DIFFERENTIALLY PRIVATE WASSERSTEIN BARYCENTERS

Anonymous authors
Paper under double-blind review

ABSTRACT

The Wasserstein barycenter is defined as the mean of a set of probability measures under the optimal transport metric, and has numerous applications in machine learning. In practice these input measures are empirical distributions, often built from sensitive datasets. We present, to our knowledge, the first algorithms for computing Wasserstein barycenters under differential privacy. Empirically, on large-scale US population datasets, our methods produce high-quality private barycenters with strong accuracy-privacy tradeoffs. As a byproduct of our analysis we give a simple reduction with provable utility guarantees that can significantly accelerate the runtime for non-private barycenters, which may be of broader interest.

1 INTRODUCTION

In the era of big data and machine learning, users are increasingly concerned about their privacy. Differential privacy (DP) (Dwork et al., 2006) has seen widespread adoption to provide guarantees for user privacy. For example, government bureaus (Abowd, 2018; Hod & Canetti, 2024) and companies (Apple, 2017; Ding et al., 2017; Rogers et al., 2020) extensively employ DP when releasing data.

Clustering, summarizing and reducing the size of datasets are fundamental tasks in unsupervised machine learning. Many of these unsupervised learning problems are NP-hard (Megiddo & Supowit, 1984; Altschuler & Boix-Adserà, 2022), leading to the development of polynomial-time approximation algorithms (Charikar et al., 1999; Charikar & Guha, 1999; Jain & Vazirani, 2001; Jain et al., 2003; Charikar & Li, 2012; Cohen-Addad et al., 2022b). A long line of works (Gupta et al., 2010; Balcan et al., 2017; Kaplan & Stemmer, 2018; Jones et al., 2021; Chaturvedi et al., 2020; Ghazi et al., 2020; La Tour et al., 2024) have further studied clustering under DP, providing polynomial time algorithms that essentially match the non-private guarantees.

Defined as the mean of a set of probability measures under the optimal transport metric, the Wasserstein barycenter is a useful notion that contains many of these unsupervised tasks as special cases, with applications to a much more general suite of problems. Specific instances of the Wasserstein barycenter problem include centroids of probability measures (Zen & Ricci, 2011) and k-means clustering (Canas & Rosasco, 2012). Consequently, Wasserstein barycenters have seen extensive applications in domain adaptation (Montesuma & Mboula, 2021), computer graphics (Pele & Werman, 2009; Solomon et al., 2015), and biology (Nadeem et al., 2020; Heinemann et al., 2022).

Similar to clustering and other unsupervised learning problems mentioned above, privatization of Wasserstein barycenters becomes important when working with sensitive data. As one of many possible examples, suppose a company wishes to train and deploy machine learning models for analysis of sensitive data for each of many countries. Prior to actually training these models, to minimize the risk of privacy breaches, a machine learning engineer must design the model architecture and tune hyperparameters on a DP synthetic dataset (see Ridgeway et al. (2021); McKenna et al. (2021); Xie et al. (2018) for more details on this emerging practice), which by the post-processing property of DP can be used and re-used *ad infinitum* without incurring any additional privacy loss. A private Wasserstein barycenter of a (sub)set of country-level datasets would be a natural candidate for this private synthetic dataset, as (1) it averages across many countries and hence should incur much less privacy cost than maintaining separate private synthetic datasets for each country, and (2) it approximately minimizes the Wasserstein distances to the true distributions, which should maximize the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Results	Additive error
Theorem 1 ($\delta = 0$)	$\tilde{O}\left(\frac{\sqrt{d}}{(\varepsilon n)^{1/d}}\right)$
Theorem 2	$\tilde{O}_\delta\left(\frac{m\sqrt{d}}{\varepsilon nk} + \frac{m}{\varepsilon n}\right)$
Theorem 3	$\Omega\left(\frac{m^2 d}{(\varepsilon nk)^2}\right)$

Table 1: Our upper and lower bounds for (ε, δ)-DP algorithms for 2-Wasserstein barycenters with excess additive error due to privacy.

chance that the designed model architecture will work well when applied to each country's data at deployment time.

More classically, recall that the Wasserstein barycenter minimizes the (weighted) average transport cost from the barycenter to each marginal. This can be an interesting optimization problem in its own right, e.g. choosing locations for distribution centers for multiple products each with its own geographic demand distribution, where each center has the same mix of products.

Motivated by these considerations, our work answers the following question in the affirmative:

Do there exist efficient[1] *algorithms for computing Wasserstein barycenters under DP?*

Challenges and techniques It is not hard to see that gradient-based approaches to Wasserstein barycenters, e.g. Cuturi & Doucet (2014) can be privatized by DP-SGD (Abadi et al., 2016), but the objective function is non-convex so it is not clear how to obtain provable utility guarantees. Another potential approach is to consider privacy over the probability manifold, e.g. Reimherr et al. (2021) consider private Fréchet means, which is a special case of the Wasserstein barycenter. However, it is not clear how to obtain optimization guarantees for this strategy as the space of discrete probability measures is not geodesically convex. This suggests we may consider private preprocessing before running a Wasserstein barycenter algorithm or postprocessing the output barycenter for privacy. It turns out data preprocessing is sufficient.

Contributions To the best of our knowledge, we provide the first algorithms for computing Wasserstein barycenters under the constraints of DP (Dwork et al., 2006). Let k be the number of probability distributions, n be the number of datapoints in each distribution, m be the number of points in the output barycenter, and d be the dimension of each datapoint. We work under the setting where each individual contributes one datapoint to one distribution.

Our results are presented in Table 1, and our contributions are as follows.

- In Theorem 1 we provide an ε-DP algorithm based on He et al. (2023) that works by generating private distributions that are close in Wasserstein distance to the sensitive distributions, which we term coresets. While this method is general (works for all values of m), it suffers from the curse of dimensionality.
- In Theorem 2 we provide an (ε, δ)-DP algorithm based on a reduction to private clustering. The guarantees of this algorithm degrade gracefully in m. As a byproduct of our analysis, we provide speedups for non-private Wasserstein barycenters as well.
- In Theorem 3 we prove lower bounds on the excess cost any private 2-Wasserstein barycenter algorithm must incur using a reduction from private mean estimation.
- We validate the efficacy of our algorithms when applied to large-scale real-world sensitive data, e.g. US census data.

We discuss related work in Appendix A.

[1] Same (asymptotic) runtime as a non-private algorithm.

2 PRELIMINARIES

Notation In big-O notation, we use $\tilde{O}$ to hide logarithmic factors and subscripts to hide dependence in those variables. We use $a \lesssim_p b$ to denote $a = O_p(b)$, e.g. a constant that depends on p. We use $[t]$ to denote the set $\{1, \ldots, t\}$. For a function $T : \mathcal{X} \to \mathcal{Y}$ and a measure μ, we use $T_\sharp \mu$ to denote the pushforward measure, e.g. $T_\sharp \mu(B) = \mu(T^{-1}(B))$ for a measurable set $B \subseteq \mathcal{Y}$. For a datapoint x, we use δ_x to denote a Dirac delta at x. We reserve the Greek letter ξ to denote a failure probability. We use $B_x(R) := \{y \mid \|x - y\|_2 \leq R\}$ to denote the closed Euclidean ball of radius R centered at point x. **(OT)** We define $\mathcal{P}_p(\mathcal{X})$ to be probability measures over $\mathcal{X}$ with finite pth moments. We use Greek letters μ for the dataset and ν for barycenter, which are probability measures. **(DP)** We use the script font $\mathcal{A}$ and $\mathcal{D}$ to denote algorithms and datasets, respectively. We reserve ε, δ for privacy parameters for DP. **(JL)** We use letters d to represent the dimension of the *ambient* space and d' to represent the dimension of the *projected* space from the JL transform. We use the Greek letter γ for the multiplicative factor in the JL and Π for the projection matrix.

We introduce two definitions, deferring additional results that we use in Appendix B.

Definition 1 ((ε, δ)-DP)**.** *Algorithm $\mathcal{A}$ is said to satisfy (ε, δ)-differential privacy if for all adjacent datasets $\mathcal{D}, \mathcal{D}'$ (datasets differing in at most one element) and all $\mathcal{S} \subseteq \operatorname{range} \mathcal{A}$, it holds $\Pr[\mathcal{A}(\mathcal{D}) \in \mathcal{S}] \leq e^\varepsilon \Pr[\mathcal{A}(\mathcal{D}') \in \mathcal{S}] + \delta$. If $\delta = 0$, we say $\mathcal{A}$ satisfies ε-differential privacy.*

Definition 2 (Wasserstein distance)**.** *For $p, q \in [1, \infty)$, the (p, q)-Wasserstein distance between probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ is defined to be*

$$W_{p,q}(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int \|x - y\|_q^p d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \mid (P_x)_\sharp \pi = \mu, (P_y)_\sharp \pi = \nu\}$ is the set of transport plans, and $P_x(x, y) := x$ and $P_y(x, y) := y$ are the projections onto the first and second coordinates, respectively.

We will be using the Euclidean metric for the cost function, e.g. $q = 2$, which we simply denote $W_p := W_{p,2}$.

3 PROBLEM STATEMENT

Agueh & Carlier (2011) introduced the notion of barycenters on Wasserstein space:

Definition 3 (Wasserstein barycenter)**.** *Given probability distributions $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathcal{X})$, the p-Wasserstein barycenter is any distribution ν^* satisfying*

$$\nu^* \in \operatorname*{arg\,min}_{\nu \in \mathcal{P}(\mathcal{X})} F(\nu) := \frac{1}{k} \sum_{i \in [k]} W_p^p(\mu_i, \nu). \tag{1}$$

We will be working with *discrete distributions*, where each distribution can be thought of as a subpopulation, and one individual contributes sensitive data to one of the distributions. For instance each of these distributions could represent the data from one country.

Formally, we have k empirical distributions μ_i for $i \in [k]$, each with n point masses, where

$$\mu_i = \frac{1}{n} \sum_{j \in [n]} \delta_{x_{i,j}}, \tag{2}$$

Our goal is to compute a distribution ν consisting of exactly m point masses and uniform weights that minimizes the objective (1) under the constraints of DP.

For any application of DP, a definition of neighboring datasets is required. We use the following.

Definition 4 (Neighboring datasets)**.** *Let $\{\mu_i\}_{i \in [k]}$ be a collection of empirical distributions. Let the dataset be $\mathcal{D} := \{(x_{i,j}, i) \mid x_{i,j} \in \mu_i, i \in [k]\}$. We say two datasets $\mathcal{D}, \mathcal{D}'$ are neighboring if they differ by exactly one element in the first coordinate, e.g. they differ by one point in one distribution.*

This definition of neighboring datasets is motivated by viewing each μ_i as a non-overlapping subpopulation, i.e. we are essentially assuming that individuals do not appear in multiple of the μ_i. Without loss of generality, we assume all of the points are distinct. We will also abuse notation and identify $\mathcal{D}$ with $\{\mu_1, \ldots, \mu_k\}$.

In order to limit the influence of any individual, we require an assumption on the support. For simplicity, we assume a support contained in a ball of radius of 1/2.

Assumption 1. *It holds that* $\cup_{i\in[k]} \operatorname{supp} \mu_i \subseteq B_0(1/2)$.

This is a standard assumption in private clustering, e.g. see Ghazi et al. (2020), and more generally private convex optimization. This assumption is used to simplify the description of the results as it is well known that the additive error of any DP algorithm scales proportionally with respect to the radius of the support of the dataset, e.g. see Altschuler et al. (2024).

We constrain the optimization objective in (1) as follows.

Assumption 2. *Assume that the objective (1) has an added constraint that the solution has m equally weighted atoms,*[2] *where m is specified. Specifically, the solution satisfies $\nu = \frac{1}{m}\sum_{j\in[m]} \delta_{y_j}$, where $m \ll n$ and $y_j \in \mathbb{R}^d$ for $j \in [m]$.*

We make a few remarks.

- From an *interpretability* and *privacy* perspective, uniform weights is a reasonable assumption so that each datapoint can be considered as data representing one "person", so the output makes sense as a synthetic dataset.
- From a *computational* perspective, many papers on Wasserstein barycenters *a priori* solve the problem under the uniform weight assumption, as optimizing weights for barycenters is much more challenging than optimizing for supports, e.g. see the discussion in Altschuler & Boix-Adsera (2021).
- For both practical and modeling considerations, we should usually think of $m \ll n$, e.g. not growing to fast with the dataset.

We will use the following definitions.

Definition 5 (Cost)**.** *Following (1), we define* $\mathsf{cost}_{(\mu_1,\ldots,\mu_k)}(\nu) := \frac{1}{k}\sum_{i\in[k]} W_p^p(\mu_i, \nu)$.

Definition 6 (Approximate Wasserstein barycenter)**.** *Let* OPT *be the minimum value of Definition 5. A (z, t)-approximation for the p-Wasserstein barycenter is probability measure ν such that*

$$\mathsf{cost}_{(\mu_1,\ldots,\mu_k)}(\nu) \le z \cdot \mathsf{OPT}_{(\mu_1,\ldots,\mu_k)} + t,$$

where OPT *is the cost of an optimal barycenter supported on m atoms with uniform weights. When it's clear, we suppress the dependence on the inputs.*

We use this notion for most generality when stating our guarantees as (1) our methods work for any Wasserstein barycenter solvers, and (2) it gives a standardized way to state utility guarantees and measure the excess cost of privacy.

4 Private Coreset Approach

In this section, we use a black-box private coreset approach; some proofs are deferred to Appendix C. In particular, we consider private distributions that are close in Wasserstein distance to the sensitive distribution. We use these coresets as input to the approximate Wasserstein barycenter algorithm to obtain the private barycenter via post-processing.

We start by introducing the notion of coresets for Wasserstein distance.

Definition 7 (Coreset for Wasserstein distance)**.** *A measure μ' is a (p, t)-coreset of μ for the p-Wasserstein distance if for every $\pi \in \mathcal{P}(\mathbb{R}^d)$, we have $|W_p(\mu', \pi) - W_p(\mu, \pi)| \le t$.*

The following lemma is a direct consequence of the triangle inequality.

[2] Our algorithms and analysis does not require uniform weights, but we present it in this way.

Algorithm 1 $\mathsf{CoresetBarycenter}^{\varepsilon}$

Require: k discrete distributions $\mu_1, \dots, \mu_k$ supported on $\mathbb{R}^d$, projection dimension d', approximate Wasserstein barycenter algorithm Approx, privacy parameter ε

1: Sample a JL transform $\Pi \in \mathbb{R}^{d \times d'}$
2: **for** $i \in [k]$ **do**
3: $\quad \mu_i' \leftarrow \mathsf{WassersteinDistanceCoreset}(\mu_i)$ with privacy ε
4: $\quad \hat{\mu}_i \leftarrow \Pi_\sharp \mu_i'$
5: **end for**
6: $\hat{\nu} \leftarrow \mathsf{Approx}(\hat{\mu}_1, \dots, \hat{\mu}_k)$ ▷ $\hat{\nu} \in \mathcal{P}(\mathbb{R}^{d'})$
7: Reconstruct ν from $\hat{\nu}$ and $\mu_1', \dots, \mu_k'$ according to Izzo et al. (2021)
8: **return** ν

Lemma 1. *If $W_p(\mu, \mu') \le t$, then μ' is a (p, t)-coreset of μ for the p-Wasserstein distance problem.*

Now our goal is to find a coreset for the p-Wasserstein distance problem. We use the algorithm from He et al. (2023), which has the following guarantee.

Proposition 1. *For every $\varepsilon > 0$ and $\xi \in (0, 1)$, there exists an ε-DP algorithm running in time $\tilde{O}(dn)$ that with probability $1 - \xi$, outputs an*

$$\left(p, O_p\left(\left(\frac{\sqrt{d}}{(\varepsilon n)^{1/d}} \cdot \operatorname{poly}\log\left(\frac{n}{\xi}\right)\right)^{1/p}\right)\right)$$

-approximate coreset for the Wasserstein distance.

Proof (sketch). He et al. (2023) provides an algorithm with guarantees in expectation for the $W_{1,\infty}$ distance. They obtain guarantees in expectation by bounding the expectation of Laplace random variables. By concentration of Laplace random variables, we obtain a high-probability guarantee while only incurring polylog factors in the error. Finally, Fact 5 and $\|\cdot\|_2 \le \sqrt{d}\|\cdot\|_\infty$ yields the claimed guarantee for the W_p distance. □

We will use the following result to bound the error using Wasserstein distance coresets instead of the true distributions.

Lemma 2. *Let $\mu_1, \dots, \mu_k$ be discrete probability measures and suppose $\mu_1', \dots, \mu_k'$ are (p, t)-coresets for each μ_i, respectively. Then, for any $\gamma \in (0, 1)$*

$$\mathsf{OPT}_{(\mu_1', \dots, \mu_k')} \le (1 + \gamma)\mathsf{OPT}_{(\mu_1, \dots, \mu_k)} + O_{p,\gamma}(t^p),$$

and symmetrically with $(\mu_1', \dots, \mu_k')$ and $(\mu_1, \dots, \mu_k)$ interchanged.

Now we can show the main result in this section.

Theorem 1. *For any $p \ge 1$, suppose that there exists a (not necessarily private) (z, t)-approximation algorithm that runs in time $\operatorname{poly}(n, k, 2^{O(d)})$ for the p-Wasserstein barycenter problem. Then, for every $\varepsilon > 0$ and $\gamma, \xi \in (0, 1)$, there exists a polynomial-time ε-DP algorithm that outputs an*

$$\left(z(1 + \gamma), \tilde{O}_{p,\gamma,z,\xi}\left(\frac{\sqrt{d}}{(\varepsilon n)^{1/d}} + t\right)\right)$$

-approximate p-Wasserstein barycenter, with probability $1 - \xi$.

One issue of this method is if $d = \Omega(\log n)$, then the additive error becomes a constant. This is problematic as the cost of *any* Dirac delta in $B_0(1/2)$ is $O(1)$. Unfortunately, any coreset-based method will suffer from the curse of dimensionality, e.g. see Corollary 9.4, Boedihardjo et al. (2024). Further, in Section 6 we will see that this is inherent if $m \ge nk$.

5 Private Clustering Approach

In this section we provide an approach that reduces the computation of Wasserstein barycenters to the problem to private clustering. In particular, we will pool the sensitive datasets together and run a clustering algorithm on it; proofs are deferred to Appendix D. Then, we will privately release counts on the number of points sent to each cluster to form weighted inputs for a barycenter solver. We will see that this strategy will have good utility when $m \ll n$.

We recall some preliminaries for clustering.

Definition 8 ((q,p)-clustering)**.** *Given $m \in \mathbb{N}$ and a dataset $\mathcal{D} = \{x_1, \dots, x_n\}$ in $B_0(1/2)$, the goal is to find q centers $c_1, \dots, c_q \in \mathbb{R}^d$ that minimizes*

$$\mathsf{cost}_{\mathcal{D}}(c_1, \dots, c_q) := \sum_{i \in [n]} \min_{j \in [q]} \|x_i - c_j\|_2^p. \tag{3}$$

Definition 9 (Approximation for clustering)**.** *A (w,t)-approximation algorithm for (q,p)-clustering outputs $c_1, \dots, c_q$ such that $\mathsf{cost}(c_1, \dots, c_q) \le w \cdot \mathsf{OPT} + t$, where OPT is the optimal cost.*

We introduce some helpful notation. For a finite set $\mathcal{C} = \{c_1, \dots, c_m\} \subseteq B_0(1/2)$, define the snapping map

$$s_{\mathcal{C}}(x) \in \arg\min_{c \in \mathcal{C}} \|x - c\|_2, \tag{4}$$

with ties broken by a fixed deterministic rule and the normalized snapping cost

$$\Phi_{\mathcal{D}}(\mathcal{C}) := \frac{1}{nk} \sum_{i \in [k]} \sum_{j \in [n]} \|x_{i,j} - s_{\mathcal{C}}(x_{i,j})\|_2^p. \tag{5}$$

The key observation we use in this section is the following result that shows that the optimal Wasserstein barycenter cost upper bounds the optimal clustering cost. This connection allows us to use clustering as a preprocessing step before applying any Wasserstein barycenter solver, while still obtaining provable guarantees on the resulting barycenter objective.

Proposition 2. *Let $\mu_1, \dots, \mu_k$ be as in (2), and let $\mathsf{OPT}_{\mathsf{WB}}$ be the optimal cost for the Wasserstein barycenter problem as in Definition 6. For any $q \ge m$, let $\mathsf{OPT}^{(q)}_{\mathsf{cluster}}$ be the optimal (q,p)-clustering cost on the pooled dataset $\cup_{i \in [k]} \mu_i$ as in Definition 8. Then,*

$$\frac{1}{nk} \mathsf{OPT}^{(q)}_{\mathsf{cluster}} = \inf_{|\mathcal{C}| \le q} \Phi_{\mathcal{D}}(\mathcal{C}) \le \mathsf{OPT}_{\mathsf{WB}}.$$

Proof. Let $\nu^* = \frac{1}{m} \sum_{r \in [m]} \delta_{y_r}$ be an optimal barycenter, and let $\mathcal{C}^* := \{y_1, \dots, y_m\}$. Then it is clear from (5) that $\Phi_{\mathcal{D}}(\mathcal{C}^*) \le \mathsf{OPT}_{\mathsf{WB}}$, which implies the inequality. The equality follows from the fact that (3) and (5) are identical up to a normalization factor. □

Our algorithm is given in Algorithm 2 and satisfies the following guarantees, using the private clustering algorithm from Proposition 4.

Theorem 2. *For any $p \ge 1$, suppose that there exists a (not necessarily private) (z,t)-approximation algorithm that runs in polynomial-time for the p-Wasserstein barycenter problem. Then, for every $\varepsilon \le O(1)$ and $\delta, \xi \in (0,1)$, there exists a polynomial-time (ε, δ)-DP algorithm that outputs an*

$$\left(O_p(z), \tilde{O}_{p,z,\xi} \left(\frac{m\sqrt{d}}{\varepsilon nk} + \frac{m}{\varepsilon n} + t \right) \right)$$

-approximate p-Wasserstein barycenter, with probability $1 - \xi$.

Here, for our theoretical analysis, we take $q \leftarrow m$ in clustering, but in our experiments, we will often use a small multiple of m. Note that as in Section 4, we can utilize dimensionality reduction, which we omit in the statement and proofs for brevity. Also, by using a pure-DP clustering algorithm, we can obtain a pure-DP algorithm with worse-dependence on dimension, e.g. see Ghazi et al. (2020).

Algorithm 2 ClusteringBarycenter

Require: k discrete distributions $\mu_1, \dots, \mu_k$ supported on $\mathbb{R}^d$, approximate Wasserstein barycenter algorithm Approx, private clustering algorithm Cluster, privacy parameters ε, δ, failure probability ξ

$\mathcal{C} = \{c_1, \dots, c_m\} \leftarrow \mathsf{Cluster}(\cup_{i\in[k]}\mu_i)$ with privacy parameters $\varepsilon/2, \delta$ and failure probability $\xi/2$
Project $\mathcal{C}$ to $B_0(1/2)$
for $i \in [k]$ and $r \in [m]$ **do**
  $h_{i,r} \leftarrow |\{j \in [n] : s_{\mathcal{C}}(x_{i,j}) = c_r\}|$, with $s_{\mathcal{C}}$ defined as in (4)
  $Z_{i,r} \sim \mathsf{Lap}(4/\varepsilon)$ i.i.d.
end for
for $i \in [k]$ **do**
  $w_i \leftarrow (h_{i,1}, \dots, h_{i,m})/n$ and $\hat{w}_i \leftarrow w_i + (Z_{i,1}, \dots, Z_{i,m})/n$
  Choose $\tilde{w}_i \in \arg\min_{w\in\Delta_m} \|w - \hat{w}_i\|_1$, where $\Delta_m := \{w \in \mathbb{R}^m_{\geq 0} \mid \sum_{r\in[m]} w_r = 1\}$
  $\tilde{\mu}_i \leftarrow \sum_{r\in[m]} \tilde{w}_{i,r}\delta_{c_r}$
end for
$\nu \leftarrow \mathsf{Approx}(\tilde{\mu}_1, \dots, \tilde{\mu}_k)$
return ν

We also require the barycenter algorithm to accept weighted input measures as inputs. This can be done without loss of generality as this is private postprocessing: rational weights can be converted to uniform weights by increasing the input measure size and duplicating respective atoms accordingly. For uniqueness, we can perturb low-order bits.

We remark that when using non-private clustering, we also speed up computation of general Wasserstein barycenter solvers. The proof essentially the same, except without the privacy components, and thus is omitted. While these ideas have appeared in the literature, but as far as we are aware, we are the first to provide provable utility guarantees. In particular, Seguy & Cuturi (2015) (empirically) use clustering to speed up barycenter computation; Gachon et al. (2025) analyze quantization for Wasserstein barycenters, but their theoretical guarantees are asymptotic (e.g. see Proposition 4.1 therein) and require absolute continuity. Our guarantees are black-box and give directly give finite-sample guarantees.

Corollary 1. *Let $p \geq 1$ and $q \geq m$. Suppose there exists a (z,t)-approximation algorithm for the p-Wasserstein barycenter problem and an (w,s)-approximation algorithm for (q,p)-clustering, respectively running in $S(n,k,m)$ and $T(n)$ time. Then, there exists an algorithm that outputs an $(O_p(zw), O_{p,z}(\frac{s}{n}+t))$-approximate p-Wasserstein barycenter running in time $S(q,k,m) + kT(n)$ time.*

Different from the privacy case, the analysis uses a clustering algorithm for each distribution. In the statement of the results we suppress dependence on dimension and p.

Note that Izzo et al. (2021); Braverman et al. (2022) consider a strong notion of coreset that improve dependence on k. In particular, Izzo et al. (2021); Braverman et al. (2022) preserve the cost of every candidate barycenter, while our method only guarantees the output will have comparable utility to that of an optimal barycenter.

Remark 1. *Often, barycenter solvers are iterative, where each iteration takes $O(mnk)$. Our method would improve the runtime per iteration to $O(m^2k)$ for $m \leq n$, when $q = O(m)$.*

6 Lower Bound

In this section we provide a lower bound (for $p = 2$) on the additive error that any private $(1,t)$-approximation algorithm must incur via a reduction from private (empirical) mean estimation; proofs are deferred to Appendix E. Recall F as defined in (1) and OPT from Definition 6.

Theorem 3. *Let $d, k, n, m \in \mathbb{N}$, where $m \mid n$ and $2 \leq m \leq \exp(cd)$ for a sufficiently small universal constant $c > 0$. Consider any $(\varepsilon \leq 1, \delta = o(\frac{m}{nk}))$ DP algorithm that takes dataset $\mathcal{D}$ (Definition 4)*

over $B_0(1/2)$ and outputs a barycenter $\hat{\nu}$. Then,

$$\sup_{\mathcal{D}\in(B_0(1/2)^n)^k} \mathbb{E}[F(\hat{\nu}) - \mathsf{OPT}] = \Omega\left(\min\left\{1, \frac{m^2 d}{\varepsilon^2 (nk)^2}\right\}\right).$$

7 Experiments



(a) Cost $F(\nu)$. (b) Runtime in seconds.

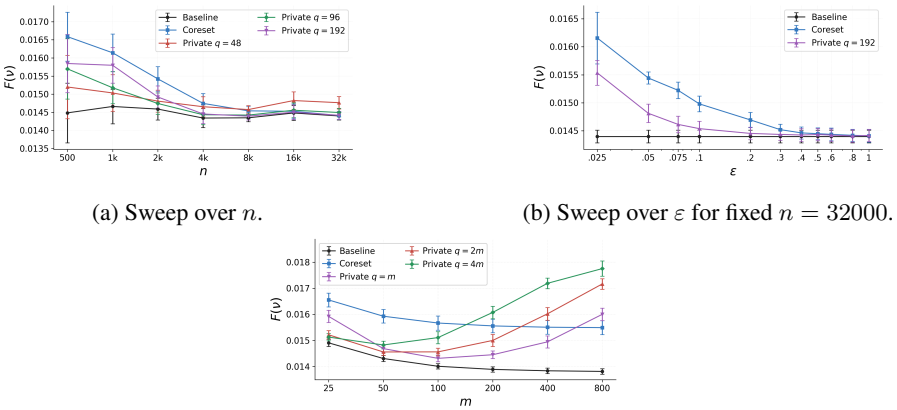
Figure 1: Exact Wasserstein barycenters on US census data.



(a) Cost $F(\nu)$. (b) Runtime in seconds.

Figure 2: Sinkhorn Wasserstein barycenters on US census data.

We test our methods on US census data and high-dimensional synthetic data. All of our experiments use the POT framework (Flamary et al., 2021). For our experiments under DP, unless otherwise stated we set $\varepsilon = 1$ and $\delta = \frac{1}{(nk)^2}$. Experiments are run on 10 independent trials, and we report results with 95% confidence intervals. For the coreset-based approach, we utilize weighted centers of the grids as described in He et al. (2023). Our experiments do not use the dimensionality reduction (as they are relatively low-dimensional). We usually use exact barycenter solver with 200 iterations, except in Figure 2, where we also use Sinkhorn barycenter solver with entropic regularization 0.01.



(a) Sweep over n.

(b) Sweep over ε for fixed $n = 32000$.

(c) Sweep over m for fixed $n = 8000$ and $\varepsilon = 0.2$.

Figure 3: Results on the US census data under DP.

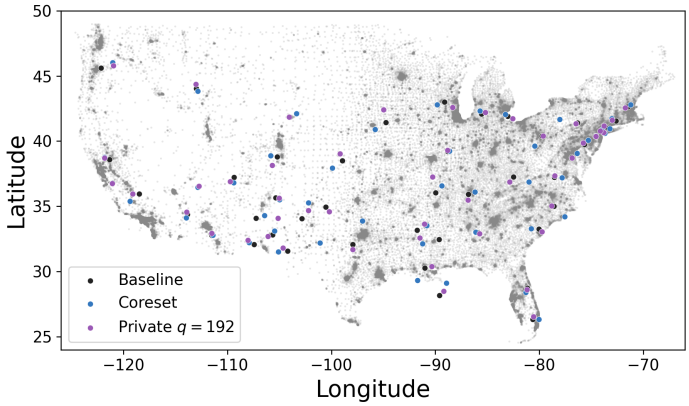


Figure 4: Example output with $n = 32000$ under $(\varepsilon = 1, \delta = \frac{1}{(nk)^2})$-DP.

In Figures 1–4, we use US census data from the 2015 American Community Survey.[3] We take the (sensitive) data to be multisets of the centers of census tracts (chosen with replacement) with $k = 4$, where the distributions corresponding to the self-reported racial groups White, Asian, Black, Hispanic. We use $m = 48$ and initialize the solver at the 48 capitols; this initialization has costs 0.0203.

In Figures 1 and 2, we show the effectiveness of our (non-private) clustering algorithms on the US census data. We consider varying number of clustering points $q \in \{6, 12, 24, 48\}$ over a range of n. We observe that for relatively clustered datasets like the US population, even $q < m$ can suffice; we recall that without the constraint of privacy, we run a clustering algorithm for each of the k distributions. On moderate size inputs $n = 4000$, our methods become an order of magnitude faster. In Figure 3, we now consider the experiments under DP, varying n, ε, and m. For the coreset-based method, we only include points that lie within the US borders (which is private by post-processing). We observe that the coreset-based approach becomes competitive with the clustering method at $n = 16000$ when $\varepsilon = 1$ (Figure 3a). In Figure 3c, we use initialization to be sampled uniformly from within the continental US borders. As Theorem 2 predicts, the guarantees of the cluster-based approach deteriorates as m increases. In Figure 4, we plot an example run for the US census data. We observe that both private barycenters put a point not too far from the non-private barycenter.

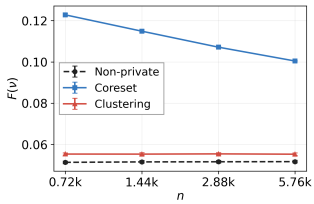


Figure 5: Results on $d = 25$-dimensional synthetic dataset. An example run with $n = 5760$ has $F(\delta_0) = 0.06334$.

Because the US census data are 2-dimensional, the coreset-based approach is competitive with the cluster-based approach. As such, in Figure 5 we consider results for a higher dimensional synthetic data. We choose $d = 25$, $m = 36$, $q = 4m$, $k = 20$ over a range of n. The data is generated as follows: for each of the distributions, we choose m' uniformly from the divisors of m, which is the number of clusters in this distribution. We put each cluster center p towards a random direction, and draw its radius uniformly from $(0, 0.2)$. Finally, we add n/m' points drawn from $\mathcal{N}(p, \frac{0.01}{\sqrt{d}} I_d)$ to the distribution. We observe that the cluster-based approach is relatively effective, while the coreset-based approach does not beat the $(0, 0)$-DP baseline δ_0.

[3] https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2015/

AI USE STATEMENT

In this work, we used ChatGPT 5.4, 5.5, and 5.6 Pro to explore proof strategies for Theorem 2, check correctness of mathematical arguments, and implement experiments. We also used ChatGPT 5.6 Pro to conduct literature search subsequent to the development of our results in Section 4, which helped us find the related works (Seguy & Cuturi, 2015; Gachon et al., 2025). The main ideas were developed by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS'16. ACM, October 2016. doi: 10.1145/2976749.2978318. URL http://dx.doi.org/10.1145/2976749.2978318.

John M Abowd. The US census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2867–2867, 2018.

Pankaj K. Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. Fast and accurate approximations of the optimal transport in semi-discrete and discrete settings. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 4514–4529, 2024. doi: 10.1137/1.9781611977912.159. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611977912.159.

Pankaj K. Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. Efficient approximation algorithm for computing Wasserstein barycenter under Euclidean metric. Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2025. doi: 10.1137/1.9781611978322.163. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611978322.163.

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

Noga Alon, Phillip B Gibbons, Yossi Matias, and Mario Szegedy. Tracking join and self-join sizes in limited storage. In *Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 10–20, 1999.

Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration, 2018. URL https://arxiv.org/abs/1705.09634.

Jason M Altschuler and Enric Boix-Adsera. Wasserstein barycenters can be computed in polynomial time in fixed dimension. *Journal of Machine Learning Research*, 22(44):1–19, 2021. URL http://jmlr.org/papers/v22/20-588.html.

Jason M. Altschuler and Enric Boix-Adserà. Wasserstein barycenters are NP-hard to compute. volume 4, pp. 179–203, 2022. doi: 10.1137/21M1390062. URL https://doi.org/10.1137/21M1390062.

Jason M. Altschuler, Jinho Bok, and Kunal Talwar. On the privacy of noisy stochastic gradient descent for convex optimization. *SIAM Journal on Computing*, 53(4):969–1001, 2024. doi: 10.1137/23M1556538. URL https://doi.org/10.1137/23M1556538.

Apple. Learning with privacy at scale. 2017. URL https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

Maria-Florina Balcan, Travis Dick, Yingyu Liang, Wenlong Mou, and Hongyang Zhang. Differentially private clustering in high-dimensional Euclidean spaces. In *International Conference on Machine Learning*, pp. 322–331. PMLR, 2017.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th annual symposium on foundations of computer science*, pp. 464–473. IEEE, 2014.

Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. volume 37, pp. A1111–A1138, 2015. doi: 10.1137/141000439. URL https://doi.org/10.1137/141000439.

Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss transform itself preserves differential privacy, 2012. URL https://arxiv.org/abs/1204.2136.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Private measures, random walks, and synthetic data. *Probability theory and related fields*, pp. 1–43, 2024.

Vladimir Braverman, Vincent Cohen-Addad, H-C Shaofeng Jiang, Robert Krauthgamer, Chris Schwiegelshohn, Mads Bech Toftrup, and Xuan Wu. The power of uniform sampling for coresets. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 462–473. IEEE, 2022.

Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 1–10, 2014.

Guillermo Canas and Lorenzo Rosasco. Learning probability measures with respect to optimal transport metrics. *Advances in Neural Information Processing Systems*, 25, 2012.

Moses Charikar and Sudipto Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pp. 378–388. IEEE, 1999.

Moses Charikar and Shi Li. A dependent LP-rounding approach for the k-median problem. In *International Colloquium on Automata, Languages, and Programming*, pp. 194–205. Springer, 2012.

Moses Charikar and Erik Waingarten. The johnson-lindenstrauss lemma for clustering and subspace approximation: From coresets to dimension reduction. pp. 3172–3209. Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, 2025. doi: 10.1137/1.9781611978322.102. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611978322.102.

Moses Charikar, Sudipto Guha, Éva Tardos, and David B Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pp. 1–10, 1999.

Anamay Chaturvedi, Huy Nguyen, and Eric Xu. Differentially private k-means clustering via exponential mechanism and max cover. *arXiv preprint arXiv:2009.01220*, 2020.

Sebastian Claici, Edward Chien, and Justin Solomon. Stochastic Wasserstein barycenters, 2018. URL https://arxiv.org/abs/1802.05757.

Vincent Cohen-Addad, Hossein Esfandiari, Vahab Mirrokni, and Shyam Narayanan. Improved approximations for Euclidean k-means and k-median, via nested quasi-independent sets. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1621–1628, 2022a.

Vincent Cohen-Addad, Anupam Gupta, Lunjia Hu, Hoon Oh, and David Saulpic. An improved local search algorithm for k-median. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1556–1612. SIAM, 2022b.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. In Eric P. Xing and Tony Jebara (eds.), *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 685–693, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/cuturi14.html.

Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014. ISSN 1551-305X. doi: 10.1561/0400000042. URL http://dx.doi.org/10.1561/0400000042.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3*, pp. 265–284. Springer, 2006.

Vitaly Feldman, Audra McMillan, Satchit Sivakumar, and Kunal Talwar. Instance-optimal private density estimation in the wasserstein distance. *Advances in Neural Information Processing Systems*, 37:90061–90131, 2024.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78): 1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Erell Gachon, Elsa Cazelles, and Jérémie Bigot. Scalable learning from probability measures with mean measure quantization. *arXiv preprint arXiv:2502.04907*, 2025.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Differentially private clustering: Tight approximation ratios. *Advances in Neural Information Processing Systems*, 33:4040–4054, 2020.

Kristjan Greenewald, Anming Gu, Mikhail Yurochkin, Justin Solomon, and Edward Chien. k-mixup regularization for deep learning via optimal transport. *arXiv preprint arXiv:2106.02933*, 2021.

Anupam Gupta, Katrina Ligett, Frank McSherry, Aaron Roth, and Kunal Talwar. Differentially private combinatorial optimization. In *Proceedings of the 2010 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1106–1125, 2010. doi: 10.1137/1.9781611973075.90. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611973075.90.

Yiyun He, Roman Vershynin, and Yizhe Zhu. Algorithmically effective differentially private synthetic data. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 3941–3968. PMLR, 2023.

Florian Heinemann, Axel Munk, and Yoav Zemel. Randomized wasserstein barycenter computation: Resampling with statistical guarantees. volume 4, pp. 229–259, 2022. doi: 10.1137/20M1385263. URL https://doi.org/10.1137/20M1385263.

Shlomi Hod and Ran Canetti. Differentially private release of Israel's national registry of live births. *arXiv preprint arXiv:2405.00267*, 2024.

Zachary Izzo, Sandeep Silwal, and Samson Zhou. Dimensionality reduction for Wasserstein barycenter, 2021. URL https://arxiv.org/abs/2110.08991.

Kamal Jain and Vijay V Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. *Journal of the ACM (JACM)*, 48(2):274–296, 2001.

Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM (JACM)*, 50(6):795–824, 2003.

William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into Hilbert space. *Contemporary mathematics*, 26:189–206, 1984.

Matthew Jones, Huy L Nguyen, and Thy D Nguyen. Differentially private clustering via maximum coverage. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11555–11563, 2021.

Haim Kaplan and Uri Stemmer. Differentially private k-means with constant multiplicative error. *arXiv preprint arXiv:1804.08001*, 2018.

Max Dupré La Tour, Monika Henzinger, and David Saulpic. Making old things new: a unified algorithm for differentially private clustering. *arXiv preprint arXiv:2406.11649*, 2024.

Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2): 129–137, 1982.

Konstantin Makarychev, Yury Makarychev, and Ilya Razenshteyn. Performance of Johnson-Lindenstrauss transform for k-means and k-medians clustering. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1027–1038, 2019.

Ryan McKenna, Gerome Miklau, and Daniel Sheldon. Winning the nist contest: A scalable and general approach to differentially private synthetic data. *arXiv preprint arXiv:2108.04978*, 2021.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pp. 94–103, 2007. doi: 10.1109/FOCS.2007.66.

Nimrod Megiddo and Kenneth J Supowit. On the complexity of some common geometric location problems. *SIAM Journal on Computing*, 13(1):182–196, 1984.

Eduardo Fernandes Montesuma and Fred Maurice Ngole Mboula. Wasserstein barycenter for multi-source domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16785–16793, 2021.

Shanmugavelayutham Muthukrishnan et al. Data streams: Algorithms and applications. *Foundations and Trends® in Theoretical Computer Science*, 1(2):117–236, 2005.

Saad Nadeem, Travis Hollmann, and Allen Tannenbaum. Multimarginal Wasserstein barycenter for stain normalization and augmentation. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz (eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, pp. 362–371, Cham, 2020. Springer International Publishing. ISBN 978-3-030-59722-1.

Shyam Narayanan, Sandeep Silwal, Piotr Indyk, and Or Zamir. Randomized dimensionality reduction for facility location and single-linkage clustering. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7948–7957. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/narayanan21b.html.

Aleksandar Nikolov. Private query release via the Johnson-Lindenstrauss transform, 2022. URL https://arxiv.org/abs/2208.07410.

Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th international conference on computer vision*, pp. 460–467. IEEE, 2009.

Matthew Reimherr, Karthik Bharath, and Carlos Soto. Differential privacy over riemannian manifolds. *Advances in Neural Information Processing Systems*, 34:12292–12303, 2021.

Diane Ridgeway, Diane Ridgeway, Mary F Theofanos, Terese W Manley, and Christine Task. *Challenge design and lessons learned from the 2018 differential privacy challenges*. US Department of Commerce, National Institute of Standards and Technology, 2021.

Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, and Parvez Ahammad. Linkedin's audience engagements API: A privacy preserving data analytics system at scale. *arXiv preprint arXiv:2002.05839*, 2020.

Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

Vivien Seguy and Marco Cuturi. Principal geodesic analysis for probability measures under the optimal transport metric. *Advances in Neural Information Processing Systems*, 28, 2015.

Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015.

Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially private generative adversarial network. *arXiv preprint arXiv:1802.06739*, 2018.

Gloria Zen and Elisa Ricci. Earth mover's prototypes: A convex learning approach for discovering activity patterns in dynamic scenes. In *CVPR 2011*, pp. 3225–3232. IEEE, 2011.

A Related Work

In the theoretical computer science community, the Wasserstein barycenter falls under the category of facility location problems. This class of problem is concerned with placing points, or "facilities," to minimize some objective given a set of input data. Note that clustering also falls under this category. Clustering (Lloyd, 1982) has seen many non-private approximation algorithms. Over the past few decades, a line of works (Charikar et al., 1999; Charikar & Guha, 1999; Jain & Vazirani, 2001; Jain et al., 2003; Charikar & Li, 2012; Cohen-Addad et al., 2022b) have pushed multiplicative approximation factors to 2.406 and 5.912 for Euclidean k-medians and k-means, respectively (Cohen-Addad et al., 2022a).

Gupta et al. (2010) initiated the study of facility location algorithms under DP, and provided an inefficient algorithm based on the exponential mechanism (McSherry & Talwar, 2007) that gave constant factor multiplicative approximation. Then a series of works (Balcan et al., 2017; Kaplan & Stemmer, 2018; Jones et al., 2021; Chaturvedi et al., 2020; Ghazi et al., 2020; La Tour et al., 2024) have achieved polynomial time algorithms for private clustering with the optimal multiplicative approximation ratio and small additive errors.

On the other hand, the Wasserstein barycenter is a much more nascent problem. Initial works provide approximations using methods such as entropic regularization (Cuturi, 2013; Cuturi & Doucet, 2014), iterative Bregman projections (Benamou et al., 2015), or stochastic optimization (Claici et al., 2018); however, these lack worst-case guarantees on the approximations, for instance to the non-entropic setting. Even theoretical guarantees for fast approximations of Wasserstein distances are recent (Altschuler et al., 2018; Agarwal et al., 2024). More recently, some works have provided theoretical guarantees for Wasserstein barycenters in the $p = 2$ setting: Altschuler & Boix-Adsera (2021); Agarwal et al. (2025) showed that additive and multiplicative (respectively) approximations for Wasserstein barycenters can be computed in polynomial time for *constant* dimension. Recently, Boedihardjo et al. (2024); He et al. (2023) provided constructions for private measures that are close to input empirical measures over $[0, 1]^d$ in 1-Wasserstein distance and Feldman et al. (2024) provided instance-optimal constructions for finite metric spaces.

The Johnson-Lindenstrauss (JL) lemma (Johnson & Lindenstrauss, 1984) is a dimensionality reduction method that provides worst-case guarantees on preserving pairwise distances between a collection of points. It has been applied to numerous problems in many areas of computer science, including streaming algorithms (Alon et al., 1999; Muthukrishnan et al., 2005) and DP (Blocki et al., 2012; Nikolov, 2022). The (lower) bound on the dimension required to approximately preserve solutions varies from problem to problem, e.g. see Narayanan et al. (2021); Charikar & Waingarten (2025) for a discussion. For facility location problems, Makarychev et al. (2019) showed that dimension $d' = O(\log k)$ suffices for preserving the cost of solutions to k-means clustering, and Izzo et al. (2021) showed that dimension $d' = O(\log n)$ suffices for Wasserstein barycenters supported on $\leq n$ points.

B Additional Results

B.1 Differential Privacy

Fact 1 (Sequential composition). *Let $\mathcal{A}_1, \mathcal{A}_2$ be $(\varepsilon_i, \delta_i)$-DP algorithms for $i = 1, 2$. Then, the output $(\mathcal{A}_1(\mathcal{D}), \mathcal{A}_2(\mathcal{D}))$ is $(\varepsilon_1 + \varepsilon_2, \delta_1 + \delta_2)$-DP, where $\mathcal{A}_2$ is allowed to depend on the output of $\mathcal{A}_1$.*

Fact 2 (Parallel composition). *Let $\mathcal{A}_1, \ldots, \mathcal{A}_k$ be ε-DP algorithms. Suppose $\mathcal{D} = S_1 \cup \cdots \cup S_k$, where $S_i \cap S_j = \emptyset$ for every $i \neq j$. Then $(\mathcal{A}_1(S_1), \ldots, \mathcal{A}_k(S_k))$ is ε-DP.*

A nice property of differential privacy is the post-processing property, which informally says that transforming private output does not incur additional privacy loss. Formally, we have the following:

Fact 3 (Post-processing). *Let $\mathcal{A}$ be an ε-DP algorithm. Then for any (possibly randomized algorithm) g, $g \circ \mathcal{A}(\mathcal{D})$ is ε-DP.*

Definition 10 (ℓ_p-sensitivity). *We define the ℓ_p-sensitivity of a function f to be*

$$\Delta_p f := \max_{\mathcal{D}, \mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_p,$$

where $\mathcal{D}, \mathcal{D}'$ are adjacent datasets.

B.2 Optimal Transport

Fact 4 (Lemma 5.4, Santambrogio (2015)). *For any $p \geq 1, \mu, \nu, \pi \in \mathcal{P}_p(\mathcal{X})$, we have $W_p(\mu, \pi) \leq W_p(\mu, \nu) + W_p(\nu, \pi)$.*

For bounded spaces, we can bound the p-Wasserstein distance by the 1-Wasserstein distance:

Fact 5 (Equation 5.1, Santambrogio (2015)). *Let $\mathcal{X}$ be bounded. Then for any $p \geq 1, \mu, \nu \in \mathcal{P}_p(\mathcal{X})$*

$$W_p(\mu, \nu) \leq \operatorname{diam}(\mathcal{X})^{(p-1)/p} W_1(\mu, \nu)^{1/p}.$$

Definition 11. *The total variation distance between two discrete measures μ, ν is defined to be*

$$\|\mu - \nu\|_{\mathsf{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|.$$

Lemma 3. *Let $(\mathcal{X}, \rho)$ be a metric space with diameter D. For any $p \geq 1$ and probability measures $\mu, \nu \in \mathcal{P}(\mathcal{X})$, it holds*

$$W_p(\mu, \nu) \leq D \|\mu - \nu\|_{\mathsf{TV}}^{1/p}.$$

Proof. Let $\omega = \mu \wedge \nu$, and rewrite $\mu = \omega + \alpha$ and $\nu = \omega + \beta$ for $\alpha, \beta \geq 0$ and $\alpha(\mathcal{X}) = \beta(\mathcal{X}) = t = \|\mu - \nu\|_{\mathsf{TV}}$. Now we construct a coupling $\pi \in \Pi(\mu, \nu)$ as follows. We couple μ and ν by coupling ω to itself and α to β with any plan $\gamma \in \Pi(\alpha, \beta)$. This yields

$$\int \rho(x, y)^p d\pi(x, y) \leq \int \rho(x, y)^p d\gamma(x, y) \leq D^p \|\mu - \nu\|_{\mathsf{TV}},$$

where we use $\rho(x, y) \leq D$. Taking the infimum over couplings and taking the pth root yields the desired result. □

B.3 Additional Lemma

Lemma 4. *For every $p \geq 1$, $\gamma \in (0, 1)$, and $a, b \geq 0$,*

$$(a + b)^p \leq (1 + \gamma) a^p + O_{p,\gamma}(b^p).$$

Proof. The case $p = 1$ is immediate. Suppose $p > 1$. For any $\theta \in (0, 1)$, convexity of $x \mapsto x^p$ gives

$$\begin{aligned}(a + b)^p &= \left((1 - \theta) \frac{a}{1 - \theta} + \theta \frac{b}{\theta}\right)^p \\ &\leq (1 - \theta)^{1-p} a^p + \theta^{1-p} b^p.\end{aligned}$$

Taking $\theta = 1 - (1 + \gamma)^{-1/(p-1)}$ gives $(1 - \theta)^{1-p} = 1 + \gamma$, so $\theta^{1-p} = O_{p,\gamma}(1)$. This proves the claim. □

C Deferred Proofs from Section 4

C.1 Proof of Lemma 2

Proof. Consider a candidate barycenter ν. Using (1), we have

$$\begin{aligned}
\mathsf{cost}_{\mu'_1,\dots,\mu'_k}(\nu) &= \frac{1}{k}\sum_{i\in[k]} W_p^p(\mu'_i,\nu) \\
&\le \frac{1}{k}\sum_{i\in[k]} (W_p(\mu'_i,\mu_i)+W_p(\mu_i,\nu))^p && (6)\\
&\le \frac{1+\gamma}{k}\sum_{i\in[k]} W_p^p(\mu_i,\nu) + \frac{O_{p,\gamma}(1)}{k}\sum_{i\in[k]} W_p^p(\mu'_i,\mu_i) && (7)\\
&= (1+\gamma)\mathsf{cost}_{\mu_1,\dots,\mu_k}(\nu) + \frac{O_{p,\gamma}(1)}{k}\sum_{i\in[k]} W_p^p(\mu'_i,\mu_i) && (8)\\
&\le (1+\gamma)\mathsf{cost}_{\mu_1,\dots,\mu_k}(\nu) + \frac{O_{p,\gamma}(1)}{k}\sum_{i\in[k]} t^p, && (9)\\
&= (1+\gamma)\mathsf{cost}_{\mu_1,\dots,\mu_k}(\nu) + O_{p,\gamma}(t^p)
\end{aligned}$$

where (6) follows from the triangle inequality, (7) follows from Lemma 4, (8) follows from (1), and (9) follows from Definition 7. The other bound follows symmetrically. □

C.2 Proof of Theorem 1

In the proof, we require the following result.

Proposition 3 (Theorem 4.1, Izzo et al. (2021))**.** *Let $\mu_1,\dots,\mu_k$ be discrete probability distributions on $\mathbb{R}^d$ such that $|\operatorname{supp}\mu_i| \le \operatorname{poly}(n)$ for all $i\in[k]$. Let $d\ge 1$, $\gamma,\xi\in(0,1)$. Let $\Pi:\mathbb{R}^d\to\mathbb{R}^{d'}$ be an i.i.d. Gaussian JL map[4] with $d' = O\left(\frac{p^4}{\gamma^2}\log\frac{n}{\gamma\xi}\right)$ and $\hat\nu$ be a barycenter over $\mathcal{P}(\mathbb{R}^{d'})$. Then, with probability $1-\xi$, it holds that*

$$\mathsf{cost}_{(\mu_1,\dots,\mu_k)}(\nu) \approx_{1+\gamma} \mathsf{cost}_{(\Pi_\sharp\mu_1,\dots,\Pi_\sharp\mu_k)}(\hat\nu),$$

where $\nu\in\mathcal{P}(\mathbb{R}^d)$ is reconstructed from $\hat\nu$.

Above, for $\gamma\ge 0$, we use $a\approx_{1+\gamma} b$ to denote $\frac{1}{1+\gamma}\le\frac{a}{b}\le 1+\gamma$.

Proof (of Theorem 1). Let d' be as in Proposition 3.

(Runtime) It suffices to bound the runtime of computing the barycenter in low dimensions as it is clear that the pre- and post-processing steps run in polynomial time. With the given d', we have that $2^{O(d')}\cdot\operatorname{poly}(n,k) = \operatorname{poly}(n)\cdot\operatorname{poly}(n,k) = \operatorname{poly}(n,k)$, as desired.

(Privacy) The privacy follows from Facts 2 and 3 as the input distributions are disjoint.

(Utility) Define $\alpha := \frac{\sqrt{d}}{(\varepsilon n)^{1/d}}\operatorname{poly}\log\frac{nk}{\xi}$. For each μ_i, we invoke Fact 1 with failure probability $\xi/(2k)$. Then by a union bound, with probability $1-\xi/2$, μ'_i is a $\left(p, O_p(\alpha^{1/p})\right)$-coreset for μ_i, for each $i\in[k]$. Assuming this event holds, Lemma 2 with $\gamma\leftarrow\gamma/10$ implies

$$\mathsf{OPT}_{(\mu'_1,\dots,\mu'_k)} \le (1+\gamma/10)\mathsf{OPT}_{(\mu_1,\dots,\mu_k)} + O_{p,\gamma}(\alpha). \tag{10}$$

Let ν be the output of the algorithm. Now we apply Fact 3 with $\gamma\leftarrow\gamma/10$, along with the guarantee of the not necessarily private approximation algorithm, which implies with probability $1-\xi/2$,

$$\begin{aligned}
\mathsf{cost}_{(\mu'_1,\dots,\mu'_k)}(\nu) &\le (1+\gamma/10)\mathsf{cost}_{(\Pi_\sharp\mu'_1,\dots,\Pi_\sharp\mu'_k)}(\hat\nu) \\
&\le z(1+\gamma/10)\mathsf{OPT}_{(\Pi_\sharp\mu'_1,\dots,\Pi_\sharp\mu'_k)} + O_{p,\gamma,z}(t) && (11)\\
&\le z(1+\gamma/10)^2\mathsf{OPT}_{(\mu'_1,\dots,\mu'_k)} + O_{p,\gamma,z}(t).
\end{aligned}$$

[4] It is also possible to use other JL maps, e.g. the fast JL transform. For details, see Appendix B, Izzo et al. (2021).

By a union bound, (10) and (11) both occur with probability $1-\xi$. When this is the case, we deduce

$$\begin{aligned}\mathsf{cost}_{(\mu_1,\dots,\mu_k)}(\nu) &\le (1+\gamma/10)\mathsf{cost}_{(\mu'_1,\dots,\mu'_k)}(\nu) + O_{p,\gamma,z}(\alpha)\\ &\le z(1+\gamma/10)^3\mathsf{OPT}_{(\mu'_1,\dots,\mu'_k)} + O_{p,\gamma,z}(\alpha+t)\\ &\le z(1+\gamma)\mathsf{OPT}_{(\mu_1,\dots,\mu_k)} + O_{p,\gamma,z}(\alpha+t)\end{aligned}$$

where we use Lemma 2 and $(1+\gamma/10)^4 \le 1+\gamma$ for $\gamma \in (0,1)$. □

D Deferred Proofs from Section 5

The private clustering algorithm we use provides the following guarantee:

Proposition 4 (Lemma F.1, La Tour et al. (2024)). *Let $p \ge 1$, $\varepsilon \le O(1)$, and $\delta, \xi < 1$. There exists an (ε,δ)-DP algorithm that runs in $q^{O(1)} n \log^2 n$ time, and with probability $1-\xi$, outputs an*

$$\left(O_p(1), O_p\left(\frac{q\sqrt{d}}{\varepsilon}\mathrm{poly}\log\left(\frac{n}{\delta\xi}\right)\right)\right)$$

-approximation for (q,p)-clustering.

Proof of Theorem 2. (Runtime) The runtime is immediate by assumption and the guarantees for Fact 4.

(Privacy) By Proposition 4 and line 1 in the algorithm, releasing the candidate set $\mathcal{C}$ is $(\varepsilon/2,\delta)$-DP. Conditioned on this set $\mathcal{C}$, changing one point in an adjacent dataset changes at most two coordinates of $(h_{i,r})_{i\in[k],r\in[m]}$. Thus, the ℓ_1 sensitivity of the counts is 2, so adding $\mathsf{Lap}(4/\varepsilon)$ noise to each coordinate is $\varepsilon/2$-DP by the Laplace mechanism (Theorem 3.6, Dwork & Roth (2014)). Then the privacy is immediate by composition and post-processing (Facts 1 and 3).

(Utility) The utility proof will utilize Proposition 2. Condition on the success of the clustering algorithm, which holds with probability $1-\xi/2$. For each $i \in [k]$, denote the snapped measure

$$\mu_i^{\mathcal{C}} := (s_{\mathcal{C}})_\sharp \mu_i = \sum_{r\in[m]} w_{i,r}\delta_{c_r}$$

supported on $\mathcal{C}$ with weights $w_{i,r}$. We observe that

$$W_p^p(\mu_i,\mu_i^{\mathcal{C}}) \le \frac{1}{n}\sum_{j\in[n]} \|x_{i,j} - s_{\mathcal{C}}(x_{i,j})\|_2^p$$

as we can take the coupling that sends $x_{i,j}$ to $s_{\mathcal{C}}(x_{i,j})$. Next, we will have

$$\|\tilde{w}_i - w_i\|_1 \le \|\tilde{w}_i - \hat{w}_i\|_1 + \|\hat{w}_i - w_i\|_1 \le 2\|\hat{w}_i - w_i\|_1 = \frac{2}{n}\sum_{r\in[m]}|Z_{i,r}|,$$

where the second inequality uses the fact that $\tilde{w}_i$ is an ℓ_1-projection of $\hat{w}_i$ onto Δ_m. Then,

$$W_p^p(\mu_i^{\mathcal{C}},\tilde{\mu}_i) \le \|\mu_i^{\mathcal{C}} - \tilde{\mu}_i\|_{\mathsf{TV}} \le \frac{1}{2}\|w_i - \tilde{w}_i\|_1 \le \frac{1}{n}\sum_{r\in[m]}|Z_{i,r}|,$$

using Lemma 3 and definition of TV distance. Combining the above displays, the triangle inequality, and $(a+b)^p \le 2^{p-1}(a^p+b^p)$ yields

$$\frac{1}{k}\sum_{i\in[k]} W_p^p(\mu_i,\tilde{\mu}_i) \lesssim_p \Phi_{\mathcal{D}}(\mathcal{C}) + \frac{1}{kn}\sum_{i\in[k]}\sum_{r\in[m]}|Z_{i,r}|.$$

A union bound over Laplace tails shows that, with probability at least $1-\xi/2$,

$$|Z_{i,r}| \le \frac{4}{\varepsilon}\log\left(\frac{2km}{\xi}\right)$$

for every $i \in [k]$, $r \in [m]$. Conditioning on this event and the success of the clustering algorithm, which holds with probability $1 - \xi$, combining the above displays and using Proposition 2,

$$\frac{1}{k}\sum_{i\in[k]} W_p^p(\mu_i, \tilde{\mu}_i) \lesssim_p O_p(\mathsf{OPT}_{\mathsf{WB}}) + \tilde{O}_{p,z,\xi}\left(\frac{m\sqrt{d}}{\varepsilon nk} + \frac{m}{\varepsilon n}\right).$$

The claimed bound now follows analogously from the proof of Lemma 2 as applied to Theorem 1. □

E Deferred Proofs from Section 6

We use the following two results.

Proposition 5 (Appendix C.2, Bassily et al. (2014) and Theorem 3.5, Bun et al. (2014)). *Let $n, d \in \mathbb{N}$, $\varepsilon \le 1$, and $\delta = o(\frac{1}{n})$. There exists a distribution[5] $\mathcal{Q}$ over $\mathcal{D} = \{x_1, \dots, x_n\} \subseteq \{\pm 1/\sqrt{d}\}^d$ such that for every (ε, δ)-DP mean estimation algorithm $\mathcal{A}$,*

$$\mathbb{E}_{\mathcal{D}\sim\mathcal{Q},\mathcal{A}} \|\mathcal{A}(\mathcal{D}) - \bar{x}\|_2^2 = \Omega\left(\min\left\{1, \frac{d}{\varepsilon^2 n^2}\right\}\right),$$

where $\bar{x} = \frac{1}{n}\sum_{i\in[n]} x_i$.

Proposition 6 (Lemma 3.3, Greenewald et al. (2021)). *Let $\{p_i\}_{i\in[N]}, \{q_i\}_{i\in[N]} \subseteq \mathbb{R}^d$ be two collections of points contained in the union of m balls of radius at most Δ, where the distance between every pair of balls is at least 2Δ. If r_j, s_j respectively denote the number of points from the two collections in the jth ball, then every optimal transport contains at least $\frac{1}{2}\sum_{j\in[m]} |r_j - s_j|$ cross-cluster matches. Moreover, if $r_j = s_j$ for every j, there exists an optimal transport with no cross-cluster matches.*

Proof (of Theorem 3). Set $N' := \frac{nk}{m}$. By a standard packing argument, there exists public points $a_1, \dots, a_m \in B_0(1/4)$ with pairwise separations $\|a_r - a_{r'}\|_2 \ge \frac{1}{4}$ by the assumption $m \le \exp(cd)$.

Let ρ be a sufficiently small constant. For each $r \in [m]$, let $u_{r,1}, \dots, u_{r,N'} \in \{\pm 1/\sqrt{d}\}^d$, and partition each block into k groups, forming $\{u_{i,r,j} \mid j \in [n/m]\}$. Define

$$\mu_i := \frac{1}{n}\sum_{r\in[m]}\sum_{j\in[n/m]} \delta_{a_r + \rho u_{i,r,j}}.$$

These are the distributions. Also, define

$$\bar{u}_r := \frac{1}{N'}\sum_{i\in[k]}\sum_{j\in[n/m]} u_{i,r,j}, \quad z_r := a_r + \rho\bar{u}_r, \quad \nu^\star := \frac{1}{m}\sum_{r\in[m]} \delta_{z_r}.$$

It is clear that these points lie in $B_0(1/2)$. Additionally, since $\|\bar{u}_r\|_2 \le 1$, each z_r lies within distance ρ of a_r. Observe by replacing each atom of ν^* with n/m identical copies of mass $1/n$, Proposition 6 yields

$$F(\nu^*) = \sum_{r\in[m]} F_r^* := \sum_{r\in[m]} \frac{\rho^2}{kn}\sum_{i\in[k]}\sum_{j\in[n/m]} \|u_{i,r,j} - \bar{u}_r\|_2^2. \tag{12}$$

Let $\hat{\nu} = \frac{1}{m}\sum_{j\in[m]} \delta_{y_j}$ be the output of the private algorithm. Consider the disjoint balls $B_r := B_{a_r}(10\rho)$ for $r \in [m]$. If ρ is sufficiently small, their pairwise separations are $\ge 20\rho$.

Suppose that there exists exactly one output atom in each of these balls, which we denote as $y_r := a_r + \rho\tilde{u}_r$. By the multiplicity argument, the two point collections have the same number of points in each cluster, so Proposition 6 implies the optimal transport map has no cross-cluster matches, and

$$\begin{aligned} F(\hat{\nu}) - F(\nu^*) &= \frac{\rho^2}{kn}\sum_{r\in[m]}\left(\sum_{i\in[k]}\sum_{j\in[n/m]} \|u_{i,r,j} - \tilde{u}_r\|_2^2 - \sum_{i\in[k]}\sum_{j\in[n/m]} \|u_{i,r,j} - \bar{u}_r\|_2^2\right) \\ &= \frac{\rho^2}{m}\sum_{r\in[m]} \|\tilde{u}_r - \bar{u}_r\|_2^2 = \frac{1}{m}\sum_{r\in[m]} \|y_r - z_r\|_2^2. \end{aligned}$$

[5] This distributional form follows from the minimax theorem.

It remains to handle outputs that do not place exactly one atom near each a_r; here Proposition 6 then forces cross-cluster transport, where each such match costs $\Omega(1)$. Formally, we call $r \in [m]$ good if B_r contains exactly one output atom and bad otherwise. Let b be the count of bad indices; among the b bad balls, let e be the number that contain no output atom. By the pigeonhole principle, every nonempty bad ball contains at least two atoms, so $e \geq \frac{b}{2}$.

Fix a μ_i. Again, by the multiplicity argument previously, each of the empty bad balls will collectively incur $\frac{en}{m} \geq \frac{nb}{2m}$ cross-cluster matches outside their ball. We claim these cross-cluster matches will incur $\geq 81\rho^2$ excess cost to the objective. Indeed, by the separation of the a_r and construction of the B_r, we will have $(\frac{1}{4} - 11\rho)^2 - (11\rho)^2 \geq 81\rho^2$ if it is sent to another ball, and $\geq 81\rho^2$ if it is sent outside $\cup_{r\in[m]} B_r$. Now using (12), $F_r^* \leq \frac{\rho^2}{m}$, and normalization yields

$$\begin{aligned} F(\hat{\nu}) - F(\nu^*) &\geq \frac{1}{m} \sum_{r \text{ good}} \|y_r - z_r\|_2^2 + \frac{81\rho^2 b}{2m} - \sum_{r \text{ bad}} F_r^* \\ &\geq \frac{1}{m} \sum_{r \text{ good}} \|y_r - z_r\|_2^2 + \left(\frac{81}{2} - 1\right) \frac{\rho^2 b}{m} \end{aligned}$$

For private decoding, we set $y_r = a_r$ for the bad points, from which the preceding display yields

$$F(\hat{\nu}) - F(\nu^*) \geq \frac{1}{m} \sum_{r\in[m]} \|y_r - z_r\|_2^2.$$

We now apply Proposition 5. Draw the m blocks independently, with each block $U_r := (u_{r,1}, \ldots, u_{r,N'}) \sim \mathcal{Q}$, where $\mathcal{Q}$ is the distribution from Proposition 5. Fix $r \in [m]$ and condition on all blocks except the rth one. Since y_r is a deterministic function of $\hat{\nu}$ and the a_r, $\frac{y_r - a_r}{\rho}$ is an (ε, δ)-DP estimator of $\bar{u}_r$ (with sample size $N' = \frac{nk}{m}$). Hence, the distributional form of Proposition 5 gives

$$\mathbb{E}_{U_r \sim \mathcal{Q}, \mathcal{A}}[\|y_r - (a_r + \rho \bar{u}_r)\|_2^2 \mid U_{-r}] = \Omega\left(\rho^2 \min\left\{1, \frac{d}{\varepsilon^2 N'^2}\right\}\right).$$

Combining the above displays and averaging then yields

$$\mathbb{E}_{U_1,\ldots,U_m \overset{\text{i.i.d.}}{\sim} \mathcal{Q}, \mathcal{A}}\left[F(\hat{\nu}) - \mathsf{OPT}\right] = \Omega\left(\min\left\{1, \frac{m^2 d}{\varepsilon^2 (nk)^2}\right\}\right),$$

using $\mathsf{OPT} \leq F(\nu^*)$. Since the bound holds on average over $\mathcal{Q}$, some deterministic input attains at least this expected cost, which completes the proof. □