

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

Department of Data Science, City University of Hong Kong, Hong Kong, wj.xie@cityu.edu.hk

Abstract

Given a collection of probability measures, the Wasserstein barycenter problem seeks a probability measure that minimizes the weighted sum of their squared type-2 Wasserstein distances. We study the discrete setting, in which every input measure has finite support; this setting arises frequently in machine learning and operations research. Since computing a discrete Wasserstein barycenter is NP-hard, we develop approximation algorithms with provable performance guarantees. The best previously known polynomial-time guarantee is a factor of 2. We improve this bound by proposing a polynomial-time approximation scheme (PTAS) that generalizes the existing 2-approximation: for any prescribed $\alpha \in (0, 1]$, the proposed algorithms return a $(1+\alpha)$ -approximate barycenter in time polynomial in $(nk)^{1/\alpha}$ and d . We provide both randomized and deterministic constructions and derive a sharper guarantee when the input measures are equally weighted. Numerical experiments on synthetic and real datasets show that the algorithms are computationally practical and produce near-optimal barycenters.

1 Introduction

A Wasserstein barycenter is a Fréchet mean in Wasserstein space: it minimizes a weighted average of optimal-transport costs over a collection of probability measures. For $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_2(\mathbb{R}^d)$, the type-2 Wasserstein distance is

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

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

Let $[k] := \{1, \dots, k\}$, and let $\mathbb{P}_1, \dots, \mathbb{P}_k$ be discrete probability measures. For each $i \in [k]$, the measure $\mathbb{P}_i$ is supported on a finite set $\Xi_i \subseteq \mathbb{R}^d$ with $|\Xi_i| \leq n$. We denote its j -th support point by $\hat{\mathbf{x}}_{ij}$ and the corresponding mass by $\hat{p}_{ij}$, for $j \in [|\Xi_i|]$. Given weights $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$ in the simplex $\Delta_k := \{\boldsymbol{\lambda} \in \mathbb{R}_+^k : \sum_{i \in [k]} \lambda_i = 1\}$, the discrete Wasserstein barycenter problem is

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

where $\mathcal{P}(\mathbb{R}^d)$ denotes the set of Borel probability measures on $\mathbb{R}^d$ with finite second moment. Discrete Wasserstein barycenters have applications in 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, Bazeille, Thirion, Cuturi &

Gramfort 2020), 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 Wasserstein barycenter was introduced by Agueh & Carlier (2011), who also connected it to the multi-marginal optimal transport (MOT) problem. Several extensions and variants have since been studied; see, for example, Bigot et al. (2012), Huang et al. (2021), Uribe et al. (2018). For finitely supported measures, Anderes et al. (2016) proved fundamental structural properties—including discreteness, sparsity, and a non-mass-splitting property—and derived an LP formulation through MOT. The resulting formulation, however, has $O(n^k)$ variables and constraints and is therefore exponential in the number of marginals k . Hardness persists under substantial restrictions: Borgwardt & Patterson (2021) proved that the sparse barycenter problem is NP-hard even for $d = 2$ and $k = 3$, while Altschuler & Boix-Adsera (2022) ruled out a polynomial-time randomized algorithm that achieves prescribed additive accuracy, even under uniform weights, unless $\text{NP} \subseteq \text{BPP}$. The latter result concerns additive approximation and thus does not preclude the multiplicative guarantees studied here. Existing multiplicative guarantees are limited to a factor of 2: Borgwardt (2022) used the union of the input supports as the candidate barycenter support, and Lindheim (2023) constructed a barycenter from a sequence of two-marginal transport plans. We improve this factor-2 bound by developing a polynomial-time approximation scheme (PTAS): for every $\alpha \in (0, 1]$, our method computes a $(1 + \alpha)$ -approximate barycenter in time polynomial in $(nk)^{1/\alpha}$ and d .

Exact algorithms are available under additional structural assumptions or through more elaborate LP techniques. In fixed dimension d , Altschuler & Boix-Adsera (2021) showed that an exact type- p Wasserstein barycenter for $p \in \{1, 2\}$ can be computed in time polynomial in n , k , and $\log U$, where $\log U$ bounds the input bit length. The same framework computes a barycenter with additive error $\epsilon > 0$ in time polynomial in n , k , and $\log(1/\epsilon)$ by solving an exponential-size LP through a separation oracle based on power diagrams. Outside the fixed-dimensional setting, Borgwardt & Patterson (2020) strengthened the standard LP formulation by exploiting the non-mass-splitting structure, aggregating transport variables by support tuples, and incorporating preprocessing rules and support-size bounds. Building on these ideas, Borgwardt & Patterson (2022) developed an exact column-generation method whose pricing problem searches for barycenter atoms with negative reduced cost and is accelerated through bounding, heuristics, stabilization, and warm starts. These approaches can be effective in practice, but they do not yield polynomial-time algorithms for all general input parameters and can become difficult to scale when n , k , and d are large.

Related literature studied the fixed-support variant, in which the barycenter support is specified *a priori* and the remaining problem is an LP. Since this LP can still be large, Cuturi & Doucet (2014) introduced entropic regularization and subgradient-based schemes, Benamou et al. (2015) developed a parallelizable iterative Bregman-projection (Sinkhorn) method, and Janati, Cuturi & Gramfort (2020) proposed a debiased variant that removes the regularization bias while retaining comparable convergence rates. Other approaches include an interior-point method (Ge et al. 2019), a deterministic Bregman-projection algorithm with improved complexity guarantees (Lin et al. 2020), and a

symmetric Gauss–Seidel ADMM derived from the dual formulation (Yang et al. 2021). Although some of these methods can be incorporated into schemes for updating non-prespecified support locations, their analysis does not provide the globally guaranteed candidate-support reduction developed in this paper.

Several methods address free-support or regularized barycenter problems directly. Clatici et al. (2018) proposed a stochastic scheme that updates the barycenter support using sampled subgradients, and Luise et al. (2019) developed a Frank–Wolfe method for barycenters defined through the Sinkhorn divergence. Lin & Ruszczýński (2025) formulated the free-support problem as an integer program and designed a federated dual-subgradient algorithm, while Kroshnin et al. (2019) analyzed iterative Bregman projections and accelerated gradient methods for the entropically regularized objective. These methods provide useful computational tools, but they do not show the polynomial-time multiplicative approximation guarantees developed in this paper.

1.2 Summary of contributions

We develop a unified candidate-support reduction framework for the discrete Wasserstein barycenter problem (WBCenter). The framework converts the design of approximation algorithms into the construction of a small set that approximates every tuplewise barycenter location. Its main results are as follows.

- (i) We prove the first PTAS for the discrete type-2 Wasserstein barycenter problem. For every $\alpha \in (0, 1]$, the proposed family of algorithms attains a $(1 + \alpha)$ multiplicative guarantee in time polynomial in $(nk)^{1/\alpha}$ and d .
- (ii) We provide both randomized sampling and deterministic enumeration variants. Under equal weights, sampling without replacement yields the sharper ratio $1 + (k - t)/(t(k - 1))$, compared with $1 + 1/t$ for general weights.
- (iii) We evaluate the algorithms on synthetic and real datasets that include many input measures and large marginal supports. The results show that the reduced-support models are computationally practical and produce high-quality barycenters in regimes where the full MOT formulation is intractable.
- (iv) We extend the support-reduction principle in two directions. For sparse type-2 barycenters, the same candidate set yields a $(1 + 1/t)$ value guarantee when the restricted mixed-integer model is solved to optimality. For type-1 barycenters, we derive exact and approximate candidate-support constructions based on coordinatewise weighted medians.

The remainder of the paper is organized as follows. Section 2 develops the candidate-support reduction framework. Section 3 presents randomized and deterministic algorithms for general weights, and Section 4 sharpens the guarantee under equal weights. Section 5 reports the numerical results. Section 6 studies sparse type-2 and type-1 extensions, and Section 7 concludes.

Notation. We write $[k] := \{1, \dots, k\}$ and use $\|\cdot\|_2$ for the Euclidean norm. The symbols $\delta_{(\cdot)}$ and $\mathbb{I}_{(\cdot)}$ denote a Dirac measure and an indicator, respectively. We use $(\mathbb{R}_+^n)^{\otimes k}$ for the k -fold

tensor product of $\mathbb{R}_+^n$, and $\mathbb{P}(S)$ for a probability measure supported on S . Finally, $O(\cdot)$ denotes standard asymptotic complexity, whereas $\tilde{O}(\cdot)$ suppresses logarithmic and polylogarithmic factors in the input size and accuracy.

2 Preliminary results and approximation scheme

This section develops the candidate-support reduction framework that underlies our algorithms and generalizes the construction of Borgwardt (2022). We first recall the equivalence between the discrete barycenter problem and multi-marginal optimal transport (MOT), which yields an exact but exponentially large candidate support S^* . We then restrict the tuplewise barycenter locations to a smaller set S . The resulting dual problem admits a separation oracle with running time polynomial in n , k , d , and $|S|$, and an optimal primal solution induces an explicit barycenter supported on S . Finally, we show that a tuplewise approximation guarantee for S implies the same multiplicative guarantee for the barycenter objective.

2.1 Optimal support

A key ingredient is the standard equivalence between the Wasserstein barycenter problem and an MOT problem; see Proposition 2.1 of Altschuler & Boix-Adsera (2022) and Section 6 of Carlier & Ekeland (2010). Specifically, (WBCenter) is equivalent to

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

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

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

Each entry $\Pi_{j_1, \dots, j_k}$ is a joint mass assigned to the tuple $(\hat{\mathbf{x}}_{1,j_1}, \dots, \hat{\mathbf{x}}_{k,j_k})$, and $C_{\vec{j}}$ is the minimum weighted squared distance from that tuple to a common barycenter location $\mathbf{w}$. The MOT representation immediately yields a finite exact candidate support consisting of tuplewise minimizers.

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

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

This characterization appears in the remark following Proposition 2.1 of Altschuler & Boix-Adsera (2022) and in Section 6 of Carlier & Ekeland (2010); it is also consistent with the finiteness results of Anderes et al. (2016). Thus, an optimal barycenter can be chosen with finite support, and every support point is a weighted least-squares minimizer associated with a tuple of input atoms.

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

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

for each index tuple $\vec{j} = (j_1, \dots, j_k)$. Although an optimal barycenter is finitely supported, the exact candidate set S^* may contain exponentially many points. Our approximation framework therefore replaces S^* with a substantially smaller candidate set.

2.2 Dual representation of the MOT formulation and candidate-support reduction

Let γ_{ij} be the dual variable associated with the marginal constraint for atom j of measure i . By LP strong duality, the barycenter value admits the representation:

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

The dual has at most nk variables, but it contains one constraint for every k -tuple $\vec{j}$ and therefore exponentially many constraints. Moreover, evaluating a constraint through the exact representation may require minimizing over the exponentially large set S^* . We address both difficulties by replacing S^* with a manageable candidate set $S \subseteq \mathbb{R}^d$. Our goal is to ensure that the transportation cost induced by the reduced support set S , for every tuple of input support points, is within a prescribed multiplicative factor of the unrestricted cost, either deterministically or in expectation.

For a finite candidate support $S = \{\mathbf{w}_\ell\}_{\ell \in [|S|]} \subseteq \mathbb{R}^d$, define the restricted tuple cost

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

for each multi-index $\vec{j} = (j_1, \dots, j_k)$. The corresponding restricted dual value is

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

Equivalently, $v(S)$ is the optimal barycenter value when all tuplewise barycenter locations are restricted to S .

When $|S|$ is polynomial in the input size, the restricted dual can be solved in polynomial time since its separation problem decomposes across marginals.

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

Proof. We observe that the fact that γ 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 following two-stage minimization problem:

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

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 $c_{ij_i\ell} = \lambda_i \|\mathbf{w}_\ell - \hat{\mathbf{x}}_{ij_i}\|_2^2 - \gamma_{ij_i}$. 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 (Section 3.1 in Grötschel et al. (2012)), the overall complexity for solving the restricted MOT dual (3) is $\tilde{O}((nk)^2 kn|S|d) = \tilde{O}(k^3 n^3 |S|d)$. $\square$

An optimal restricted MOT solution also yields an explicit barycenter supported on S .

Lemma 3. Fix a candidate support $S \subseteq \mathbb{R}^d$. Let $\Pi^*(S)$ 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

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

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

The tensor $\Pi^*(S)$ specifies the joint mass assigned to each tuple of input atoms. Assigning that mass to a best-fitting point $\mathbf{w}_{\vec{j}} \in S$ and aggregating over tuples produces the measure $\mathbb{P}(S)$. We next quantify how the restriction from S^* to a smaller candidate set $\tilde{S}$ affects the optimal value.

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 $\mathbf{x}_i \in \Xi_i$ for all $i \in [k]$, we have

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

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 $\mathbf{C}$ by the approximate one $\mathbf{C}(\tilde{S})$, the value of $v(\tilde{S})$ can be obtained by solving the following restricted MOT problem:

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

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

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

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

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

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).

Theorem 1 reduces the global barycenter approximation problem to a local one: it suffices to approximate the optimal one-center cost for every tuple of input atoms. The next two sections construct candidate supports satisfying (4), either deterministically or in expectation.

3 Approximation algorithms for the Wasserstein barycenter problem with general weights

Under squared Euclidean loss, the tuplewise minimizer in Lemma 1 is the weighted centroid. Hence, an optimal barycenter can be supported on

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

This set contains at most $\prod_{i \in [k]} |\Xi_i| = O(n^k)$ points and is therefore exponential in k . For general weights $\boldsymbol{\lambda} \in \Delta_k$, we construct smaller supports by averaging t sampled input atoms. Sampling with replacement preserves the weighted centroid in expectation, which leads to both a randomized algorithm and a deterministic enumeration counterpart with the same approximation ratio.

3.1 Sampling with replacement

Draw $T_1, \dots, T_t \in [k]$ independently according to $\boldsymbol{\lambda}$, so that

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

Let $\mathcal{T}$ be the resulting multiset. We construct

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

Thus, every candidate atom is the average of t input atoms, with one atom chosen from each sampled marginal. The support size is at most n^t , and hence polynomial in n for fixed t . Solving the restricted MOT problem over $S_1^{\mathcal{T}}$ and applying Lemma 3 produces the approximate barycenter $\mathbb{P}(S_1^{\mathcal{T}})$.

Algorithm 1 Sampling-with-replacement algorithm for the Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in [k]$.
 - 2: Independently draw $T_1, \dots, T_t \in [k]$ with $\text{Prob}[T_j = i] = \lambda_i$, and let $\mathcal{T}$ be the resulting multiset.
 - 3: $S_1^{\mathcal{T}} := \left\{ \frac{1}{t} \sum_{i \in [t]} \mathbf{x}_{T_i} : \mathbf{x}_{T_i} \in \Xi_{T_i} \right\}$.
 - 4: Solve the restricted MOT dual (3) and recover a corresponding optimal primal solution $\boldsymbol{\Pi}^*(S_1^{\mathcal{T}})$.
 - 5: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_1^{\mathcal{T}})$ using 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

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

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

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

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 $\mathbf{x}_{T_j} = \sum_{i \in [k]} \mathbf{x}_i \mathbb{I}_{\{T_j=i\}}$.

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

closest to $\mathbf{c}$. We have

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

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

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

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

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

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

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

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

Therefore, we have

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

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

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

This completes the proof. $\square$

Remark. Choosing $t = \lceil 1/\alpha \rceil$ gives an expected $(1+\alpha)$ -approximation. Since $|S_1^\mathcal{T}| = O(n^t)$, Lemma 2 yields the running-time bound $\tilde{O}(n^{3+\lceil 1/\alpha \rceil} k^3 d)$. Thus, for every fixed $\alpha > 0$, the randomized algorithm runs in polynomial time.

3.2 Deterministic counterpart based on multiset enumeration

The randomized construction can be derandomized by enumerating all multisets of t marginal indices and collecting every corresponding average. Define

$$S_1^t := \left\{ \frac{1}{t} \sum_{i \in [t]} \mathbf{x}_{T_i} : \mathbf{x}_{T_i} \in \Xi_{T_i}, T_i \in [k], i \in [t] \right\}. \quad (7)$$

Solving the restricted MOT problem over S_1^t and applying Lemma 3 yields the deterministic approximate barycenter $\mathbb{P}(S_1^t)$.

Algorithm 2 Multiset-enumeration algorithm for the Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in [k]$.
 - 2: $S_1^t := \left\{ \frac{1}{t} \sum_{i \in [t]} \mathbf{x}_{T_i} : \mathbf{x}_{T_i} \in \Xi_{T_i}, T_i \in [k], i \in [t] \right\}$.
 - 3: Solve the restricted MOT dual (3) and recover a corresponding optimal primal solution $\mathbf{\Pi}^*(S_1^t)$.
 - 4: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_1^t)$ using Lemma 3.
-

Since S_1^t contains the candidate support generated by every realization of $\mathcal{T}$, for any reference point $\mathbf{c}$, we have

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

The same argument as in Theorem 2 therefore gives the deterministic guarantee below.

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

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

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

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

holds for any $\mathbf{x}_i \in \Xi_i$ 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 at most $\binom{k+t-1}{t} n^t$ points in the set S_1^t by stars and bars method. Therefore, the size of S_1^t is $O(nk)^t$. $\square$

Remark. For $t = 1$, S_1^t is the union of the input supports, so Algorithm 2 reduces to the 2-approximation of Borgwardt (2022). More generally, choosing $t = \lceil 1/\alpha \rceil$ gives a deterministic $(1 + \alpha)$ -approximation. Since $|S_1^t| = O((nk)^t)$, the resulting running time is $\tilde{O}(n^{3+\lceil 1/\alpha \rceil} k^{3+\lceil 1/\alpha \rceil} d)$.

4 Approximation algorithms for the Wasserstein barycenter problem with equal weights

We now assume equal weights, $\lambda_i = 1/k$ for all $i \in [k]$. Sampling marginals without replacement reduces the variance of the sampled centroid and consequently improves the approximation ratio relative to the general-weight construction.

4.1 Sampling without replacement

Uniformly sample a subset $\mathcal{T} \subseteq [k]$ of cardinality t and define

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

The restricted MOT problem over $S_2^{\mathcal{T}}$ yields an approximate barycenter through Lemma 3.

Algorithm 3 Sampling-without-replacement algorithm for the Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in [k]$.
 - 2: Uniformly sample a subset $\mathcal{T} \subseteq [k]$ with $|\mathcal{T}| = t$.
 - 3: $S_2^{\mathcal{T}} := \left\{ \frac{1}{t} \sum_{i \in \mathcal{T}} \mathbf{x}_i : \mathbf{x}_i \in \Xi_i \right\}$.
 - 4: Solve the restricted MOT dual (3) and recover a corresponding optimal primal solution $\Pi^*(S_2^{\mathcal{T}})$.
 - 5: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_2^{\mathcal{T}})$ using 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

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

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

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\mathbf{s} \in S_2^{\mathcal{T}}} \frac{1}{k} \sum_{i \in [k]} \|\mathbf{s} - \mathbf{x}_i\|_2^2 \right] \leq \left(1 + \frac{k-t}{t(k-1)} \right) \frac{1}{k} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2. \quad (8a)$$

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

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

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

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\mathbf{s} \in S_2^{\mathcal{T}}} \frac{1}{k} \sum_{i \in [k]} \|\mathbf{s} - \mathbf{x}_i\|_2^2 \right] \leq \frac{1}{k} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2 + \mathbb{E}_{\mathcal{T}} \left[\min_{\mathbf{s} \in S_2^{\mathcal{T}}} \|\mathbf{s} - \mathbf{c}\|_2^2 \right]. \quad (8b)$$

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

Since $\hat{\boldsymbol{\mu}}_{\mathcal{T}} \in S_2^{\mathcal{T}}$, we have

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\mathbf{s} \in S_2^{\mathcal{T}}} \|\mathbf{s} - \mathbf{c}\|_2^2 \right] \leq \mathbb{E}_{\mathcal{T}} \left[\|\hat{\boldsymbol{\mu}}_{\mathcal{T}} - \mathbf{c}\|_2^2 \right]. \quad (8c)$$

Note that

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

In addition, we also have

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

where

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

by direct computation.

Therefore, we further have

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

which implies that

$$\begin{aligned} \text{Var} [\hat{\boldsymbol{\mu}}_{\mathcal{T}}] &= \mathbb{E}_{\mathcal{T}} [\|\hat{\boldsymbol{\mu}}_{\mathcal{T}} - \mathbf{c}\|_2^2] \\ &= \frac{1}{tk} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2 + \frac{t-1}{tk(k-1)} \sum_{i \neq j \in [k]} (\mathbf{x}_i - \mathbf{c})^{\top} (\mathbf{x}_j - \mathbf{c}) \\ &= \frac{1}{tk} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2 - \frac{t-1}{tk(k-1)} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2 \\ &= \frac{k-t}{t(k-1)} \frac{1}{k} \sum_{i \in [k]} \|\mathbf{c} - \mathbf{x}_i\|_2^2 \end{aligned} \quad (8d)$$

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

Combining (8b)-(8d), we have

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

□

Remark. A $(1 + \alpha)$ guarantee is obtained by choosing

$$t = \left\lceil \frac{k}{1 + \alpha(k-1)} \right\rceil,$$

which ensures $(k-t)/(t(k-1)) \leq \alpha$. Since $|S_2^{\mathcal{T}}| = O(n^t)$, the corresponding running time is $\tilde{O}(n^{3+t}k^3d) = \tilde{O}(n^{3+\lceil 1/\alpha \rceil}k^3d)$.

4.2 Deterministic counterpart based on subset enumeration

Enumerating all size- t subsets of marginals gives the deterministic candidate support

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

We solve the restricted MOT problem over S_2^t and recover $\mathbb{P}(S_2^t)$ through Lemma 3.

Algorithm 4 Subset-enumeration algorithm for the Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in [k]$.
 - 2: $S_2^t := \left\{ \frac{1}{t} \sum_{i \in \mathcal{T}} \mathbf{x}_i : \mathbf{x}_i \in \Xi_i, \mathcal{T} \subseteq [k], |\mathcal{T}| = t \right\}$.
 - 3: Solve the restricted MOT dual (3) and recover a corresponding optimal primal solution $\mathbf{\Pi}^*(S_2^t)$.
 - 4: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_2^t)$ using Lemma 3.
-

Since S_2^t contains the support generated by every realization of $\mathcal{T}$, for any reference point $\mathbf{c}$ we have

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

Applying the argument of Theorem 4 yields the following deterministic guarantee.

Theorem 5. Let $t \in [k]$ be a fixed constant, and let S_2^t be the candidate support returned by Algorithm 4. Then

$$v(S_2^t) \leq \left(1 + \frac{k-t}{t(k-1)}\right) v^*,$$

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

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

holds for any $\mathbf{x}_i \in \Xi_i$ 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 $O((nk)^t)$. $\square$

Remark. Since

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

the equal-weight algorithm is never worse than the general-weight construction for the same t and is strictly better whenever $1 < t < k$.

5 Numerical experiments

We evaluate the computational efficiency and solution quality of the proposed algorithms. The experiments are designed to (i) assess the accuracy–complexity tradeoff controlled by t ; (ii) demonstrate scalability on instances for which the extensive MOT formulation is intractable; and (iii) compare against the 2-approximation of Borgwardt (2022), which coincides with our deterministic construction at $t = 1$. All methods were implemented in Python with Gurobi 11.0.3 and run on a MacBook Pro equipped with an Apple M4 Pro chip and 48 GB of memory.

5.1 Synthetic nested ellipse dataset

We first consider the nested-ellipse benchmark used in Cuturi & Doucet (2014), Janati, Cuturi & Gramfort (2020), Altschuler & Boix-Adsera (2021), for which a reference optimal barycenter is available (Altschuler & Boix-Adsera 2021). The dataset contains ten probability measures, each represented on a 60×60 grid and corresponding to one nested ellipse; see Figure 1. The marginal supports contain between 139 and 192 equally weighted atoms. The exact barycenter need not to be on the original grid. A direct MOT formulation can involve on the order of $192^{10} \approx 10^{22}$ tuple variables and is therefore computationally intractable.

We run 10 runs of randomized sampling Algorithm 3 and pick the best solution. We also use deterministic subset enumeration Algorithm 4. Note that the deterministic candidate support is the union of the candidate supports generated by all possible sampling outcomes. It therefore contains the candidate support associated with every individual realization and yields an objective value no larger than that obtained from any single realization. Figure 2 illustrates this distinction: even the best of the ten possible $t = 1$ samples is visibly worse than the deterministic $t = 1$ solution. Accordingly, the remaining experiments focus on deterministic enumeration with $t = 1$ and $t = 2$.

The empirical performance is substantially better than the worst-case guarantees in Theorem 5. Moreover, small improvements in the objective can correspond to meaningful visual improvements:

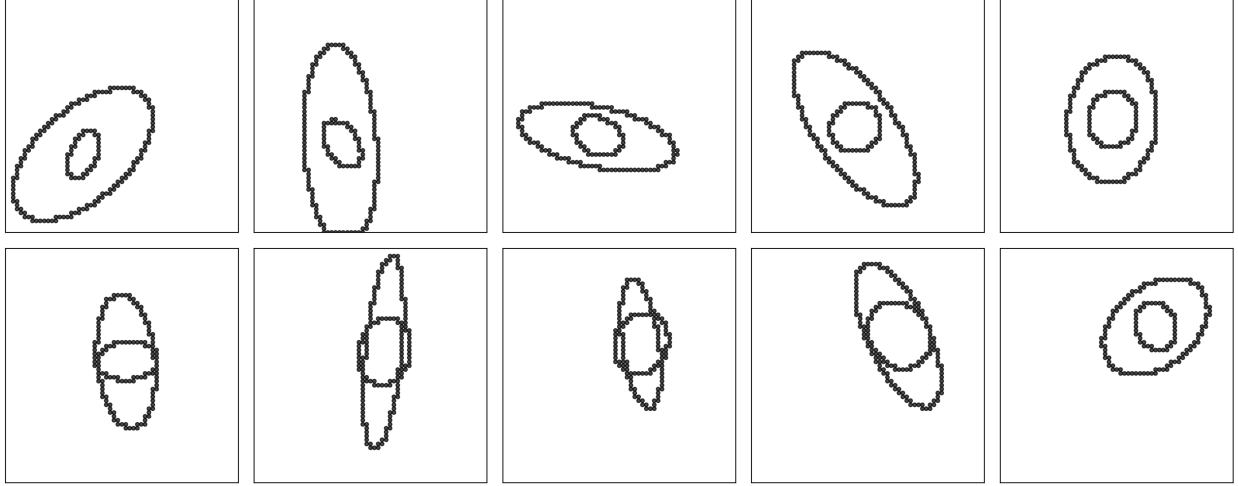
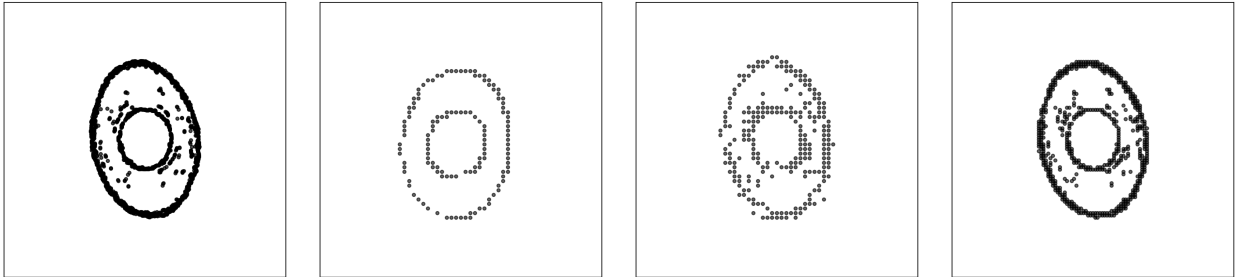


Figure 1: Ten measures from the nested-ellipse dataset.

although $t = 1$ is faster, $t = 2$ produces a barycenter that is nearly indistinguishable from the reference solution.



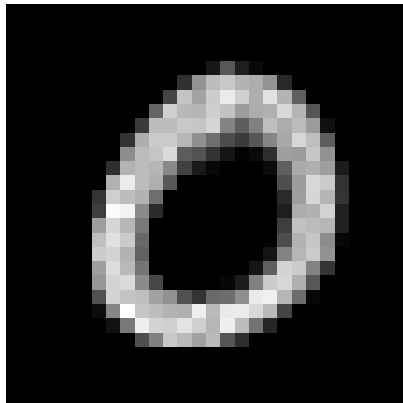
(a) Exact ellipse (Altschuler & Boix-Adsera 2021), objective=0.02666, time=8865.03s. (b) 10 runs of Algorithm 3 with $t = 1$, objective=0.02709, time=17.46s. (c) Algorithm 4 with $t = 1$, objective=0.02673, time=21.08s. (d) Algorithm 4 with $t = 2$, objective=0.02667, time=212.44s.

Figure 2: Barycenters produced by different algorithms on the nested-ellipse dataset.

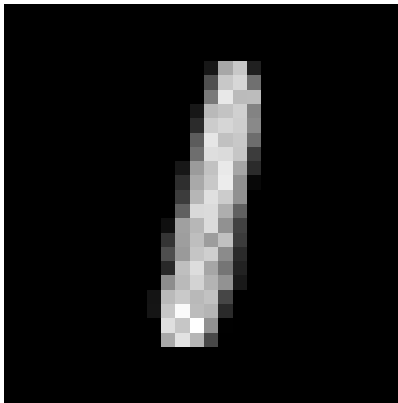
5.2 MNIST dataset

To assess scalability in the number of marginals, we use the MNIST dataset (LeCun et al. 2002). For each of three digit classes, we randomly select $k = 50$ empirical measures on a 28×28 grid ($d = 2$). Their mass distributions differ, and the number of positive-mass pixels ranges from 165 to 263. The full MOT formulation can contain on the order of $263^{50} \approx 10^{120}$ tuple variables.

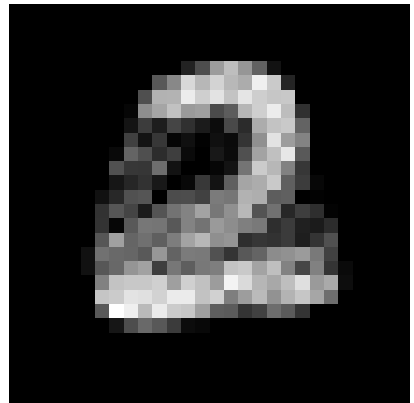
We run Algorithm 2 with $t = 1$ and $t = 2$ and report the objective values and running times in Figure 3. Both settings are computationally viable, while $t = 2$ consistently improves the objective value and preserves the characteristic shape of each digit more clearly.



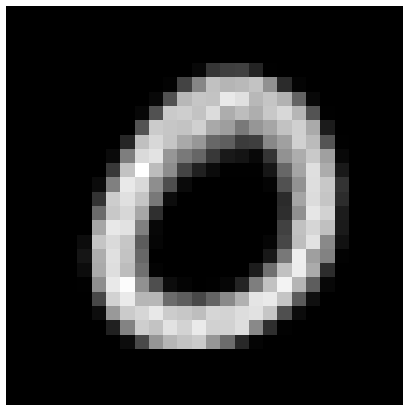
(a) Algorithm 2 with $t = 1$,
objective=0.0132, time=556s.



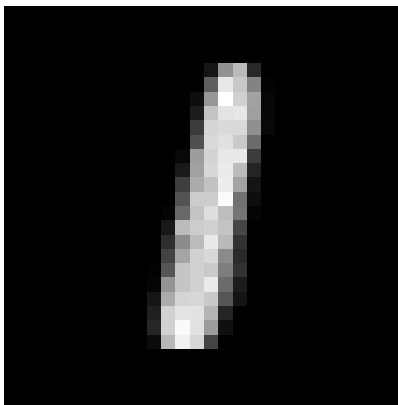
(b) Algorithm 2 with $t = 1$,
objective=0.0174, time=56s.



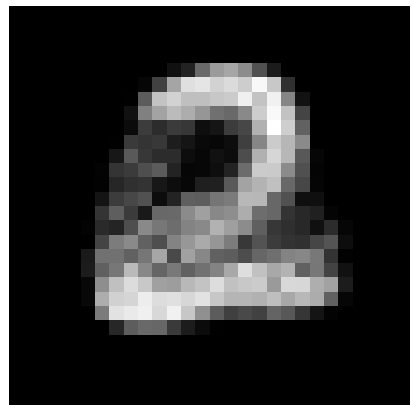
(c) Algorithm 2 with $t = 1$,
objective=0.0279, time=253s.



(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.

5.3 Sign language dataset

The final experiment studies measures with large supports. Each 64×64 RGB image is represented as a five-dimensional empirical measure: two coordinates encode spatial location, and three encode color. For each of three hand gestures, we randomly select three training images as the input measures.

Column generation implementation. Since every image has a large support, explicitly constructing the full candidate support and solving the resulting LP in Algorithm 4 is computationally expensive.

For $t = 1$, the candidate support reduces to the union of the marginal supports:

$$S_2^1 = \bigcup_{i \in [k]} \Xi_i.$$

Thus, S_2^1 contains at most $\sum_{i \in [k]} |\Xi_i|$ candidate atoms. For $t = 2$, the candidate support contains all pairwise centroids:

$$S_2^2 = \left\{ \frac{\mathbf{x}_r + \mathbf{x}_s}{2} : \mathbf{x}_r \in \Xi_r, \mathbf{x}_s \in \Xi_s, 1 \leq r < s \leq k \right\}.$$

Thus,

$$|S_2^2| \leq \sum_{1 \leq r < s \leq k} |\Xi_r| |\Xi_s|,$$

so the size of S_2^2 grows quadratically with the marginal support sizes. To control memory usage, we combine Algorithm 4 with a column-generation procedure in the spirit of Borgwardt & Patterson (2022). The restricted master problem retains only a small set of active columns, and the pricing problem adds a column only when its reduced cost is negative. Let

$$S_{\text{cand}} \subseteq S_1^1 \cup S_2^2$$

be a mixed candidate support containing both individual input atoms and pairwise centroids. A column is indexed by a candidate barycenter atom $\mathbf{w} \in S_{\text{cand}}$ and a tuple $\vec{j} = (j_1, \dots, j_k)$ with $j_i \in [|\Xi_i|]$. Its cost is

$$c(\mathbf{w}, \vec{j}) = \sum_{i \in [k]} \lambda_i \|\mathbf{w} - \hat{\mathbf{x}}_{ij_i}\|_2^2.$$

For a current column set $\mathcal{C}$, the restricted master problem is

$$\begin{aligned} \min_{\theta \geq 0} \quad & \sum_{(\mathbf{w}, \vec{j}) \in \mathcal{C}} c(\mathbf{w}, \vec{j}) \theta_{\mathbf{w}, \vec{j}} \\ \text{s.t.} \quad & \sum_{(\mathbf{w}, \vec{j}) \in \mathcal{C}: j_i = j} \theta_{\mathbf{w}, \vec{j}} = \hat{p}_{ij}, \quad \forall i \in [k], j \in [|\Xi_i|]. \end{aligned} \tag{CG-RMP}$$

The variable $\theta_{\mathbf{w}, \vec{j}}$ is the mass assigned to barycenter atom $\mathbf{w}$ through the tuple $(\hat{\mathbf{x}}_{1j_1}, \dots, \hat{\mathbf{x}}_{kj_k})$.

The induced mass of atom $\mathbf{w}$ and the corresponding barycenter are

$$m_{\mathbf{w}} = \sum_{\vec{j}:(\mathbf{w},\vec{j}) \in \mathcal{C}} \theta_{\mathbf{w},\vec{j}}, \quad \mathbb{P}_{\mathcal{C}} = \sum_{\mathbf{w} \in S_{\text{cand}}: m_{\mathbf{w}} > 0} m_{\mathbf{w}} \delta_{\mathbf{w}}.$$

Let γ_{ij} be the dual variable for the marginal constraint associated with $\hat{\mathbf{x}}_{ij}$. The reduced cost of $(\mathbf{w}, \vec{j})$ is

$$\bar{c}(\mathbf{w}, \vec{j}) = \sum_{i \in [k]} \lambda_i \|\mathbf{w} - \hat{\mathbf{x}}_{ij_i}\|_2^2 - \sum_{i \in [k]} \gamma_{ij_i}.$$

For a fixed $\mathbf{w}$, the minimizing tuple decomposes across marginals:

$$j_i(\mathbf{w}) \in \arg \min_{j \in [\Xi_i]} \left\{ \lambda_i \|\mathbf{w} - \hat{\mathbf{x}}_{ij}\|_2^2 - \gamma_{ij} \right\}, \forall i \in [k].$$

Hence, the pricing value for $\mathbf{w}$ is

$$r(\mathbf{w}) = \sum_{i \in [k]} \min_{j \in [\Xi_i]} \left\{ \lambda_i \|\mathbf{w} - \hat{\mathbf{x}}_{ij}\|_2^2 - \gamma_{ij} \right\}.$$

If $\min_{\mathbf{w} \in S_{\text{cand}}} r(\mathbf{w}) < 0$, we add the corresponding column $(\mathbf{w}, j_1(\mathbf{w}), \dots, j_k(\mathbf{w}))$ and reoptimize the restricted master problem. If no negative-reduced-cost column exists, the current solution is optimal for the LP restricted to S_{cand} .

For the sign-language instances, even scanning all of S_2^2 is expensive. We therefore construct a hybrid support intended to optimize the accuracy of $t = 2$ at a computational cost closer to $t = 1$. First, we solve the $t = 1$ restricted problem and collect its active atoms $A_1 = \{\mathbf{w} : m_{\mathbf{w}} > 0\}$. For each $\mathbf{a} \in A_1$ and marginal $i \in [k]$, let $\mathcal{N}_i^K(\mathbf{a}) \subseteq \Xi_i$ contain the K nearest atoms of $\mathbb{P}_i$ to $\mathbf{a}$; we set $K = 5$. Define

$$S_{\text{hyb}} = A_1 \cup \bigcup_{\mathbf{a} \in A_1} \left\{ \frac{\mathbf{x}_r + \mathbf{x}_s}{2} : \mathbf{x}_r \in \mathcal{N}_r^K(\mathbf{a}), \mathbf{x}_s \in \mathcal{N}_s^K(\mathbf{a}), \forall 1 \leq r < s \leq k \right\}.$$

The pricing step is performed over $S_{\text{cand}} = S_{\text{hyb}}$ rather than the full candidate set S_2^2 . The hybrid set contains the active atoms of the $t = 1$ solution and pairwise centroids formed from their neighboring input atoms. This localized construction substantially reduces the number of candidate atoms while retaining potentially useful $t = 2$ support points.

We warm-start column generation with the active $t = 1$ columns, which provide a feasible restricted master problem. The procedure terminates when either no negative-reduced-cost column is found or the time limit is reached. Under a time limit, the final restricted-master solution remains a feasible barycenter, and we report its objective value. The worst-case guarantee of Algorithm 4 requires optimizing over the full enumerated support; it does not apply to this hybrid acceleration. We refer to the implementation as ‘‘Hybrid Algorithm 4 with $t = 2$.’’

Results. Figures 4–6 show representative results for the three gestures. Relative to the best $t = 1$ solutions reported in the figures, the best hybrid $t = 2$ solutions reduce the objective values

by approximately 2.8%, 4.3%, and 5.4% for gestures 1, 2, and 3, respectively. The hybrid $t = 2$ barycenters also provide visually clearer representations of the corresponding hand shapes. For gesture 1, Figures 4a and 4b show the $t = 1$ solutions obtained after one and two hours, respectively. Figures 4c and 4d show the hybrid $t = 2$ solutions obtained after one and two additional hours, respectively, using the one-hour $t = 1$ solution as a warm start. The hybrid $t = 2$ barycenters form fuller and more clearly delineated hand shapes than their $t = 1$ counterparts. Figures 5 and 6 exhibit similar improvements for gestures 2 and 3.

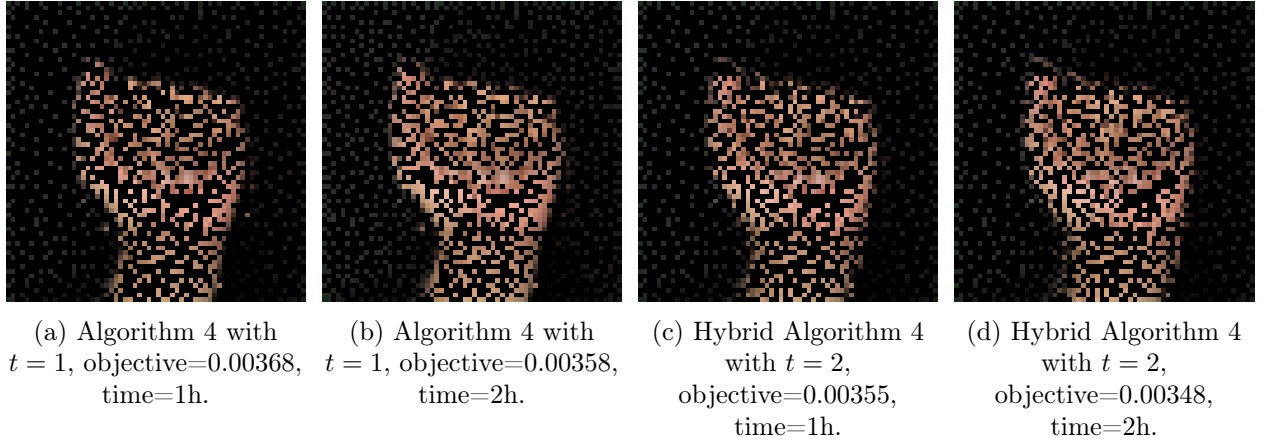


Figure 4: Barycenters produced by different algorithms for gesture 1.

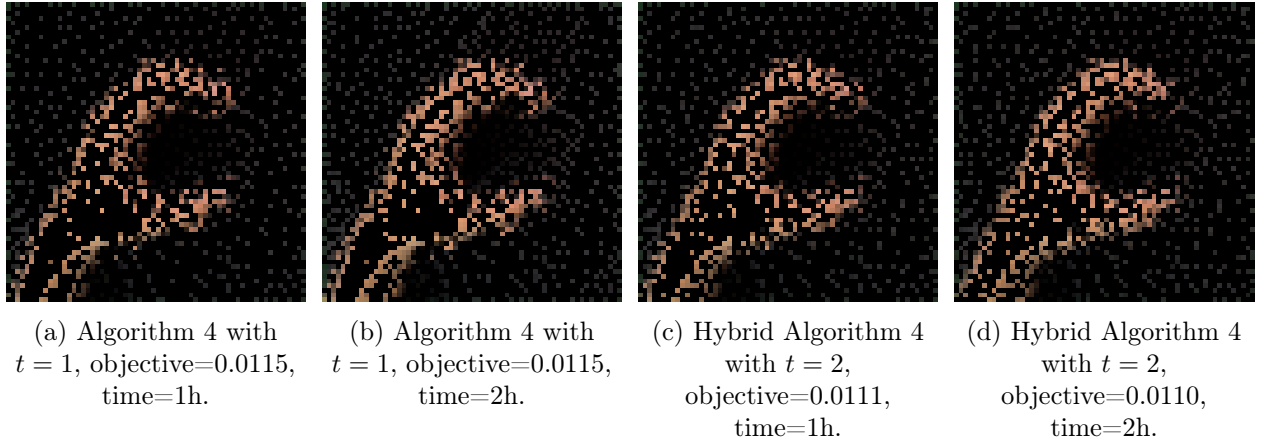


Figure 5: Barycenters produced by different algorithms for gesture 2.

We also evaluate the barycenters in a nearest-barycenter classification task. Each gesture class is represented by its training-sample barycenter, and each of the 25 test images per gesture is assigned to the class with the smallest Wasserstein distance. The two-hour $t = 1$ barycenters yield 96% accuracy, whereas the one-hour hybrid $t = 2$ barycenters classify all 75 images correctly. Thus, the improvement in the barycenter objective also translates into better downstream classification performance; see Table 1.

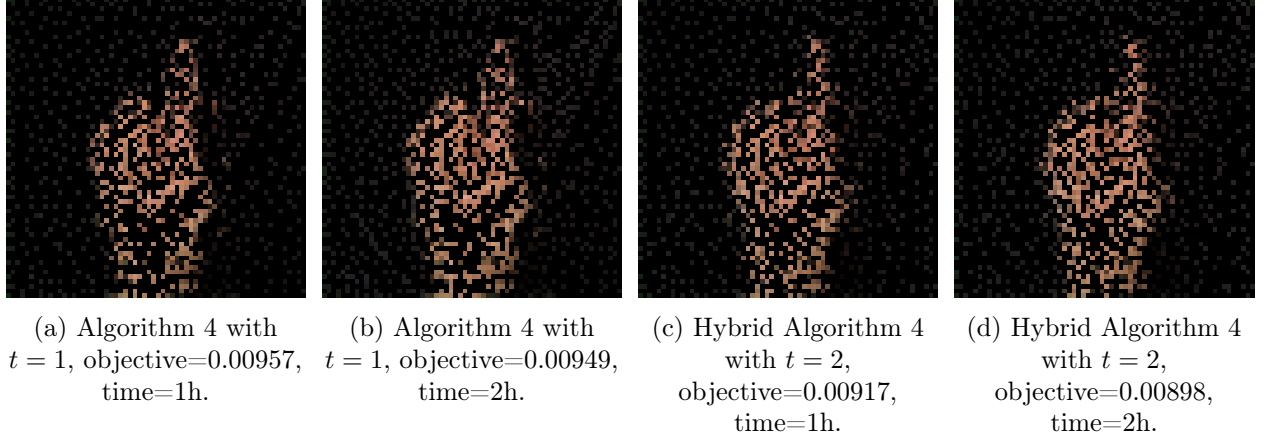


Figure 6: Barycenters produced by different algorithms for gesture 3.

Table 1: Gesture-classification results. Each entry is the number of test images in the indicated category. For example, Algorithm 4 with $t = 1$ correctly classifies 22 of the 25 gesture-2 images and misclassifies the remaining 3.

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

6 Extensions

We consider two extensions of the candidate-support framework. First, we show that the deterministic support S_1^t yields a value guarantee for sparse type-2 barycenters, although the resulting restricted problem remains a mixed-integer program. Second, we exploit coordinatewise weighted medians to construct exact and approximate supports for type-1 barycenters. Extending comparable guarantees to general type- p barycenters with $p \notin \{1, 2\}$ remains open.

6.1 Approximation guarantee for sparse type-2 Wasserstein barycenters

Consider the support-size-constrained Wasserstein barycenter problem:

$$v_m^* := \inf_{\substack{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d) \\ |\text{supp}(\mathbb{P})| \leq m}} \sum_{i \in [k]} \lambda_i W_2^2(\mathbb{P}, \mathbb{P}_i).$$

For a finite candidate set $G \subseteq \mathbb{R}^d$, define the value of the corresponding restricted sparse Wasserstein barycenter problem as

$$v_m(G) := \inf_{\substack{\mathbb{P} \in \mathcal{P}(\mathbb{R}^d) \\ \text{supp}(\mathbb{P}) \subseteq G, |\text{supp}(\mathbb{P})| \leq m}} \sum_{i \in [k]} \lambda_i W_2^2(\mathbb{P}, \mathbb{P}_i).$$

Mixed-integer formulation of the restricted problem. Let $S = \{\mathbf{s}_1, \dots, \mathbf{s}_M\} \subseteq \mathbb{R}^d$ be a finite candidate support, and write $\mathbb{P}_i = \sum_{j \in [\Xi_i]} \hat{p}_{ij} \delta_{\hat{\mathbf{x}}_{ij}}$ for each $i \in [k]$. For each candidate atom $\mathbf{s}_\ell$, let z_ℓ denote its barycenter mass and let $y_\ell \in \{0, 1\}$ be a binary activation variable for that atom. For each $i \in [k]$, $j \in [\Xi_i]$, and $\ell \in [M]$, let $\pi_{ij\ell}$ denote the amount of mass transported from the barycenter atom $\mathbf{s}_\ell$ to the input atom $\hat{\mathbf{x}}_{ij}$ in the transport plan between the barycenter and $\mathbb{P}_i$. The restricted m -sparse problem over S can then be formulated as the following mixed-integer linear program:

$$\begin{aligned}
v_m(S) = \min_{\boldsymbol{\pi}, \mathbf{z}, \mathbf{y}} \quad & \sum_{i \in [k]} \lambda_i \sum_{j \in [\Xi_i]} \sum_{\ell \in [M]} \|\mathbf{s}_\ell - \hat{\mathbf{x}}_{ij}\|_2^2 \pi_{ij\ell} \\
\text{s.t.} \quad & \sum_{\ell \in [M]} \pi_{ij\ell} = \hat{p}_{ij}, & \forall i \in [k], j \in [\Xi_i], \\
& \sum_{j \in [\Xi_i]} \pi_{ij\ell} = z_\ell, & \forall i \in [k], \ell \in [M], \\
& 0 \leq z_\ell \leq y_\ell, & \forall \ell \in [M], \\
& \sum_{\ell \in [M]} y_\ell \leq m, \\
& \pi_{ij\ell} \geq 0, \quad y_\ell \in \{0, 1\}, & \forall i \in [k], j \in [\Xi_i], \ell \in [M].
\end{aligned} \tag{RWBCenter}$$

The first constraint family ensures that the transport plan associated with each marginal $\mathbb{P}_i$ has the prescribed marginal masses $\hat{p}_{ij}$. The second constraint family requires all k transport plans to share the same barycenter mass vector $\mathbf{z} = (z_1, \dots, z_M)$. The linking constraints $z_\ell \leq y_\ell$ deactivate unselected atoms, and $\sum_{\ell \in [M]} y_\ell \leq m$ limits the support size. The final constraints impose nonnegativity and integrality.

Given an optimal solution $(\boldsymbol{\pi}^*, \mathbf{z}^*, \mathbf{y}^*)$, the corresponding restricted sparse barycenter is

$$\mathbb{P}_{S,m} = \sum_{\ell \in [M]} z_\ell^* \delta_{\mathbf{s}_\ell}.$$

Candidate atoms with $z_\ell^* = 0$ can be omitted from its support. We next analyze the value obtained when this MILP is solved to optimality over the candidate support S_1^t .

Approximation guarantee. We next show that the deterministic candidate support S_1^t defined in (7) can also be used to approximate sparse barycenters. The following lemma provides the key ingredient: it shows that S_1^t yields a $(1 + 1/t)$ -approximation for the weighted one-center problem induced by any probability distribution over the indexed input atoms.

Lemma 4. For every matrix $\mathbf{q} \in \mathbb{R}_+^{k \times n}$ satisfying $\sum_{i \in [k]} \sum_{j \in \Xi_i} q_{ij} = 1$, we have

$$\min_{\mathbf{w} \in S_1^t} \sum_{i \in [k]} \sum_{j \in \Xi_i} q_{ij} \|\mathbf{w} - \hat{\mathbf{x}}_{ij}\|_2^2 \leq \left(1 + \frac{1}{t}\right) \min_{\mathbf{w} \in \mathbb{R}^d} \sum_{i \in [k]} \sum_{j \in \Xi_i} q_{ij} \|\mathbf{w} - \hat{\mathbf{x}}_{ij}\|_2^2.$$

Proof. Let $\mathbf{Y}$ be a random variable supported on $\{\hat{\mathbf{x}}_{ij} : i \in [k], j \in \Xi_i\}$ with $\mathbb{P}(\mathbf{Y} = \hat{\mathbf{x}}_{ij}) = q_{ij}$.

Let $\boldsymbol{\mu} := \mathbb{E}[\mathbf{Y}]$. Then $\boldsymbol{\mu} \in \arg \min_{\mathbf{w} \in \mathbb{R}^d} \mathbb{E} \|\mathbf{w} - \mathbf{Y}\|_2^2$, and $\min_{\mathbf{w} \in \mathbb{R}^d} \mathbb{E} \|\mathbf{w} - \mathbf{Y}\|_2^2 = \mathbb{E} \|\boldsymbol{\mu} - \mathbf{Y}\|_2^2$. Now draw $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ independently from the distribution of $\mathbf{Y}$, and define $\bar{\mathbf{Y}}_t := \frac{1}{t} \sum_{r=1}^t \mathbf{Y}_r$. By taking expectation over $\mathbf{Y}, \mathbf{Y}_1, \dots, \mathbf{Y}_t$, and expand the squares similar to (6b), the approximation follows from the proof of Theorem 2. $\square$

We next transfer this local centroid guarantee to the sparse barycenter objective.

Theorem 6. For every integer $t \geq 1$,

$$v_m(S_1^t) \leq \left(1 + \frac{1}{t}\right) v_m^*.$$

Proof. Choose an optimal m -sparse barycenter $\mathbb{P} = \sum_{\ell=1}^m a_\ell \delta_{\mathbf{w}_\ell}$ such that $\sum_{i \in [k]} \lambda_i W_2^2(\mathbb{P}, \mathbb{P}_i) = v_m^*$ and $\mathbf{a} \in \mathbb{R}_{++}^m$, $\sum_{\ell=1}^m a_\ell = 1$. For each $i \in [k]$, let $\boldsymbol{\pi}_i = \{\pi_{i\ell j}\}_{\ell \in [m], j \in [\Xi_i]}$ be an optimal transport plan between $\mathbb{P}$ and $\mathbb{P}_i$. Hence,

$$\sum_{j \in [\Xi_i]} \pi_{i\ell j} = a_\ell, \quad \sum_{\ell=1}^m \pi_{i\ell j} = \hat{p}_{ij}.$$

For each active barycenter atom $\mathbf{w}_\ell$, define the probability mass $q_{ij}^\ell := \frac{\lambda_i \pi_{i\ell j}}{a_\ell}$ for each $i \in [k]$ and $j \in [\Xi_i]$. This is indeed a probability vector, since

$$\sum_{i \in [k]} \sum_{j \in [\Xi_i]} q_{ij}^\ell = \frac{1}{a_\ell} \sum_{i \in [k]} \lambda_i \sum_{j \in [\Xi_i]} \pi_{i\ell j} = \frac{1}{a_\ell} \sum_{i \in [k]} \lambda_i a_\ell = 1.$$

By Lemma 4, for each $\ell \in [m]$, there exists $\mathbf{s}_\ell \in S_1^t$ such that

$$\sum_{i \in [k]} \sum_{j \in [\Xi_i]} q_{ij}^\ell \|\mathbf{s}_\ell - \hat{\mathbf{x}}_{ij}\|_2^2 \leq \left(1 + \frac{1}{t}\right) \min_{\mathbf{w} \in \mathbb{R}^d} \sum_{i \in [k]} \sum_{j \in [\Xi_i]} q_{ij}^\ell \|\mathbf{w} - \hat{\mathbf{x}}_{ij}\|_2^2.$$

In particular, since $\mathbf{w}_\ell \in \mathbb{R}^d$ is feasible to the right-hand minimization problem, we have

$$\sum_{i \in [k]} \sum_{j \in [\Xi_i]} q_{ij}^\ell \|\mathbf{s}_\ell - \hat{\mathbf{x}}_{ij}\|_2^2 \leq \left(1 + \frac{1}{t}\right) \sum_{i \in [k]} \sum_{j \in [\Xi_i]} q_{ij}^\ell \|\mathbf{w}_\ell - \hat{\mathbf{x}}_{ij}\|_2^2. \quad (9)$$

Now define $\tilde{\mathbb{P}} := \sum_{\ell=1}^m a_\ell \delta_{\mathbf{s}_\ell}$. If several $\mathbf{s}_\ell$'s coincide, we merge their masses. Therefore, $\tilde{\mathbb{P}}$ is supported on S_1^t and has support size at most m . We construct a feasible transportation plan from $\tilde{\mathbb{P}}$ to each $\mathbb{P}_i$ by transporting the mass a_ℓ at $\mathbf{s}_\ell$ to the points $\hat{\mathbf{x}}_{ij}$ using the same coefficients $\pi_{i\ell j}$. Thus,

$$v_m(S_1^t) \leq \sum_{i \in [k]} \lambda_i W_2^2(\tilde{\mathbb{P}}, \mathbb{P}_i) \leq \sum_{i \in [k]} \lambda_i \sum_{\ell=1}^m \sum_{j \in [\Xi_i]} \pi_{i\ell j} \|\mathbf{s}_\ell - \hat{\mathbf{x}}_{ij}\|_2^2$$

$$\begin{aligned}
&= \sum_{\ell=1}^m a_{\ell} \sum_{i \in [k]} \sum_{j \in \Xi_i} q_{ij}^{\ell} \|\mathbf{s}_{\ell} - \hat{\mathbf{x}}_{ij}\|_2^2 \leq \left(1 + \frac{1}{t}\right) \sum_{\ell=1}^m a_{\ell} \sum_{i \in [k]} \sum_{j \in \Xi_i} q_{ij}^{\ell} \|\mathbf{w}_{\ell} - \hat{\mathbf{x}}_{ij}\|_2^2 \\
&= \left(1 + \frac{1}{t}\right) \sum_{i \in [k]} \lambda_i \sum_{\ell=1}^m \sum_{j \in [\Xi_i]} \pi_{i\ell j} \|\mathbf{w}_{\ell} - \hat{\mathbf{x}}_{ij}\|_2^2 \\
&= \left(1 + \frac{1}{t}\right) \sum_{i \in [k]} \lambda_i W_2^2(\mathbb{P}, \mathbb{P}_i) = \left(1 + \frac{1}{t}\right) v_m^*,
\end{aligned}$$

where the first inequality is since $\text{supp}(\tilde{\mathbb{P}}) \subseteq S_1^t$, the second inequality is since $\boldsymbol{\pi}_i$ is a feasible transportation plan from $\tilde{\mathbb{P}}$ to $\mathbb{P}_i$, the first and second equalities are by the definition of q_{ij}^{ℓ} , the third inequality follows (9), the third equality is by the definition of $(\pi_{i\ell j}, \mathbf{w}_{\ell})$, and the last inequality is by the optimality of $\mathbb{P}$. This completes the proof. $\square$

Theorem 6 shows that restricting the barycenter support to S_1^t increases the optimal m -sparse barycenter value by at most a factor of $1 + 1/t$. This result should be interpreted as a candidate-support reduction guarantee rather than a PTAS for the sparse barycenter problem. Indeed, unlike the nonsparse problem studied earlier, restricting the candidate support does not reduce the sparse problem to an LP: computing $v_m(S_1^t)$ still requires solving the MILP (RWBCenter) to optimality. Nevertheless, the reduced candidate support can substantially decrease the size of the resulting formulation and therefore improve its practical tractability.

6.2 Approximation guarantee for the type-1 Wasserstein barycenter problem

The discrete type-1 Wasserstein barycenter problem is

$$v_I^* = \inf_{\mathbb{P} \in \mathcal{P}_1(\mathbb{R}^d)} \sum_{i \in [k]} \lambda_i W_1(\mathbb{P}, \mathbb{P}_i),$$

where $\mathcal{P}_1(\mathbb{R}^d)$ is the set of Borel probability measures with finite first moment and

$$W_1(\mathbb{P}, \mathbb{Q}) := \inf_{\Pi \in \mathcal{M}(\mathbb{P}, \mathbb{Q})} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\|_1 d\Pi(\mathbf{x}, \mathbf{y}).$$

The MOT representation underlying Lemma 1 extends directly to the ℓ_1 transportation cost. For each tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k)$, its tuplewise transportation cost is

$$\min_{\mathbf{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1.$$

Unlike the squared Euclidean case, the minimizer need not be unique. Nevertheless, the coordinate-wise separability of the ℓ_1 norm allows us to construct a finite exact candidate support. Indeed, we have

$$\min_{\mathbf{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1 = \sum_{j \in [d]} \min_{s_j \in \mathbb{R}} \sum_{i \in [k]} \lambda_i |s_j - x_{ij}|.$$

Thus, a tuplewise minimizer can be obtained by solving d one-dimensional weighted-median problems (Vazler et al. 2012, Sabo & Scitovski 2008).

Lemma 5 (Weighted median in one dimension; Lemma 3.1 in Sabo & Scitovski 2008). Let $a_1 \leq a_2 \leq \dots \leq a_k$ be real numbers with corresponding weights $\lambda_1, \dots, \lambda_k$, where $\sum_{i \in [k]} \lambda_i = 1$. Define $J = \left\{ \nu \in [k] : \sum_{i \in [\nu]} \lambda_i \leq \frac{1}{2} \right\}$. If $J \neq \emptyset$, let $\nu_0 = \max_{\nu \in J} \nu$. Then:

- (i) If $J = \emptyset$, then $\min_{s \in \mathbb{R}} \sum_{i \in [k]} \lambda_i |a_i - s|$ is attained at $s^* = a_1$.
- (ii) If $J \neq \emptyset$ and $\sum_{i=1}^{\nu_0} \lambda_i < \frac{1}{2}$, then the minimum is attained at $s^* = a_{\nu_0+1}$.
- (iii) If $J \neq \emptyset$ and $\sum_{i=1}^{\nu_0} \lambda_i = \frac{1}{2}$, then every point $s^* \in [a_{\nu_0}, a_{\nu_0+1}]$ is optimal.

For each coordinate $j \in [d]$, define the set of observed coordinate values

$$U_j := \left\{ x_j : \mathbf{x} \in \bigcup_{i \in [k]} \Xi_i \right\}.$$

We define the coordinate-product candidate support

$$\hat{S}_1 = \left\{ \mathbf{y} \in \mathbb{R}^d : y_j \in U_j \text{ for every } j \in [d] \right\}.$$

By Lemma 5, for every input tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k)$, each coordinate of a tuplewise minimizer can be chosen from the corresponding coordinate values appearing in that tuple. Such values belong to U_j , and therefore $\hat{S}_1$ contains a tuplewise minimizer for every input tuple. Equivalently, for every $(\mathbf{x}_1, \dots, \mathbf{x}_k)$, we have

$$\min_{\mathbf{s} \in \hat{S}_1} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1 = \min_{\mathbf{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1.$$

Thus, restricting the type-1 MOT formulation to $\hat{S}_1$ leaves every tuplewise cost unchanged and therefore preserves the optimal barycenter value. Moreover, since

$$|U_j| \leq \sum_{i \in [k]} |\Xi_i| \leq nk, \quad j \in [d],$$

we have

$$|\hat{S}_1| = \prod_{j \in [d]} |U_j| \leq \left(\sum_{i \in [k]} |\Xi_i| \right)^d \leq (nk)^d.$$

The coordinate-product structure of $\hat{S}_1$ motivates the reduced candidate supports introduced next.

Subspace-sampling algorithm. Unlike the type-2 constructions, which sample marginals, the type-1 method samples coordinates. Let $\hat{U} := \bigcup_{i \in [k]} \Xi_i$ and, for each $\ell \in [d]$, let $U_\ell := \{x_\ell : \mathbf{x} \in \hat{U}\}$.

For $\mathcal{T} \subseteq [d]$, write $\mathbf{x}_{\mathcal{T}}$ for the corresponding subvector. Fix $t \in \{0, \dots, d\}$ and uniformly sample $\mathcal{T} \subseteq [d]$ with $|\mathcal{T}| = d - t$. Define

$$S_3^{\mathcal{T}} := \left\{ \mathbf{y} \in \mathbb{R}^d : \text{there exists } \mathbf{x} \in \hat{U} \text{ such that } \mathbf{y}_{\mathcal{T}} = \mathbf{x}_{\mathcal{T}}, \text{ and } y_{\ell} \in U_{\ell} \text{ for every } \ell \notin \mathcal{T} \right\}.$$

The coordinates in $\mathcal{T}$ are inherited jointly from one observed support point, whereas the remaining t coordinates are selected independently from their observed values. Since $|\hat{U}| \leq nk$ and $|U_{\ell}| \leq nk$, the support size satisfies

$$|S_3^{\mathcal{T}}| \leq (nk)(nk)^t = (nk)^{t+1}.$$

Algorithm 5 Subspace-sampling algorithm for the type-1 Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in \{0, \dots, d\}$.
 - 2: Uniformly sample $\mathcal{T} \subseteq [d]$ with $|\mathcal{T}| = d - t$.
 - 3: Construct the candidate support $S_3^{\mathcal{T}}$.
 - 4: Solve the type-1 restricted MOT problem over $S_3^{\mathcal{T}}$ and recover an optimal primal solution $\Pi^*(S_3^{\mathcal{T}})$.
 - 5: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_3^{\mathcal{T}})$ using Lemma 3.
-

Theorem 7. Let $S_3^{\mathcal{T}}$ be the random candidate support returned by Algorithm 5. Then

$$\mathbb{E}_{\mathcal{T}} [v_I(S_3^{\mathcal{T}})] \leq \left(1 + \frac{d-t}{d}\right) v_I^*.$$

Proof. By the type-1 analogue of Theorem 1, it suffices to prove the corresponding approximation guarantee for arbitrary support points $\mathbf{x}_i \in \Xi_i$ for all $i \in [k]$. Let $\mathbf{c} \in \arg \min_{\mathbf{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1$.

Since the ℓ_1 objective is separable across coordinates, $\mathbf{c}$ can be chosen such that each coordinate c_j is a weighted median of $\{x_{ij}\}_{i \in [k]}$ and, in particular, is equal to one of the observed coordinate values. Thus, $c_j \in U_j$ for all $j \in [d]$.

For a fixed realization of $\mathcal{T}$, define

$$D_{\mathcal{T}}(\mathbf{c}) := \min_{\mathbf{s} \in S_3^{\mathcal{T}}} \|\mathbf{s} - \mathbf{c}\|_1,$$

and choose

$$\mathbf{s}_{\mathcal{T}} \in \arg \min_{\mathbf{s} \in S_3^{\mathcal{T}}} \|\mathbf{s} - \mathbf{c}\|_1.$$

By the triangle inequality and $\sum_{i \in [k]} \lambda_i = 1$, we obtain

$$\min_{\mathbf{s} \in S_3^{\mathcal{T}}} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1 \leq \sum_{i \in [k]} \lambda_i \|\mathbf{s}_{\mathcal{T}} - \mathbf{x}_i\|_1 \leq \sum_{i \in [k]} \lambda_i (\|\mathbf{s}_{\mathcal{T}} - \mathbf{c}\|_1 + \|\mathbf{c} - \mathbf{x}_i\|_1)$$

$$= D_{\mathcal{T}}(\mathbf{c}) + \sum_{i \in [k]} \lambda_i \|\mathbf{c} - \mathbf{x}_i\|_1. \quad (10a)$$

We next bound the expected value of $D_{\mathcal{T}}(\mathbf{c})$. For each $i \in [k]$, define the mixed point $\tilde{\mathbf{x}}_i^{\mathcal{T}} \in \mathbb{R}^d$ coordinatewise by

$$\tilde{\mathbf{x}}_{ij}^{\mathcal{T}} = \begin{cases} x_{ij}, & j \in \mathcal{T}, \\ c_j, & j \notin \mathcal{T}. \end{cases}$$

The coordinates indexed by $\mathcal{T}$ are inherited from the single point $\mathbf{x}_i \in \hat{U}$, while each remaining coordinate c_j belongs to U_j . Therefore, $\tilde{\mathbf{x}}_i^{\mathcal{T}} \in S_3^{\mathcal{T}}$. It follows that

$$D_{\mathcal{T}}(\mathbf{c}) \leq \|\tilde{\mathbf{x}}_i^{\mathcal{T}} - \mathbf{c}\|_1 = \sum_{j \in \mathcal{T}} |x_{ij} - c_j|, \quad \forall i \in [k]. \quad (10b)$$

Since $\mathcal{T}$ is sampled uniformly from all subsets of $[d]$ of cardinality $d-t$, every coordinate belongs to $\mathcal{T}$ with probability $(d-t)/d$. Hence, for every $i \in [k]$,

$$\begin{aligned} \mathbb{E}_{\mathcal{T}} [\|\tilde{\mathbf{x}}_i^{\mathcal{T}} - \mathbf{c}\|_1] &= \mathbb{E}_{\mathcal{T}} \left[\sum_{j \in \mathcal{T}} |x_{ij} - c_j| \right] = \sum_{j \in [d]} \Pr[j \in \mathcal{T}] |x_{ij} - c_j| = \frac{d-t}{d} \sum_{j \in [d]} |x_{ij} - c_j| \\ &= \frac{d-t}{d} \|\mathbf{x}_i - \mathbf{c}\|_1. \end{aligned} \quad (10c)$$

Taking expectations in (10b), multiplying the resulting inequality by λ_i , and summing over $i \in [k]$ yield

$$\mathbb{E}_{\mathcal{T}} [D_{\mathcal{T}}(\mathbf{c})] = \sum_{i \in [k]} \lambda_i \mathbb{E}_{\mathcal{T}} [D_{\mathcal{T}}(\mathbf{c})] \leq \sum_{i \in [k]} \lambda_i \mathbb{E}_{\mathcal{T}} [\|\tilde{\mathbf{x}}_i^{\mathcal{T}} - \mathbf{c}\|_1] = \frac{d-t}{d} \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1 \quad (10d)$$

Taking expectations in (10a) and applying (10d), we obtain

$$\mathbb{E}_{\mathcal{T}} \left[\min_{\mathbf{s} \in S_3^{\mathcal{T}}} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1 \right] \leq \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1 + \mathbb{E}_{\mathcal{T}} [D_{\mathcal{T}}(\mathbf{c})] \leq \left(1 + \frac{d-t}{d} \right) \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1.$$

Since the tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ was arbitrary, the result follows from Theorem 1. $\square$

Deterministic counterpart: subspace enumeration. Enumerating all subsets $\mathcal{T} \subseteq [d]$ with $|\mathcal{T}| = d-t$ gives $S_3^t := \bigcup_{\substack{\mathcal{T} \subseteq [d] \\ |\mathcal{T}| = d-t}} S_3^{\mathcal{T}}$. Its cardinality is bounded by

$$|S_3^t| \leq \binom{d}{d-t} (nk)^{t+1} = \binom{d}{t} (nk)^{t+1} = O(d^t (nk)^{t+1})$$

for fixed t . Then we have the following deterministic performance guarantee.

Algorithm 6 Subspace-enumeration algorithm for the type-1 Wasserstein barycenter problem

- 1: **Input:** Probability measures $\{\mathbb{P}_i\}_{i \in [k]}$ with supports $\{\Xi_i\}_{i \in [k]}$ and an integer $t \in \{0, \dots, d\}$.
 - 2: Construct $S_3^t := \bigcup_{\substack{\mathcal{T} \subseteq [d] \\ |\mathcal{T}|=d-t}} S_3^{\mathcal{T}}$.
 - 3: Solve the type-1 restricted MOT problem over S_3^t and recover an optimal primal solution $\mathbf{\Pi}^*(S_3^t)$.
 - 4: **Output:** Construct and return the approximate barycenter $\mathbb{P}(S_3^t)$ using Lemma 3.
-

Theorem 8. Let S_3^t be the candidate support returned by Algorithm 6. Then

$$v_I(S_3^t) \leq \left(1 + \frac{d-t}{d}\right) v_I^*.$$

Proof. As in the proof of Theorem 7, fix arbitrary support points $\mathbf{x}_i \in \Xi_i$, $i \in [k]$, and let

$$\mathbf{c} \in \arg \min_{\mathbf{s} \in \mathbb{R}^d} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1$$

be chosen such that c_j is an observed coordinate value for every $j \in [d]$. For each $i \in [k]$, choose a subset

$$\mathcal{T}_i \in \arg \min_{\substack{\mathcal{T} \subseteq [d] \\ |\mathcal{T}|=d-t}} \sum_{j \in \mathcal{T}} |x_{ij} - c_j|.$$

Thus, $\mathcal{T}_i$ consists of the $d-t$ coordinates having the smallest values of $|x_{ij} - c_j|$. Therefore,

$$\sum_{j \in \mathcal{T}_i} |x_{ij} - c_j| \leq \frac{d-t}{d} \sum_{j \in [d]} |x_{ij} - c_j| = \frac{d-t}{d} \|\mathbf{x}_i - \mathbf{c}\|_1. \quad (11a)$$

Define $\tilde{\mathbf{x}}_i \in \mathbb{R}^d$ by

$$\tilde{x}_{ij} = \begin{cases} x_{ij}, & j \in \mathcal{T}_i, \\ c_j, & j \notin \mathcal{T}_i. \end{cases}$$

By construction, we have $\tilde{\mathbf{x}}_i \in S_3^{\mathcal{T}_i} \subseteq S_3^t$. Therefore, for every $i \in [k]$,

$$\min_{\mathbf{s} \in S_3^t} \|\mathbf{s} - \mathbf{c}\|_1 \leq \|\tilde{\mathbf{x}}_i - \mathbf{c}\|_1 = \sum_{j \in \mathcal{T}_i} |x_{ij} - c_j| \leq \frac{d-t}{d} \|\mathbf{x}_i - \mathbf{c}\|_1. \quad (11b)$$

Multiplying (11b) by λ_i and summing over $i \in [k]$ gives

$$\min_{\mathbf{s} \in S_3^t} \|\mathbf{s} - \mathbf{c}\|_1 \leq \frac{d-t}{d} \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1 = \frac{d-t}{d} \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1. \quad (11c)$$

Let $\mathbf{s}^* \in \arg \min_{\mathbf{s} \in S_3^t} \|\mathbf{s} - \mathbf{c}\|_1$. Using the triangle inequality and (11c), we obtain

$$\begin{aligned} \min_{\mathbf{s} \in S_3^t} \sum_{i \in [k]} \lambda_i \|\mathbf{s} - \mathbf{x}_i\|_1 &\leq \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{s}^*\|_1 \leq \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1 + \|\mathbf{s}^* - \mathbf{c}\|_1 \\ &= \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1 + \min_{\mathbf{s} \in S_3^t} \|\mathbf{s} - \mathbf{c}\|_1 \leq \left(1 + \frac{d-t}{d}\right) \sum_{i \in [k]} \lambda_i \|\mathbf{x}_i - \mathbf{c}\|_1. \end{aligned}$$

Since this inequality holds for every tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k)$, the result follows from Theorem 1. $\square$

At $t = 0$, the candidate set is the union of the marginal supports, and the guarantee is a factor of 2. At $t = d$, the construction recovers the full coordinate-product set $\hat{S}_1$ and is exact. More generally, a $(1 + \alpha)$ guarantee follows whenever $t \geq \lceil d(1 - \alpha) \rceil$. The parameter t therefore provides an explicit tradeoff between candidate-set size and approximation quality.

7 Conclusion and future work

We developed a candidate-support reduction framework for discrete Wasserstein barycenters. For type-2 barycenters with general weights, the framework yields a randomized approximation scheme with an expected guarantee and a deterministic PTAS, both of which generalize the previously known factor-2 method. Under equal weights, a sampling procedure without replacement leads to a sharper approximation ratio. The same candidate-support principle yields a $(1 + 1/t)$ support-reduction guarantee for sparse type-2 barycenters and both exact and approximate coordinate-based constructions for type-1 barycenters. The numerical results show the efficiency of the proposed approximation algorithms. Important directions for future work include extending the approximation framework to type- p costs with $p \notin \{1, 2\}$, developing more scalable methods for solving the restricted MOT problems, and identifying other optimization problems for which tuplewise candidate-support reduction provides provable approximation guarantees.

Acknowledgments

This research was supported in part by National Science Foundation grant 2246414 and Office of Naval Research grant N00014-24-1-2066. The second author thanks Zedong Wang and Shuai Li for helpful discussions.

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* **49**.
- 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* **32**.
- 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), ‘Multi-subject MEG/EEG source imaging with sparse multi-task regression’, *NeuroImage* **220**, 116847.
- Janati, H., Cuturi, M. & Gramfort, A. (2020), 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* **32**.
- Sabo, K. & Scitovski, R. (2008), ‘The best least absolute deviations line-properties and two efficient methods for its derivation’, *The ANZIAM journal* **50**(2), 185–198.
- 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.
- Vazler, I., Sabo, K. & Scitovski, R. (2012), ‘Weighted median of the data in solving least absolute deviations problems’, *Communications in Statistics-Theory and Methods* **41**(8), 1455–1465.

- 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.