example, Figures 4a and 4b show the barycenters computed by Algorithm 4 with $t = 1$, terminated after 1 hour and 2 hours, respectively. Figures 4c and 4d show the barycenters computed with $t = 2$, initialized from the solution in Figure 4a, and terminated after 1 hour and 2 hours, respectively. We see that the barycenters with $t = 2$ are denser and more clear than those with $t = 1$. The same behavior is observed for the



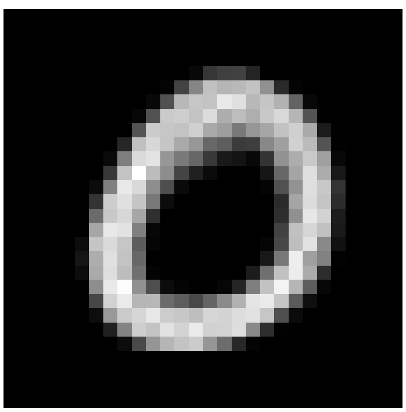
(a) Algorithm 2 with $t = 1$, objective=0.0132, time=556s.



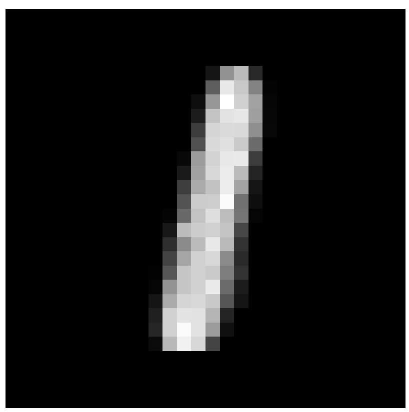
(b) Algorithm 2 with $t = 1$. objective=0.0174, time=56s.



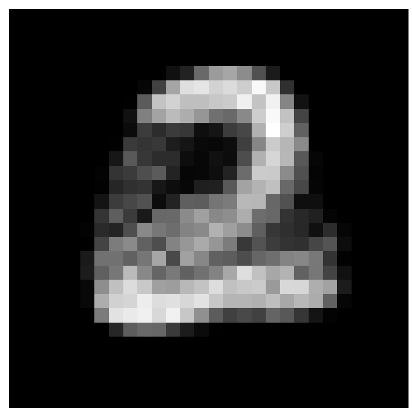
(c) Algorithm 2 with $t = 1$. objective=0.0279, time=253s.



(d) Algorithm 2 with $t = 2$, objective=0.0127, time=9213s.



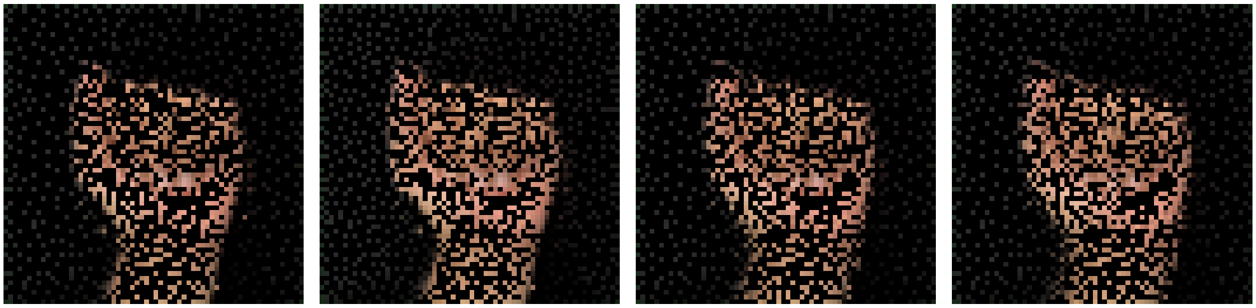
(e) Algorithm 2 with $t = 2$. objective=0.0169, time=1148s.



(f) Algorithm 2 with $t = 2$. objective=0.0274, time=2716s.

Figure 3: Barycenters produced by different algorithms for MNIST.

remaining gestures, as illustrated in Figures 5–6.



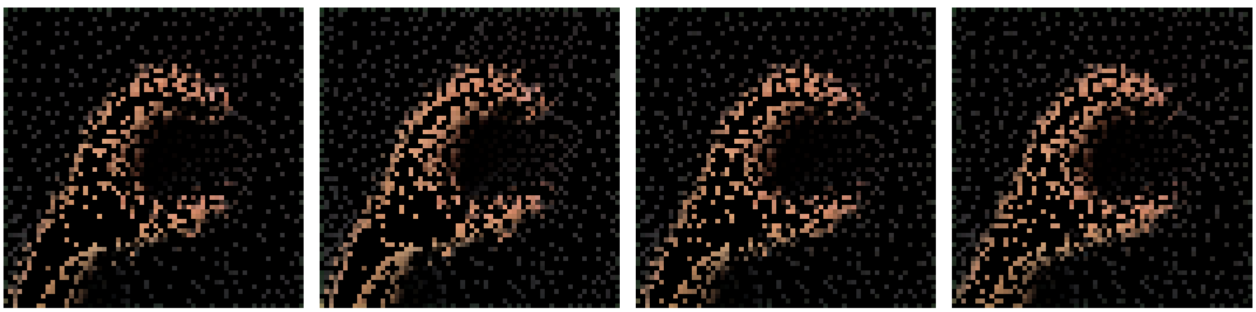
(a) Algorithm 4 with $t = 1$, objective=0.00368, time=1h.

(b) Algorithm 4 with $t = 1$, objective=0.00358, time=2h.

(c) Hybrid Algorithm 4 with $t = 2$, objective=0.00355, time=1h.

(d) Hybrid Algorithm 4 with $t = 2$, objective=0.00348, time=2h.

Figure 4: Barycenters produced by different algorithms for gesture 1.



(a) Algorithm 4 with $t = 1$, objective=0.0115, time=1h.

(b) Algorithm 4 with $t = 1$, objective=0.0115, time=2h.

(c) Hybrid Algorithm 4 with $t = 2$, objective=0.0111, time=1h.

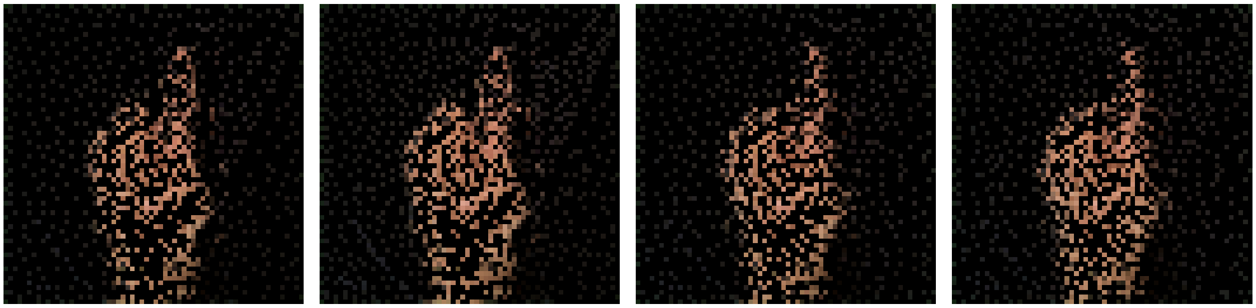
(d) Hybrid Algorithm 4 with $t = 2$, objective=0.0110, time=2h.

Figure 5: Barycenters produced by different algorithms for gesture 2.

Finally, we evaluate the quality of the computed barycenters through a classification task. Each class is represented by the Wasserstein barycenter of its corresponding training samples. For each test input (25 testing images for each gesture), we compute its Wasserstein distance to each class barycenter, and assign it to the class with the smallest distance. Using barycenters computed with $t = 1$ after 2 hours yields a classification accuracy of 96%, while using $t = 2$ after 1 hour improves the accuracy to 100%. This demonstrates that the proposed approximation algorithm for $t = 2$ is also more efficient in terms of testing accuracy. The details are described in Table 1.

6 Conclusion and future work

In this paper, we present efficient approximation algorithms for computing discrete Wasserstein barycenters. Our approach generalizes the best-known 2-approximation algorithm and yields a polynomial-time approximation scheme (PTAS) that offers an explicit tradeoff between solution quality and computational complexity.



(a) Algorithm 4 with $t = 1$, objective=0.00957, time=1h.

(b) Algorithm 4 with $t = 1$, objective=0.00949, time=2h.

(c) Hybrid Algorithm 4 with $t = 2$, objective=0.00917, time=1h.

(d) Hybrid Algorithm 4 with $t = 2$, objective=0.00898, time=2h.

Figure 6: Barycenters produced by different algorithms for gesture 3.

Table 1: Gesture classification task. Each entry means the number of images that falls in the case. For example, Algorithm 4 with $t = 1$ classifies gesture 2 correct for 22 out of 25 images, and wrong for 3 out of 25 images.

Methods	gesture 1	gesture 2	gesture 3
Algorithm 4 with $t = 1$ correct	25	22	25
Algorithm 4 with $t = 1$ wrong	0	3	0
Hybrid Algorithm 4 with $t = 2$ correct	25	25	25
Hybrid Algorithm 4 with $t = 2$ wrong	0	0	0

Our numerical results show that the proposed methods achieve near-optimal objective values in challenging datasets, including cases with a large number of measures and cases with very large supports. In particular, our methods consistently outperform existing 2-approximation approaches in terms of accuracy, while maintaining reasonable running times.

Several directions remain open. From a theoretical standpoint, extending our guarantees to general type-p barycenters for $p \neq 2$ is an important next step. On the algorithmic side, further optimization of the implementation may lead to additional computational improvement. Finally, the candidate-support reduction principle developed in this work may be applicable beyond Wasserstein barycenters, and investigating such extensions is a promising avenue for future research.

Aknowlegdements

This research has been supported in part by National Science Foundation grants 2246414 and the Office of Naval Research grant N00014-24-1-2066. The second author would like to thank Ph.D. students Zedong Wang and Shuai Li for their discussions on the problem.

References

Agueh, M. & Carlier, G. (2011), 'Barycenters in the wasserstein space', *SIAM Journal on Mathematical Analysis* **43**(2), 904–924.

Altschuler, J. M. & Boix-Adsera, E. (2021), 'Wasserstein barycenters can be computed in polynomial time in fixed dimension', *Journal of Machine Learning Research* **22**(44), 1–19.

Altschuler, J. M. & Boix-Adsera, E. (2022), 'Wasserstein barycenters are np-hard to compute', *SIAM Journal on Mathematics of Data Science* **4**(1), 179–203.

Anderes, E., Borgwardt, S. & Miller, J. (2016), 'Discrete wasserstein barycenters: Optimal transport for discrete data', *Mathematical Methods of Operations Research* **84**(2), 389–409.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L. & Peyré, G. (2015), 'Iterative bregman projections for regularized transportation problems', *SIAM Journal on Scientific Computing* **37**(2), A1111–A1138.

Bigot, J., Klein, T. et al. (2012), 'Consistent estimation of a population barycenter in the wasserstein space', *ArXiv e-prints*.

Bonneel, N., Peyré, G. & Cuturi, M. (2016), 'Wasserstein barycentric coordinates: histogram regression using optimal transport.', *ACM Trans. Graph.* **35**(4), 71–1.

Bonneel, N., Rabin, J., Peyré, G. & Pfister, H. (2015), 'Sliced and radon wasserstein barycenters of measures', *Journal of Mathematical Imaging and Vision* **51**(1), 22–45.

Borgwardt, S. (2022), 'An lp-based, strongly-polynomial 2-approximation algorithm for sparse wasserstein barycenters', *Operational Research* **22**(2), 1511–1551.

Borgwardt, S. & Patterson, S. (2020), 'Improved linear programs for discrete barycenters', *Informs Journal on Optimization* **2**(1), 14–33.

Borgwardt, S. & Patterson, S. (2021), 'On the computational complexity of finding a sparse wasserstein barycenter', *Journal of Combinatorial Optimization* **41**(3), 736–761.

Borgwardt, S. & Patterson, S. (2022), 'A column generation approach to the discrete barycenter problem', *Discrete Optimization* **43**, 100674.

Carlier, G. & Ekeland, I. (2010), 'Matching for teams', *Economic Theory* **42**(2), 397–418.

Cheng, K., Aeron, S., Hughes, M. C. & Miller, E. L. (2021), 'Dynamical wasserstein barycenters for time-series modeling', *Advances in Neural Information Processing Systems* **34**, 27991–28003.

Claici, S., Chien, E. & Solomon, J. (2018), Stochastic wasserstein barycenters, *in* 'International Conference on Machine Learning', PMLR, pp. 999–1008.

Cuturi, M. & Doucet, A. (2014), Fast computation of wasserstein barycenters, *in* 'International conference on machine learning', PMLR, pp. 685–693.

Ge, D., Wang, H., Xiong, Z. & Ye, Y. (2019), 'Interior-point methods strike back: Solving the wasserstein barycenter problem', *Advances in neural information processing systems*.

Grötschel, M., Lovász, L. & Schrijver, A. (2012), *Geometric algorithms and combinatorial optimization*, Vol. 2, Springer Science & Business Media.

Huang, M., Ma, S. & Lai, L. (2021), Projection robust wasserstein barycenters, *in* 'International Conference on Machine Learning', PMLR, pp. 4456–4465.

Janati, H., Bazeille, T., Thirion, B., Cuturi, M. & Gramfort, A. (2020*a*), 'Multi-subject MEG/EEG source imaging with sparse multi-task regression', *NeuroImage* **220**, 116847.

Janati, H., Cuturi, M. & Gramfort, A. (2020*b*), Debiased sinkhorn barycenters, *in* 'International Conference on Machine Learning', PMLR, pp. 4692–4701.

Kroshnin, A., Tupitsa, N., Dvinskikh, D., Dvurechensky, P., Gasnikov, A. & Uribe, C. (2019), On the complexity of approximating wasserstein barycenters, *in* 'International conference on machine learning', PMLR, pp. 3530–3540.

Lau, T. T.-K. & Liu, H. (2022), 'Wasserstein distributionally robust optimization with wasserstein barycenters', *arXiv preprint arXiv:2203.12136*.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (2002), 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE* **86**(11), 2278–2324.

Lin, T., Ho, N., Chen, X., Cuturi, M. & Jordan, M. (2020), 'Fixed-support wasserstein barycenters: Computational hardness and fast algorithm', *Advances in neural information processing systems* **33**, 5368–5380.

Lin, Z. & Ruszczyński, A. (2025), 'Federated calculation of the free-support transportation barycenter by single-loop dual decomposition', *arXiv preprint arXiv:2507.19627*.

Lindheim, J. v. (2023), 'Simple approximative algorithms for free-support wasserstein barycenters', *Computational Optimization and Applications* **85**(1), 213–246.

Luise, G., Salzo, S., Pontil, M. & Ciliberto, C. (2019), 'Sinkhorn barycenters with free support via frank-wolfe algorithm', *Advances in neural information processing systems*.

Schmitz, M. A., Heitz, M., Bonneel, N., Ngole, F., Coeurjolly, D., Cuturi, M., Peyré, G. & Starck, J.-L. (2018), 'Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning', *SIAM Journal on Imaging Sciences* **11**(1), 643–678.

Simon, D. & Aberdam, A. (2020), Barycenters of natural images constrained wasserstein barycenters for image morphing, *in* 'Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition', pp. 7910–7919.

Uribe, C. A., Dvinskikh, D., Dvurechensky, P., Gasnikov, A. & Nedić, A. (2018), Distributed computation of wasserstein barycenters over networks, *in* '2018 IEEE Conference on Decision and Control (CDC)', IEEE, pp. 6544–6549.

Yan, Y., Duffner, S., Phutane, P., Berthelier, A., Blanc, C., Garcia, C. & Chateau, T. (2021), '2d wasserstein loss for robust facial landmark detection', *Pattern Recognition* **116**, 107945.

Yang, L., Li, J., Sun, D. & Toh, K.-C. (2021), 'A fast globally linearly convergent algorithm for the computation of wasserstein barycenters', *Journal of Machine Learning Research* **22**(21), 1–37.

Ye, J., Wu, P., Wang, J. Z. & Li, J. (2017), 'Fast discrete distribution clustering using wasserstein barycenter with sparse support', *IEEE Transactions on Signal Processing* **65**(9), 2317–2332.