# Differentially Private Sinkhorn Algorithm


Jiaqi Wang
*Cornell University*
jw2382@cornell.edu

Ziv Goldfeld
*Cornell University*
goldfeld@cornell.edu


2024 60th Annual Allerton Conference on Communication, Control, and Computing (Allerton) | 979-8-3315-4103-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ALLERTON63246.2024.10735319


***Abstract*—Optimal transport (OT) theory play a central role in the design and analysis of various machine learning algorithms. As such, approximate computation of the OT cost between large-scale dataset via the popular Sinkhorn algorithms forms a basic primitive. However, this approach may lead to privacy violations when dealing with datasets that contain sensitive information. To address this predicament, we propose a differentially private variant of the Sinkhorn algorithm and couple it with formal guarantees by deriving its privacy utility tradeoff (PUT). To that end, the Sinkhorn algorithm is treated as a block coordinate descent algorithm scheme, which we privatize by injecting Gaussian noise to the iterates. We establish a linear convergence rate for our private Sinkhorn algorithm and analyze its privacy by controlling the Rényi divergence between outputs corresponding to neighboring input dataset. Combining these results we obtain the desired PUT. In doing so, this work also closes an existing gap in formal guarantees for private constrained nonlinear optimization. As an application, we employ the noisy Sinkhorn algorithm for differentially private (approximate) computation of OT cost and derive insights from its PUT.**


## I. Introduction

Optimal transport (OT) theory [1] provides a natural framework for comparing probability distributions. Specifically, the OT problem with cost function $c$ between two probability measures $\mu, \nu$ on $\mathbb{R}^d$ is

$$\mathsf{OT}_c(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}_\pi[c(X,Y)],$$

where $\Pi(\mu,\nu)$ is the set of couplings between $\mu$ and $\nu$. The special case is the $p$-Wasserstein distance for $p \in [1,\infty)$, is given by $\mathsf{W}_p(\mu,\nu) := \big(\mathsf{OT}_{\|\cdot\|^p}(\mu,\nu)\big)^{1/p}$. The Wasserstein distance has found applications in various fields, encompassing machine learning [2, 3], statistics [4, 5], and applied mathematics [6]. This widespread applicability is driven by an array of desirable properties that the Wasserstein distance possesses, including its convenient dual form, robustness to support mismatch, and a rich geometry it induces on a space of probability measures.

Despite the aforementioned empirical progress, the OT problem suffers from the statistical and computational hardness issues. The estimation rate of the OT cost between distributions on $\mathbb{R}^d$ is generally $n^{-1/d}$ [7], which deteriorates exponentially with dimensions—a phenomenon known as the curse of dimensionality. Computationally, OT is a linear program (LP), solvable in $O(n^3 \log n)$ time for distribution on $n$ points [8], but this complexity becomes prohibitive when $n$ is large (as the statistical considerations mandate it to be). To circumvent these issues, entropic regularization has emerged as a popular alternative [9]

$$\mathsf{OT}_c^\eta(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}_\pi[c(X,Y)] + \eta \mathsf{D}_{\mathsf{KL}}(\pi\|\mu\otimes\nu), \quad (1)$$

where $\mathsf{D}_{\mathsf{KL}}$ is the Kullback-Leibler (KL) divergence and $\eta > 0$ is a regularization parameter. Empirical estimation of EOT enjoys the parametric $n^{-1/2}$ convergence rate in arbitrary dimension, under several settings [10, 11]. Computationally, entropic OT (EOT) between discrete distributions can be efficiently solved via the Sinkhorn algorithm in $O(n^2 \log n)$ time [9].

Motivated by pressing privacy concerns, which arise due to the increase in personal data shared online and advancements in data mining techniques, the goal of this work is to devise a method for private computation of EOT/OT subject to formal guarantees. We adopt the differential privacy (DP) paradigm [12]—a gold standard in statistical privacy—which preserves the privacy of individual records while enabling aggregate queries about a database.

### *A. Contributions*

We propose a differentially private Sinkhorn algorithm, which casts the optimization as block coordinate descent (BCD) and injects adaptive noise to the iterates. To couple the approach with formal guarantees, we derive the privacy utility tradeoff (PUT) by studying the convergence rate and the resulting privacy of our method. We prove a linear convergence rate up to a noisy neighborhood of the global solution, which matches the optimal convergence rate of stochastic gradient descent (SGD) of a strongly convex function [13]. The privacy analysis accounts for adaptive noise proportional to the progress of the algorithm. Combining these results we obtain the PUT for noisy Sinkhorn. Remarkably, while noisy Sinkhorn solves a constrained convex optimization problem, the obtained PUT coincides with that of unconstrained convex optimization under DP, up to a logarithmic factor [14].

As an application of our private algorithm, we consider DP computation of classical OT. We propose a mechanism that combines the noisy Sinkhorn algorithm with the Laplace noise injection. Specifically, after computing the private EOT coupling, the mechanism outputs the transportation cost, perturbed by Laplace noise to further boost privacy. Compared to exactly solving DP OT as an LP, entropic regularization enables an approximate solution with improved computational efficiency. Using PUT, we characterize this accuracy - effiency trade-off and derive the optimal entropic regularization parameter.

Our analysis employs several key technical tools. The utility analysis draws upon EOT structure [15] and convergence rate derivations for SGD algorithms [13]. The privacy analysis results from tracking the $\alpha$-Rényi divergence [16] along the noisy iterations and then converting the accumulated $\alpha$-Rényi DP (RDP) bound into an $(\varepsilon,\delta)$-DP guarantee. For the application to differentially private OT computation, our analysis employs the asymptotic expansion of EOT around OT [17] and sensitivity analysis of the linear component of EOT [18] to derive the combined PUT.

### B. Related work

Several works explore differentially private optimization. For linear programming (LP) under DP requirements, [19, 20] address private optimization by perturbing the objective function, establishing PUT under exact minimization. In [21, 22, 23], efficient algorithms via SGD or the exponential mechanism are presented and the resulting PUT is characterized. However, these works focus on unconstrained problems, rendering them inapplicable to the EOT setting.

Although differentially private OT and EOT have also been explored before, its application in large-scale machine learning is hindered by the absence of PUT analyses and efficient algorithms [24]. In [25, 26], privacy guarantees for input perturbation were derived under the assumption of exact minimization, without studying algorithms. EOT computation via private SGD algorithms was considered in [27, 28, 29] and privacy guarantees were derived, but no account of the convergence rate was provided. In sum, none of these existing works present both efficient algorithm and PUT analysis for OT/EOT computation, which is essential from theoretical and practical considerations alike.

## II. Background and Preliminaries

### A. Entropic Optimal Transport

EOT is a convexification of the (originally linear) OT problem by means of an entropic penalty, which lends itself well for efficient computation via Sinkhorn's fixed point iteration algorithm [30]. Given a lower semi continuous cost function $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}_{\geq 0}$ and a regularization parameter $\eta\geq 0$, the EOT problem between two probability distributions $(\mu,\nu)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ is[1] given by $\mathsf{OT}_c^\eta(\mu,\nu)$ in (1), where the KL divergence is defined as $\mathsf{D}_{\mathsf{KL}}(\alpha\|\beta)=\int\log\left(\frac{d\alpha}{d\beta}\right)d\alpha$ if $\alpha\ll\beta$ and $+\infty$ otherwise. Note that the regularizer can be equivalently written as the mutual information $\mathsf{I}_\pi(X;Y)$, which is therefore understood as encouraging weakly dependent couplings. OT is obtained from EOT by setting $\eta=0$.

When $c\in L^1(\mu\otimes\nu)$, EOT admits the following dual formulation,

$$\mathsf{OT}_c^\eta(\mu,\nu)=\sup_{\varphi,\psi}\mathbb{E}_\mu[\varphi]+\mathbb{E}_\nu[\psi]-\eta\Big(\mathbb{E}_{\mu\otimes\nu}\big[e^{\frac{\varphi\oplus\psi-c}{\eta}}\big]+1\Big),\tag{2}$$

[1] $\mathcal{P}(\mathcal{X})$ denotes the class of Borel probability measures on $\mathcal{X}\subseteq\mathbb{R}^d$.

where the supremum is over $(\varphi,\psi)\in L^1(\mu)\otimes L^1(\nu)$ and $\varphi\oplus\psi(x,y)=\varphi(x)+\psi(y)$. Dual potentials are unique up to additive constants, in the sense that any other solution $(\bar\varphi,\bar\psi)$ satisfies $\bar\varphi=\varphi+a$ $\mu$-a.s. and $\bar\psi=\psi-a$ $\nu$-a.s., for some $a\in\mathbb{R}$. This dual form is central for computing EOT via the Sinkhorn algorithm and deriving its differentially private version herein.

The standard Sinkhorn algorithm for EOT can equivalently be viewed as a block coordinate descent algorithm [15]. When $\mu,\nu$ are discrete distributions supported on $\{x_1,\cdots,x_n\}\times\{y_1,\cdots,y_n\}$, respectively, the EOT problem becomes

$$\mathsf{OT}_c^\eta(\mu,\nu)=\min_{\pi\in\Pi(\mu,\nu)}\langle c,\pi\rangle-\eta(\mathsf{H}(\pi)-1),\tag{3}$$

where $c_{ij}=c(x_i,y_j)$, $\langle c,\pi\rangle:=\sum_{i,j=1}^n c_{ij}\pi_{ij}$, and $\mathsf{H}(\pi):=\sum_{i,j=1}^n\pi_{ij}\log(\pi_{ij})$, Setting $K_{ij}=e^{-c_{ij}/\eta}$, the first-order optimality condition dictates that the optimal coupling $\pi^\star\in[0,1]^{n\times n}$ is given in terms of optimal dual potentials $(\varphi^\star,\psi^\star)\in\mathbb{R}^n\times\mathbb{R}^n$ by $\pi^\star_{ij}=e^{\varphi^\star_i/\eta}K_{ij}e^{\psi^\star_j/\eta},\quad i,j=1,\ldots,n$, subject to the marginal constraints

$$\begin{cases}\exp(\varphi^\star/\eta)\odot(K\exp(\psi^\star/\eta))=\mu\\ \exp(\psi^\star/\eta)\odot K^\intercal(\exp(\varphi^\star/\eta))=\nu\end{cases}\tag{4}$$

where $\odot$ denotes to entrywise multiplication. Thus, finding $\pi^\star$ boils down to solving the latter fixed point equations, which can be done via Sinkhorn's algorithm [30].

### B. Differential Privacy

DP allows answering queries about aggregate quantities, while protecting the individual data entries [12]. To that end, the output of differentially private mechanism should be indistinguishable for neighboring datasets—those that differ only in a single sample. Formally, two datasets $D_n=\{x_i\}_{i=1}^n$ and $D'_n=\{x'_i\}_{i=1}^n$ are called neighbors, denoted $D_n\sim D'_n$, if there exists $j\in\{1,\ldots,n\}$ such that $x_j\neq x'_j$ and $x_i=x'_i$ for all $i\neq j$.

**Definition 1** (Differential privacy)**.** *Fix $\varepsilon,\delta>0$. A randomized mechanism*[2] *$M:\mathcal{X}^n\to\mathcal{Y}$ is $(\varepsilon,\delta)$-differentially private (DP) if for all $D_n\sim D'_n$ and $\mathcal{A}\subseteq\mathcal{Y}$ measurable, we have*

$$\mathbb{P}\big(M(D_n)\in\mathcal{A}\big)\leq e^\varepsilon\,\mathbb{P}\big(M(D'_n)\in\mathcal{A}\big)+\delta.\tag{5}$$

**Definition 2** (Function sensitivity)**.** *We define the $\ell^p$-sensitivity of $f:\mathcal{X}^n\to\mathbb{R}^d$ as*

$$\Delta_p[f]:=\max_{D_n\sim D'_n}\|f(D_n)-f(D'_n)\|_p$$

## III. Differentially Private Entropic OT

We consider differentially private computation of an EOT cost between two distribution, both of them contain sensitive data. We start by formulating the problem and then propose a noisy BCD scheme to solve it.

[2] A randomized mechanism is described by a (regular) conditional probability distribution given the data, i.e., $P_{M|X}$.

## A. Problem Formulation

Consider a dataset $D_n = \{x_1, \cdots, x_n, y_1, \cdots, y_n\}$, where $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ for $i = 1, \ldots, n$, and let $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ and $\nu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$ be discrete, uniform distributions on the corresponding $n$ points; in a statistical setting, these are empirical measures over sampled data from two populations. We assumes that $D_n$ contains sensitive data, and is subject to $(\varepsilon, \delta)$-DP requirement.

Our goal is design an $(\varepsilon, \delta)$-DP mechanism for evaluating the EOT cost $\mathsf{OT}_c^{\eta}(\mu_n, \nu_n)$, for a given cost function $c$ with $\|c\|_\infty := \max_{x,y} |c(x, y)| < \infty$, and regularization parameter $\eta \geq 0$. The quality of a mechanism is measured by PUT: the maximum utility for a fixed privacy constraint. We use DP as the privacy metric while approximation error of EOT/OT for the utility metric. We define the PUT function as follows.

**Definition 3** (PUT for EOT/OT computation)**.** *The PUT function $u_\eta : \mathbb{R}^2_{\geq 0} \to \mathbb{R}_{\geq 0}$ for differentially private computation of $\mathsf{OT}_c^{\eta}(\mu_n, \nu_n)$, with $\eta \geq 0$, is given by*

$$u_\eta(\varepsilon, \delta) = \inf_{M \in \mathcal{M}_{\varepsilon,\delta}} \left|\mathbb{E}[M(D_n)] - \mathsf{OT}_c^{\eta}(\mu_n, \nu_n)\right|,$$

*where $\mathcal{M}_{\varepsilon,\delta}$ is the set of all $(\varepsilon, \delta)$-DP mechanisms $M : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$ acting on the dataset $D_n = \{\mathbf{x}, \mathbf{y}\}$, i.e., satisfying (5) with $D_n'$, where $D_n' \sim D_n$ are neighbours.*

Different from the existing literature, we will propose an efficient algorithm with both convergence rate and PUT guarantees.

# IV. Differentially Privacy Sinkhorn Algorithm

We propose a noisy Sinkhorn algorithm to solve the DP EOT problem and analyze its performance (convergence rate and the PUT). We then consider an application of our algorithm to approximate the unregularized OT cost between $\mu_n$ and $\nu_n$ in a differentially private manner. Lastly, motivated by practical considerations, we formulate an optimization problem to determine the entropic regularization parameter that achieves the best PUT.

## A. The Algorithm

We present a differentially private variant of the popular Sinkhorn algorithm for EOT [9]. The key idea is to view it as a BCD scheme and inject noise at an appropriate scale to the iterates.

To describe the approach, we first express EOT through its dual form

$$\mathsf{OT}_c^{\eta}(\mu_n, \nu_n) = \max_{\varphi, \psi \in \mathbb{R}^n} \mathsf{D}_\eta(\varphi, \psi),$$

where $\mathsf{D}_\eta$ is the functional on the right-hand side (RHS) of (2); note that the dual potentials reduce to vectors in $\mathbb{R}^n$ in this finitely-supported case. For notational convenience, we also define the alternative objective function $\mathsf{F}_\eta(\varphi, \psi) := -\mathsf{D}_\eta(\varphi, \psi) + \eta$, and equivalently, consider the minimization problem

$$\min_{\varphi, \psi \in \mathbb{R}^n} \mathsf{F}_\eta(\varphi, \psi).$$

With that, we treat Sinkhorn's algorithm as a BCD scheme and employ stochastic programming techniques, as summarized below.

**Algorithm 1** Noisy Sinkhorn algorithm

1: Initialize $\varphi_0, \psi_0$ to be two arbitrary vectors in $\mathbb{R}^n$. Let iteration $K$, noise $\sigma, M$ be given.
2: **for** $k = 1, \cdots K$ **do**
——— Sinkhorn + centering ———
3: $\tilde{\varphi}_{k+1} = \operatorname{argmin}_\varphi \mathsf{F}_\eta(\varphi, \psi_k)$
4: $\varphi'_{k+1} = \tilde{\varphi}_{k+1} - \mathbb{E}_{\mu_n}[\tilde{\varphi}_{k+1}]$
5: $\psi'_{k+1} = \operatorname{argmin}_\psi \mathsf{F}_\eta(\varphi_{k+1}, \psi)$
——— Noise injection ———
6: $s_k^2 = \sigma^2(\|\varphi'_{k+1} - \varphi_k\|^2 + \|\psi'_{k+1} - \psi_k\|^2) + M$
7: $Z_{k+1} = (Z^1_{k+1}, Z^2_{k+1}) \sim N(0, s_k^2 \mathbf{I}_{2n})$
8: $\varphi_{k+1} = \varphi'_{k+1} + Z^1_{k+1}$
9: $\psi_{k+1} = \psi'_{k+1} + Z^2_{k+1}$
10: **end for**
11: Output $(\varphi_K, \psi_K)$

Evidently, the algorithm first computes the Sinkhorn iterates, and upon centering them, injects zero-mean Gaussian noise. The noise variance $s_k^2$, as given in Line 6 of Algorithm 1, is adapted to the step size: as the step size gets smaller, the noise level decreases. Such adaptive noise allocations achieve better utility under a fixed privacy constraint. However, at optimality the adaptive noise nullifies, which can violate DP. To correct for that, we add a noise floor level of $M > 0$ to all iterations. We will determine the choice of $K, M, \sigma$ that optimizes PUT in the following content. We also note that the minimizers in Lines 3 and 5 can be obtained in closed-form as the $(c, \eta)$-transform of the original function [31], namely

$$\varphi^{c,\eta}(y) = -\eta \log\left(\mathbb{E}_{\mu_n}\left[\exp\left(\frac{\varphi(X) - c(X, y)}{\eta}\right)\right]\right).$$

## B. Convergence Analysis and the PUT

Let $\tau_\eta = \frac{1}{2\eta}\exp(-\frac{6\|c\|_\infty}{\eta})$ be the strong convexity parameter, where $\|c\|_\infty = \max_{x,y} c(x, y) < \infty$ is assumed throughout. For simplicity, we denote the exact iterates (after centering) as $\theta'_k = (\varphi'_k, \psi'_k)$, and their noisy version by $\theta_k = (\varphi_k, \psi_k) = \theta'_k + Z_k$, where $Z_k = (Z^1_k, Z^2_k) \sim N(0, s_k^2 \mathbf{I}_{2n})$ with $s_k^2 = \sigma^2\|\theta'_k - \theta_{k-1}\|_2^2 + M$. Let $\theta^* = (\varphi^*, \psi^*) = \arg\min_\theta \mathsf{F}_\eta(\theta)$ be minimizer, which is unique due to the strong convexity of $\mathsf{F}_\eta, \forall \eta > 0$. We first analyze the utility of noisy Sinkhorn in terms of the accuracy with which it approximates the EOT cost.

**Theorem 1** (Utility of noisy Sinkhorn)**.** *For any $0 \leq \sigma^2 < 1$, $M \geq 0$, $\theta_0 \in \mathbb{R}^n \times \mathbb{R}^n$ and $k \in \mathbb{N}$, Algorithm 1 satisfies*

$$\begin{aligned} &|\mathsf{OT}_c^{\eta}(\mu_n, \nu_n) + \mathbb{E}[\mathsf{F}_\eta(\theta_k)] - \eta| \\ &\leq (1 - \rho)^k \left(\mathsf{F}_\eta(\theta_0) - \mathsf{F}_\eta(\theta^*) - \frac{\tau_\eta M}{2\rho}\right) + \frac{\tau_\eta M}{2\rho}, \end{aligned} \tag{6}$$

*where $\rho = \frac{(1-\sigma^2)}{8} e^{-24\|c\|_\infty/\eta}$.*

Theorem 1 is proven in the Section V via several key lemmas. The argument first shows that $\mathsf{F}_\eta$ is component-wise Lipschitz continuous and jointly convex. From that we can lower bound the progress per iteration and upper bound the progress to achieve optimality. Combining the upper and lower bound, we find the convergence rate.

**Remark 1** (Rate and dependence on $\rho$)**.** *This results matches the optimal convergence rate of stochastic programming of a strongly convex function. The gap to EOT converges at a linear rate $(1-\rho)^k$ up to a noisy neighborhood $\frac{\tau_\eta M}{2\rho}$. The noisy perturbation of the iterates results in an interesting pattern that differs from traditional stochastic gradient optimization [13]. Namely, increasing $\rho$ favors both a faster convergence rate and a smaller noisy neighborhood.*

**Theorem 2** (Privacy of noisy Sinkhorn)**.** *For any $\delta > 0$, and $K \in \mathbb{N}$, the output $(\varphi_K, \psi_K)$ of Algorithm 1 satisfies $\left(KC/M + \sqrt{KC\log(1/\delta)/M}, \delta\right)$-DP, where $C = \frac{\eta}{2}\log\left(1+\frac{4\eta\|c\|_\infty}{n}e^{\frac{6}{\eta}\|c\|_\infty}\right)$.*

The proof is given in Section V, where we first bound the sensitivity of the iteration function

$$(\varphi'_{k+1}, \psi'_{k+1}) = \operatorname{argmin}_{\varphi,\psi}\left(\mathsf{F}_\eta(\varphi,\psi_k), \mathsf{F}_\eta(\varphi_{k+1},\psi)\right),$$

and then leverage the Gaussian mechanism and the composition rule to establish the privacy guarantee.

**Remark 2** (Improved privacy analysis)**.** *The worst-case complexity of strongly convex optimization [32] implies $|\theta_k - \theta'_{k+1}| \geq p^k$ for some $0 < p < 1$. This leads to tighter privacy bound by replacing $M$ with $\sigma^2 p^K + M$. Given that this substitution doesn't affect the PUT where $\sigma = 0$ is optimal, we narrow our attention to the previous result and skip the explicit statement.*

Combining Theorems 1 and 2, we arrive at the PUT for the noisy Sinkhorn algorithm.

**Corollary 1** (PUT of noisy Sinkhorn)**.** *Fix $\varepsilon, \delta, \eta > 0$ and let $\alpha > 1$ be such that $(\alpha-1)^2 = \frac{\log(1/\delta)\alpha}{\varepsilon}$. The PUT function for differentialy private computation of $\mathsf{OT}_c^\eta(\mu_n,\nu_n)$ is given by*

$$u_\eta(\varepsilon,\delta) = \frac{C_\eta}{\varepsilon - \underline{\varepsilon}}\left[a_0 - \log\left(\frac{C_\eta}{\varepsilon-\underline{\varepsilon}}\right)\right], \tag{7}$$

*where $\underline{\varepsilon} = \frac{\log(1/\delta)}{\alpha-1}$, $a_0 = \mathsf{F}_\eta(\theta_0) - \mathsf{F}_\eta^*$, $C_\eta = \frac{32\eta\nu^{\frac{8}{\eta}}\ln\nu}{3n}$, and $\nu = e^{6\|c\|_\infty}$.*

This result follows by jointly optimizing the noise levels $\sigma, M$, and the number of iteration $K$ from Theorems 1 and 2. The optimal values are $\sigma = 0$, $M = O(\log(\frac{1}{\epsilon-\underline{\varepsilon}}))$, and $K = \frac{(\varepsilon-\underline{\varepsilon})M}{\alpha C}$. The constant $C_\eta$ is a simplified upper bound, whose exact expression can be found in Section V-C Eq. (17).

**Remark 3** (Nullifying adaptive noise)**.** *Choosing $\sigma = 0$ isn't just about loose privacy bounds but also about algorithm behavior. As the step size in Algorithm 1 diminishes exponentially, adaptive noise's privacy gains can't offset utility loss. Even with the improved bound in Remark 2, $\sigma = 0$ remains optimal. A promising avenue for future research is slowing down convergence to achieve more privacy gains by adding adaptive noise.*

**Remark 4** (Comparison to unconstrained optimization)**.** *To the best of our knowledge, our PUT analysis for the noisy Sinkhorn algorithm is the first to treat differentially private convex optimization with constraints. The PUT we achieve in Corollary 1 coincides with that of unconstrainted convex optimization $O(\frac{1}{\epsilon})$ [14], up to a log factor, and is worse than that of unconstrained strongly convex optimization $O(\frac{1}{\varepsilon^2})$ [22]. As we only present an upper bound on the PUT attained by noisy Sinkhorn, generic lower bounds for differentially private, constrained convex optimization remain an interesting question for future work.*

### *C. Application to Differentially Private Optimal Transport*

OT is widely used in large-scale machine learning problems, requiring algorithms with low computation complexity. Differentially private OT can be computed via an LP mechanism [20], with time complexity $O(n^3\log n)$ for distribution on $n$ points. This cubic complexity, however, can be prohibitively slow when $n$ is large. To accelerate this time, we propose computing the differentially private entropic approximation of OT using the noisy Sinkhorn algorithm. Entropic regularization makes the problem strongly convex, allowing faster computation, at the cost of some accuracy loss due to EOT approximation. We study this trade-off via a PUT analysis, leaving time complexity quantification for future work. Given the same dataset $D_n$, privacy parameters $(\epsilon,\delta)$, and a cost function $c$, we design an $(\epsilon,\delta)$-DP mechanism for evaluating the OT cost $\mathsf{OT}_c^0(\mu_n,\nu_n)$ as follows.

- **Stage 1: Noisy Sinkhorn.** Run Algorithm 1 to compute the optimal coupling for $\mathsf{OT}_c^\eta(\mu_n,\nu_n)$. Let $\hat\theta = (\hat\varphi,\hat\psi)$ be the output of the algorithm and define $\hat\pi_\eta$ via $(\hat\pi_\eta)_{ij} = \exp\left(\frac{(\hat\varphi)_i + (\hat\psi)_j - c_{ij}}{\eta}\right)$ as the algorithmic proxy of the true EOT coupling $\pi_\eta$.
- **Stage 2: Noise injection.** Let $Z$ follow a Laplace distribution with parameter $b$. The output of the overall mechanism is the transportation cost associated with $\hat\pi_\eta$ perturbed by the Laplace noise:

$$\hat M_\eta^b(D_n) = \langle c, \hat\pi_\eta\rangle + Z, \quad Z \sim \mathrm{Lap}(b). \tag{8}$$

For analysis, we introduce the following notation. Let $M_1 = \langle c,\pi_\eta\rangle + Z$, $g_1 = \mathsf{OT}_c^0(\mu_n,\nu_n)$, $M_2 = \langle c,\hat\pi_\eta\rangle$, and $g_2 = \langle c,\pi_\eta\rangle$. Then

$$\hat M_\eta^b(D_n) - \mathsf{OT}_c^0(\mu_n,\nu_n) = \underbrace{M_1 - g_1}_{\text{EOT approx.}} + \underbrace{M_2 - g_2}_{\text{Noisy Sinkrhon}} \tag{9}$$

Note that the subscript $\hat\pi_\eta$ indicates that it is the algorithmic solution to EOT rather than the exact optimal optimal coupling $\pi_\eta$.

The PUT function is defined as

$$u_\eta^{(i)} = \inf_{M_i \in \mathcal{M}_{\varepsilon,\delta}} \left|\mathbb{E}\left[M_i(D_n)\right] - g_i\right|, i = 1,2$$

where $\mathcal{M}_{\varepsilon,\delta}$ is the set of all $(\varepsilon,\delta)$-DP mechanisms $M : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$. First we characterize the PUT for EOT approximation and Noisy Sinkhorn respectively.

**Theorem 3** (PUT of two-stage DP OT)**.** *For any $\varepsilon, \delta \geq 0$, initial Sinkhorn variable $\theta_0$, and $\eta \leq \min\{\frac{\Delta}{1+\log n}, \frac{2\|c\|_\infty}{n \log n}\}$,*

$$u_\eta^{(1)}(\epsilon,\delta) = \underline{u} + \frac{4\sqrt{2}\|c\|_\infty}{n\epsilon} \tag{10}$$

*Where $\underline{u} = \Delta \exp(-\frac{\Delta}{\eta} + 1 + \log n)$, with $\Delta$ being the suboptimality gap of $\mathsf{OT}_c^0(\mu_n,\nu_n)$ defined in (18).*

$$u_\eta^{(2)}(\varepsilon,\delta) = C'_\eta \sqrt{\frac{C_\eta}{\varepsilon - \underline{\varepsilon}}\left[a_0 - \log\left(\frac{C_\eta}{\varepsilon - \underline{\varepsilon}}\right)\right]}, \tag{11}$$

*Where $\underline{\varepsilon} = \frac{\log(1/\delta)}{\alpha - 1}$, $a_0 = \mathsf{F}_\eta(\theta_0) - \mathsf{F}^*_\eta$, $C_\eta = \frac{32\eta\nu^{\frac{8}{\eta}}\ln\nu}{3n}$, $C'_\eta = n\|c\|_\infty\sqrt{\frac{8}{\eta}}v^{\frac{3}{2\eta}}$, and $\nu = e^{6\|c\|_\infty}$.*

The PUT function $u_\eta^{(1)}$ comprises two elements: a lower bound on utility stemming from the approximation error of EOT to OT, and another term from the Laplace mechanism's noise. We present a simplified case for small $\eta$, postponing a comprehensive depiction of the PUT function to Section V-D. Meanwhile, $u_\eta^{(2)}$ derives from converting EOT to its linear component.

Using Eq. (9) and triangle inequality to combine the above two PUT functions, we obtain the overall PUT function.

**Corollary 2** (Overall PUT function)**.** *The PUT function of DP OT mechanism $\mathcal{M} = \{\hat{M}_\eta^b(D_n)\}$ (defined in Definition 3 with $\eta = 0$) satisfies*

$$u_0(\varepsilon,\delta) = \min_{\varepsilon_1,\varepsilon_2,\eta} \; u_\eta^{(1)}(\varepsilon_1,\delta) + u_\eta^{(2)}(\varepsilon_2,\delta) \quad s.t.\ \varepsilon_1 + \varepsilon_2 = \varepsilon \tag{12}$$

Given the data set and privacy budgets, the key question is how to tune the entropic regularization parameter $\eta$ to achieve the best PUT. The following optimization problem, derived by highlighting $\eta$ in Eq. (12), provides guidelines for selecting the optimal $\eta$.

$$\min_{\eta,\varepsilon_2} C_1 \exp\left(-\frac{\Delta}{\eta}\right) + 8C_2\sqrt{\frac{\nu^{\frac{11}{\eta}}}{C_3\eta}\left(a_0 + \frac{8}{\eta}\log\left(\frac{\eta\nu}{C_3}\right)\right)}, \tag{13}$$

where $C_1 = n\Delta e$, $C_2 = n\|c\|_\infty$, $C_3 = \frac{1}{\epsilon_2 - \underline{\epsilon}}$

We now interpret Eq. (13) intuitively. The first term, resulting from the EOT approximation of OT, increases with $\eta$, indicating accuracy loss due to larger regularization. The second term, representing noisy Sinkhorn, decreases with $\eta$, as larger regularization accelerates the convergence of the Sinkhorn algorithm. This leads to fewer iterations and, consequently, less privacy loss for the same utility.

# V. Proofs

## A. Proof of Theorem 1

The argument relies on the following technical lemmas. Throughout, we assume that $\|c\|_\infty < \infty$.

**Lemma 1** (Bounded dual variables [31])**.** *For all $k > 0$, the iterates from Algorithm 1 satisfies $\|\varphi'_k\|_\infty \leq 2\|c\|_\infty$, $\|\psi'_k\|_\infty \leq 3\|c\|_\infty$*

Bounded dual variables leads to the Lipschitz smoothness and strong convexity of the objective $\mathsf{F}_\eta$. To state the result, set $L_\eta = \frac{1}{2\eta}e^{\frac{6\|c\|_\infty}{\eta}}$ and $\tau_\eta = \frac{1}{2\eta}e^{-\frac{6\|c\|_\infty}{\eta}}$.

**Lemma 2** (Lipschitz smoothness and joint convexity)**.** *The objective $\mathsf{F}_\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfies:*

*(i) blockwise $L_\eta$-Lipschitz smoothness, i.e., fixing $\varphi \in \mathbb{R}^n$, we have that $\mathsf{F}_\eta(\varphi,\cdot)$ has $L_\eta$-Lipschitz gradients and similarly for the other variable;*

*(ii) if further $\varphi,\varphi' \in \mathbb{R}^n$ are such that $\mathbb{E}_{\mu_n}[\varphi] = \mathbb{E}_{\mu_n}[\varphi'] = 0$, then $\mathsf{F}_\eta$ is jointly strongly convex, i.e.,*

$$\mathsf{F}_\eta(\varphi',\psi') - \mathsf{F}_\eta(\varphi,\psi) \geq \frac{\tau_\eta}{2}\|(\varphi',\psi') - (\varphi,\psi)\|^2 + \nabla\mathsf{F}_\eta(\varphi,\psi)^\intercal\big((\varphi',\psi') - (\varphi,\psi)\big) \tag{14}$$

Using the above properties, we devise the following two key conditions for convergence rate. Recall that $\theta_k = (\varphi_k,\psi_k)$ denotes the $k$-th iterate of Algorithm 1.

**Lemma 3** (Optimality gap bound and sufficient descent)**.** *Algorithm 1 with $\sigma = M = 0$, which reduces to the standard Sinkhorn algorithm, satisfies:*

*(i) upper bound on optimality gap:*

$$\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta^*) \leq \frac{4L_\eta^2}{\tau_\eta}\|\theta_k - \theta_{k+1}\|^2$$

*(ii) sufficient descent:*

$$\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta_{k+1}) \geq \frac{\tau_\eta}{2}\|\theta_k - \theta_{k+1}\|^2$$

*Proof.* The following argument is inspired by Lemma 3.2, Theorem 3.3 in [15]. We reformulate here in a systematic way and include the dependence on $\eta$. Let $\tilde{\theta}_k = (\varphi_{k+1},\psi_k)$ be the partial update. The fact that each iterate is an exact minimization implies

$$\nabla_\varphi\mathsf{F}_\eta(\tilde{\theta}_k) = \nabla_\psi\mathsf{F}_\eta(\theta_{k+1}) = 0 \tag{15}$$

For each fixed $k$, by combining the strong convexity from (14) and the optimality condition from (15), we obtain

$$\begin{aligned}
&\mathsf{F}_\eta(\theta^*) - \mathsf{F}_\eta(\theta_k)\\
&\quad\geq \nabla\mathsf{F}_\eta(x_k)^\intercal(\theta^* - \theta_k) + \frac{\tau_\eta}{2}\|\theta^* - \theta_k\|^2\\
&\quad= \big(\nabla_\varphi\mathsf{F}_\eta(\theta_k) - \nabla_\varphi\mathsf{F}_\eta(\tilde{\theta}_k)\big)^\intercal(\varphi^* - \varphi_k)\\
&\qquad+ \big(\nabla_\psi\mathsf{F}_\eta(\theta_k) - \nabla_\psi\mathsf{F}_\eta(\theta_{k+1})\big)^\intercal(\psi^* - \psi_k)\\
&\qquad+ \frac{\tau_\eta}{2}\big(\|\varphi^* - \varphi_k\|^2 + \|\psi^* - \psi_k\|^2\big)\\
&\quad\geq -\frac{2}{\tau_\eta}\big(\|\nabla_\varphi\mathsf{F}_\eta(\theta_k) - \nabla_\varphi\mathsf{F}_\eta(\tilde{\theta}_k))\|^2
\end{aligned}$$

$$+ \|\nabla_{\psi}\mathsf{F}_{\eta}(\theta_k) - \nabla_{\psi}\mathsf{F}_{\eta}(\theta_{k+1})\|^2\big)$$

By blockwise $L_\eta$-Lipschitz smoothness of $\mathsf{F}_\eta$, we have $\|\nabla_\varphi \mathsf{F}_\eta(\varphi,\psi) - \nabla_\varphi \mathsf{F}_\eta(\varphi',\psi)\| \leq L_\eta\|\varphi - \varphi'\|$. Combining this and above we get the Claim (i).

We move to derive the sufficient descent condition. Again, employing (14) and (15), consider

$$\begin{aligned}
&\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta_{k+1}) \\
&\quad= \mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\tilde{\theta}_k) + \mathsf{F}_\eta(\tilde{\theta}_k) - \mathsf{F}_\eta(\theta_{k+1}) \\
&\quad\geq \nabla_\varphi \mathsf{F}_\eta(\tilde{\theta}_k)^\intercal(\varphi_k - \varphi_{k+1}) + \frac{\tau_\eta}{2}\|\varphi_k - \varphi_{k+1}\|^2 \\
&\quad+ \nabla_\psi \mathsf{F}_\eta(\theta_{k+1})^\intercal(\psi_k - \psi_{k+1}) + \frac{\tau_\eta}{2}\|\psi_k - \psi_{k+1}\|^2 \\
&\quad= \frac{\tau_\eta}{2}\left(\|\varphi_k - \varphi_{k+1}\|^2 + \|\psi_k - \psi_{k+1}\|^2\right) \\
&\quad= \frac{\tau_\eta}{2}\|\theta_k - \theta_{k+1}\|^2
\end{aligned}$$

□

We now extend the result to the noisy Sinkhorn algorithm. Recall that $\theta_k' = (\varphi_k', \psi_k')$, and $\theta_k = \theta_k' + Z_k$, where $Z_k \sim N(0, \sigma^2\|\theta_{k-1} - \theta_k'\|^2 + M)$.

**Lemma 4** (Extension to noisy Sinkhorn). *The noisy Sinkhorn algorithm satisfies the following optimality gap bound and sufficient descent condition:*

$$\mathbb{E}[\mathsf{F}_\eta(\theta_k)] - \mathsf{F}_\eta(\theta^*) \leq \frac{4L_\eta^2}{\tau_\eta}\|\theta_k - \theta_{k+1}'\|^2$$

$$\mathbb{E}[\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta_{k+1})] \geq \frac{\tau_\eta}{2}(1-\sigma^2)\|\theta_k - \theta_{k+1}'\|^2 - \frac{\tau_\eta M_k}{2}.$$

*Proof.* Upper bounding the optimality gap by adjacent iterate directly follows from Claim (i) of Lemma 3

By the strong convexity of $\mathsf{F}_\eta$ from (14), expanding it around $\theta_{k+1}'$ yields

$$\begin{aligned}
&\mathsf{F}_\eta(\theta_{k+1}) - \mathsf{F}_\eta(\theta_{k+1}') \\
&\leq \frac{\tau_\eta}{2}\|\theta_{k+1} - \theta_{k+1}'\|^2 + \nabla \mathsf{F}_\eta(\theta_{k+1}')^\intercal(\theta_{k+1} - \theta_{k+1}')
\end{aligned}$$

Since $\mathbb{E}[\theta_{k+1}] = \theta_{k+1}'$, the linear term vanishes after taking the expectation of both sides. Combining this with Claim (ii) from Lemma 3, we get

$$\begin{aligned}
&\mathbb{E}[\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta_{k+1})] \\
&= \mathbb{E}[\mathsf{F}_\eta(\theta_k) - \mathsf{F}_\eta(\theta_{k+1}') + \mathsf{F}_\eta(\theta_{k+1}') - \mathsf{F}_\eta(\theta_{k+1})] \\
&\geq \frac{\tau}{2}(1-\sigma^2)\|\theta_k - \theta_{k+1}'\|^2 - \frac{\tau M_k}{2}
\end{aligned}$$

□

Putting the lemmas together, we are ready to prove Theorem 1. Similar to [13], let $a_k = \mathbb{E}[\mathsf{F}_\eta(y_k)] - \mathsf{F}_\eta(y^*)$ be the optiamlity gap at iteration $k$. Canceling $\|\theta_k - \theta_{k+1}'\|$ in Lemma 4, we have

$$a_{k+1} \leq (1-\rho)a_k + \frac{\tau_\eta M}{2} \tag{16}$$

This implies $a_k \leq (1-\rho)^k(a_0 - \frac{\tau_\eta M}{2\rho}) + \frac{\tau_\eta M}{2\rho}$, Where $\rho = \frac{1-\sigma^2}{8}e^{(-24\|c\|_\infty/\eta)}$.

### B. Proof of Theorem 2

First, employing Lemma 1 and the inequality $|e^a - e^b| \leq e^M|a-b|, \forall |a|,|b| \leq M$, we argue that the $\ell_2$ sensitivity of the iteration function, denoted as

$$\begin{aligned}
\mathsf{T}\begin{bmatrix}\varphi\\ \psi\end{bmatrix} &= \begin{bmatrix}\mathsf{T}_1(\psi)(x)\\ \mathsf{T}_2(\varphi,\psi)(y)\end{bmatrix} = \begin{bmatrix}\operatorname{argmin}_\varphi \mathsf{F}_\eta(\varphi,\psi)\\ \operatorname{argmin}_\psi \mathsf{F}_\eta(T_1(\psi),\psi)\end{bmatrix} \\
&= \begin{bmatrix}-\eta\log\mathbb{E}_{\nu_n}[\exp(\frac{\psi(Y)-c(x,Y)}{\eta})]\\ -\eta\log\mathbb{E}_{\mu_n}[\exp(\frac{T_1(\psi)(X)-c(X,y)}{\eta})]\end{bmatrix}
\end{aligned}$$

is $\Delta_2[\mathsf{T}] = \eta\log\left(1 + \frac{4\eta}{n}\|c\|_\infty\exp(\frac{6}{\eta}\|c\|_\infty)\right)$.

Next, leveraging the Gaussian mechanism and the composition theorem from [33], we establish that the algorithm, after $K$ iterations, satisfies $(\alpha, \frac{\alpha KC}{M})$-RDP, which is $(\frac{\alpha KC}{M} + \frac{\log(1/\delta)}{\alpha-1}, \delta)$-DP[33], where $C = \frac{1}{2}\Delta_2[\mathsf{T}]$. Finally, substituting the optimal $\alpha^* = 1 + \sqrt{\frac{\log(1/\delta)M}{KC}}$, we derive the desired result.

### C. Proof of Corollary 1

Let $\underline{\epsilon} = \frac{\log(1/\delta)}{\alpha-1}$, $a_0 = \mathsf{F}_\eta(\theta_0) - \mathsf{F}_\eta^*$ and $\nu = e^{6\|c\|_\infty}$. To ensure privacy constraint, we need $\frac{\alpha KC}{M} + \underline{\epsilon} = \epsilon$. Solving for $K$ and substituting it into the utility function in Theorem 1, we obtain $u(\epsilon,\delta,M) = a_0(1-\rho)^{\frac{(\epsilon-\underline{\epsilon})M}{\alpha C}} + \frac{\tau_\eta M}{2\rho}$. Minimizing this with respect to $M$, we find the optimal $M^*$ such that $(1-\rho)^{\frac{(\epsilon-\underline{\epsilon})M^*}{\alpha C}} = -\frac{\tau_\eta}{2\rho\log(1-\rho)}\frac{\alpha C}{\epsilon-\underline{\epsilon}}$. Substituting $M^*$ back into the utility function, and optimize $\rho$, we get $\sigma = 0$, and

$$u_\eta(\varepsilon,\delta) = u(\epsilon,\delta,M^*) = \frac{\tilde{C}_\eta}{\varepsilon-\underline{\varepsilon}}\left[a_0 - \log\left(\frac{\tilde{C}_\eta}{\varepsilon-\underline{\varepsilon}}\right)\right], \tag{17}$$

where $\tilde{C}_\eta = -2\alpha\nu^{\frac{3}{\eta}}\log\left(1 + \frac{2\eta\nu^{\frac{1}{\eta}}\ln\nu}{3n}\right)/\log\left(1 - \frac{1}{8}\nu^{-\frac{4}{\eta}}\right)$, with $\nu = e^{6\|c\|_\infty}$. Further simplification using $1 - \frac{1}{x} \leq \log x \leq x, \forall x > 0$ yields the bound in $C_\eta$. We omit the computations due to space constraint.

### D. Proof of Theorem 3

*1) Notations and preparation lemmas:* We first introduce related notations.

**Definition 4.** *We define the **linear component of EOT** as $f_c^\eta(D_n) = \langle c, \pi_\eta(D_n)\rangle$, where $\pi_\eta(D_n)$ is the optimal coupling to $\mathsf{OT}_c^\eta(\mu_n,\nu_n)$.*

*Let $P = \{\pi : \pi\mathbf{1} = \mu_n, \pi^T\mathbf{1} = \nu_n\}$, and $V$ be the vertex set of $P$. Entropy radius is $R_H \coloneqq \sup_{\pi,\pi'\in P}\mathsf{H}(\pi) - \mathsf{H}(\pi')$. The suboptimal vertices are $S = \{v \in V \mid \langle c,\pi\rangle > \min_{\pi\in V}\langle c,\pi\rangle\}$ The suboptimality gap $\Delta$ of $\mathsf{OT}_c^0(\mu_n,\nu_n)$ is*

$$\Delta = \min_{\pi\in S}\langle c,\pi\rangle - \min_{\pi\in V}\langle c,\pi\rangle \tag{18}$$

*2) **Proof of** $u_\eta^{(1)}(\varepsilon,\delta)$:* First, we compute the utility and privacy of $M_1 = \langle c,\pi_\eta\rangle + Z, Z \sim \mathrm{Lap}(b)$, respectively for fixed $b,\eta$.

**Theorem 4** (Utility of EOT approximation). *For all $b$ and $\eta \leq \frac{\Delta}{1+\log n}$, the following holds: $\left|\mathbb{E}[M_1] - g_1\right| \leq \underline{u} + \sqrt{2}b$, where $\underline{u} = \Delta\exp\left(-\frac{\Delta}{\eta} + 1 + \log n\right)$*

*Theorem 4*. First, we use asymptotic expansion of EOT around OT [17] and specify some constants. Since $\sum_{ij} \pi_{ij} = 1$, $R_1 := \sup_{\pi \in P} \|\pi\|_1 = 1$. Using the fact that the uniform distribution maximizes entropy for the finite support $\mathcal{Z} = \{(x_i, y_j), i, j = 1, \ldots, n\}$, we have $\mathsf{H}(\pi) \le \log |\mathcal{Z}| = 2 \log n$. Since $P$ is a scaled Birkhoff polytope, $\forall \pi \in P$, $\exists \alpha_i \ge 0$ with $\sum_{i=1}^{n!} \alpha_i = 1$ such that $\pi = \sum_i \alpha_i \pi_i$ for permutation matrices $\pi_i$ where each entry is either 0 or $\frac{1}{n}$. The concavity of $\mathsf{H}(\cdot)$ implies $\mathsf{H}(\pi) = \mathsf{H}(\sum_i \alpha_i \pi_i) \ge \sum_i \alpha_i \mathsf{H}(\pi_i) = \log n$. Combining the upper and lower bounds, we get $R_H \le \log n$. Plugging $R_1, R_H$ in Theorem 5 of [17], we have

$$\left|f_c^\eta(D_n) - g_1\right| \le \Delta \exp\left(-\frac{\Delta}{\eta} + 1 + \log n\right) := \underline{\mathrm{u}}$$

Lastly, triangle inequality implies $\left|\mathbb{E}[M_1(D_n)] - g_1\right| \le \left|f_c^\eta(D_n) - g_1\right| + \sqrt{\mathbb{E}[Z^2]} \le \underline{\mathrm{u}} + \sqrt{2}b$. □

The following privacy guarantee is a direct consequence of Laplace mechanism [12].

**Theorem 5** (Privacy of EOT approximation). *For all $\eta, b > 0$, $M_1$ satisfies $(\frac{\Delta_1[f_c^\eta]}{b}, 0)$-DP.*

Substituting $b$ in the utility function in Theorem 4 using privacy constraint in Theorem 5 $b = \frac{\Delta_1[f_c^\eta]}{\varepsilon}$, we have $u_\eta^{(1)}(\epsilon, \delta) = \underline{\mathrm{u}} + \frac{\sqrt{2}\Delta_1[f_c^\eta]}{\epsilon}$. In the remaining, we specify the function sensitivity $\Delta_1[f_c^\eta]$.

**Lemma 5** (Sensitivity of the $f_c^\eta$). *For all $\eta > 0$, $f_c^\eta$ in Definition 4 satisfies*

$$\Delta_1[f_c^\eta] \le \frac{2\|c\|_\infty}{n} + \min\{\eta \log n, \|c\|_\infty, \frac{1}{\sqrt{2n\eta}} \|c\|_\infty^{\frac{3}{2}} e^{\frac{\|c\|_\infty}{\eta}}\}$$

*Lemma 5*. Assume without loss of generality that in $D_n'(\sim D_n)$, $x_1$ changes to $x_1'$. Let $c = [c_{ij}]$ and $\tilde{c} = [c_{ij}']$ where $c_{1j}$ is replaced by $c_{1j}' = c(x_1', y_j)$ for all $j$, with all other entries unchanged. Then $\|c - \tilde{c}\|_{L_1} := \frac{1}{n^2} \sum_{i,j} |c_{ij} - c_{ij}'| \le \frac{1}{n} \|c\|_\infty$. Let $\pi_\eta, \tilde{\pi}_\eta$ be the optimal coupling to $\mathsf{OT}_c^\eta(\mu_n, \nu_n), \mathsf{OT}_{\tilde{c}}^\eta(\mu_n, \nu_n)$. Using Proposition 3.12 from [18] with $p = 1$, $c = \frac{c}{\eta}$, $\mu_1 = \mu_2 = [\frac{1}{n}, \cdots, \frac{1}{n}]$. we have $\|\pi_\eta - \tilde{\pi}_\eta\|_1 \le \frac{1}{\sqrt{2\eta}} \|c - \tilde{c}\|_{L_1}^{\frac{1}{2}} e^{\frac{\|c\|_\infty}{\eta}}$. Hence

$$\begin{aligned}
\Delta_1[f_c^\eta] =& \sup_{\tilde{c}} |\langle c, \pi_\eta\rangle - \langle \tilde{c}, \tilde{\pi}_\eta\rangle| \\
\le& \sup_{\tilde{c}} |\langle c, \pi_\eta - \tilde{\pi}_\eta\rangle| + |\langle c - \tilde{c}, \tilde{\pi}_\eta\rangle| \\
\le& \|c\|_\infty \|\pi_\eta - \tilde{\pi}_\eta\|_1 + \frac{1}{n}\|c\|_\infty \\
\le& \frac{1}{\sqrt{2n\eta}} \|c\|_\infty^{\frac{3}{2}} e^{\frac{\|c\|_\infty}{\eta}} + \frac{1}{n}\|c\|_\infty
\end{aligned}$$

To enhance the result, we argue as follows. Let $h(D_n) = \eta \mathsf{H}(\pi_\eta)$. Then $\Delta_1[h] \le \min\{\eta R_H, \mathsf{OT}_c^0(\mu_n, \nu_n)\} \le \min\{\eta \log n, \|c\|_\infty\}$. Using similar ideas in theorem 3.7 of [18], we can prove $\Delta_1[\mathsf{OT}_c^\eta] \le \frac{2\|c\|_\infty}{n}$ through shadow coupling. Since $f_c^\eta(D_n) = \mathsf{OT}_c^\eta(\mu_n, \nu_n) - h(D_n)$, triangle inequality implies

$$\Delta_1[f_c^\eta] \le \Delta_1 h + \Delta_1 \mathsf{OT}_c^\eta \le \frac{2\|c\|_\infty}{n} + \min\{\eta \log n, \|c\|_\infty\}$$

Combining the above yields the desired result. □

**Remark 5.** *When $\eta$ is large, the sensitivity of linear component of EOT is upper bounded by $O\left(\frac{1}{\sqrt{n}}\right)$. When $\eta$ is small, the sensitivity remains bounded at $O\left(\frac{1}{n}\right)$.*

*3) **Proof of** $u_\eta^{(2)}(\varepsilon, \delta)$:* With the same privacy guarantee and a different utility function, the analysis is similar to PUT of noisy Sinkhorn in Corollary 1. Due to space limit, we only presents the new utility function.

Let $\theta_k = (\varphi_k, \psi_k) \in \mathbb{R}_n \times \mathbb{R}_n$ be the $k^{th}$ iterate of Algorithm 1, $(\pi_k)_{ij} = \exp\left(\frac{(\varphi_k)_i + (\psi_k)_j - c_{ij}}{\eta}\right)$. Let $K_{ij} = e^{-c_{ij}/\eta}$, then $\pi_k = diag(e^{\varphi_k/\eta}) K diag(e^{\psi_k/\eta})$. The following theorem states the utility of computing the linear component of EOT via noisy Sinkhorn.

**Theorem 6** (Utility of noisy Sinkhorn). *For any $0 \le \sigma^2 < 1$, $M \ge 0$, $\theta_0 \in \mathbb{R}^n \times \mathbb{R}^n$ and $k \in \mathbb{N}$, the following holds*

$$\mathbb{E}[\langle c, \pi_k - \pi_\eta\rangle] \le \tilde{C}_\eta \sqrt{\frac{2}{\tau_\eta} a_0 (1 - \rho)^k + \frac{M}{\rho}}$$

*With $\rho = \frac{(1-\sigma^2)}{8} e^{(-24\|c\|_\infty)}, a_0 = \mathsf{F}_\eta(\theta_0) - \mathsf{F}_\eta(\theta^*), \tilde{C}_\eta = \frac{\sqrt{2}n\|c\|_\infty}{\eta} \cdot e^{\frac{6\|c\|_\infty}{\eta}}$*

*Theorem 6*. First, we upper bound the optimality gap by the $l_2$ distance between the couplings.

$$\langle c, \pi_k - \pi_\eta\rangle \le \|c\|_2 \|\pi_k - \pi_\eta\|_2 \le n\|c\|_\infty \|\pi_k - \pi_\eta\|_2$$

Then we upper bound coupling distance via $\theta_k$. Note that $\|K diag(e^{(\frac{\psi_k}{\eta})})\|_\infty \le e^{(\frac{\|\psi_k\|_\infty + \|c\|_\infty}{\eta})}$, and that $\|e^{(\frac{\varphi_k}{\eta})} - e^{(\frac{\varphi^*}{\eta})}\|_2 \le \frac{1}{\eta} e^{(\frac{\|\varphi_k\|_\infty}{\eta})} \|\varphi_k - \varphi^*\|_2$, we have

$$\begin{aligned}
&\|\pi_k - \pi_\eta\|_2 \le \|K e^{\frac{\psi_k}{\eta}}\|_\infty \|e^{\frac{\varphi_k}{\eta}} - e^{\frac{\varphi^*}{\eta}}\|_2 \\
&+ \|K e^{\frac{\varphi^*}{\eta}}\|_\infty \|e^{\frac{\psi_k}{\eta}} - e^{\frac{\psi^*}{\eta}}\|_2 \\
&\le \frac{1}{\eta} e^{\frac{6\|c\|_\infty}{\eta}} (\|\varphi_k - \varphi^*\|_2 + \|\psi_k - \psi^*\|_2) \\
&\le \frac{\sqrt{2}}{\eta} e^{\frac{6\|c\|_\infty}{\eta}} (\|\varphi_k - \varphi^*\|_2^2 + \|\psi_k - \psi^*\|_2^2)^{\frac{1}{2}}
\end{aligned} \tag{19}$$

where the last step follows from $a + b \le \sqrt{2(a^2 + b^2)}$

Using strong convexity in Lemma 2 around $\theta^*$ and $\nabla \mathsf{F}_\eta(\theta^*) = 0$,

$$\begin{aligned}
&\mathbb{E}[\|\varphi_k - \varphi^*\|_2^2 + \|\psi_k - \psi^*\|_2^2] \\
=& \mathbb{E}[\frac{\tau_\eta}{2} \|\theta_k - \theta^*\|_2^2] \le \mathbb{E}[\mathsf{F}_\eta(\theta_k)] - \mathsf{F}_\eta(\theta^*) \\
=& \mathbb{E}[\mathsf{F}_\eta(\theta_k)] + \mathsf{OT}_c^\eta(\mu_n, \nu_n) - \eta
\end{aligned}$$

Combining the above with Theorem 1, we obtain the desired result. □

## References


[1] C. Villani, *Optimal Transport: Old and New*, ser. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008. [Online]. Available: https://books.google.com/books?id=hV8o5R7_5tkC

[2] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning*, Sydney, Australia, 2017, pp. 214–223.

[3] N. Courty, R. Flamary, A. Habrard, and A. Rakotomamonjy, "Joint distribution optimal transportation for domain adaptation," *Advances in neural information processing systems*, vol. 30, 2017.

[4] G. Carlier, V. Chernozhukov, and A. Galichon, "Vector quantile regression: An optimal transport approach," *The Annals of Statistics*, vol. 44, no. 3, pp. 1165 – 1192, 2016. [Online]. Available: https://doi.org/10.1214/15-AOS1401

[5] P. Ghosal and B. Sen, "Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing," *The Annals of Statistics*, vol. 50, no. 2, pp. 1012–1037, 2022.

[6] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the fokker–planck equation," *SIAM journal on mathematical analysis*, vol. 29, no. 1, pp. 1–17, 1998.

[7] N. Fournier and A. Guillin, "On the rate of convergence in wasserstein distance of the empirical measure," *Probability theory and related fields*, vol. 162, no. 3-4, pp. 707–738, 2015.

[8] G. Peyré and M. Cuturi, *Computational Optimal Transport: With Applications to Data Science*, ser. Foundations and trends in machine learning, 2019. [Online]. Available: https://books.google.com/books?id=J0BiwgEACAAJ

[9] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," *Advances in neural information processing systems*, vol. 26, 2013.

[10] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré, "Sample complexity of sinkhorn divergences," in *The 22nd international conference on artificial intelligence and statistics*. PMLR, 2019, pp. 1574–1583.

[11] G. Mena and J. Niles-Weed, "Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[12] C. Dwork, "Differential privacy," *International Colloquium on Automata, Languages, and Programming*, pp. 1–12, 2006. [Online]. Available: https://link.springer.com/chapter/10.1007/11787006_1

[13] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018. [Online]. Available: https://doi.org/10.1137/16M1080173

[14] S. Gopi, Y. T. Lee, D. Liu, R. Shen, and K. Tian, *Private Convex Optimization in General Norms*, pp. 5068–5089. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611977554.ch185

[15] G. Carlier, "On the linear convergence of the multi-marginal Sinkhorn algorithm," *SIAM Journal on Optimization*, vol. 32, no. 2, pp. 786–794, 2022. [Online]. Available: https://hal.science/hal-03176512

[16] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, Aug. 2017. [Online]. Available: http://dx.doi.org/10.1109/CSF.2017.11

[17] J. Weed, "An explicit analysis of the entropic penalty in linear programming," 2018. [Online]. Available: https://arxiv.org/abs/1806.01879

[18] S. Eckstein and M. Nutz, "Quantitative stability of regularized optimal transport and convergence of sinkhorn's algorithm," *SIAM Journal on Mathematical Analysis*, vol. 54, no. 6, pp. 5922–5948, 2022. [Online]. Available: https://doi.org/10.1137/21M145505X

[19] J. Hsu, A. Roth, T. Roughgarden, and J. Ullman, *Privately Solving Linear Programs*. Springer Berlin Heidelberg, 2014, p. 612–624. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-43948-7_51

[20] A. M. Medina, U. Syed, S. Vassilvitskii, and E. Vitercik, "Private optimization without constraint violations," in *International Conference on Artificial Intelligence and Statistics*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220302679

[21] S. Gopi, Y. T. Lee, D. Liu, R. Shen, and K. Tian, *Private Convex Optimization in General Norms*, pp. 5068–5089. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611977554.ch185

[22] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds." USA: IEEE Computer Society, 2014, p. 464–473. [Online]. Available: https://doi.org/10.1109/FOCS.2014.56

[23] S. Gopi, Y. T. Lee, and D. Liu, "Private convex optimization via exponential mechanism," in *Proceedings of Thirty Fifth Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, vol. 178. PMLR, 2022, pp. 1948–1989. [Online]. Available: https://proceedings.mlr.press/v178/gopi22a.html

[24] N. Ponomareva, H. Hazimeh, A. Kurakin, Z. Xu, C. Denison, H. B. McMahan, S. Vassilvitskii, S. Chien, and A. G. Thakurta, "How to dp-fy ml: A practical guide to machine learning with differential privacy," *Journal of Artificial Intelligence Research*, vol. 77, 2023. [Online]. Available: http://dx.doi.org/10.1613/jair.1.14649

[25] N. LeTien, A. Habrard, and M. Sebban, "Differentially private optimal transport: Application to domain adaptation." International Joint Conferences on Artificial Intelligence Organization, 2019, pp. 2852–2858. [Online]. Available: https://doi.org/10.24963/ijcai.2019/395

[26] A. Rakotomamonjy and L. Ralaivola, "Differentially private sliced wasserstein distance," ser. Proceedings of Machine Learning Research, vol. 139. PMLR, 2021, pp. 8810–8820. [Online]. Available: http://proceedings.mlr.press/v139/rakotomamonjy21a.html

[27] T. Cao, A. Bie, A. Vahdat, S. Fidler, and K. Kreis, "Don't generate me: Training differentially private generative models with sinkhorn divergence," in *Advances in Neural Information Processing Systems*, vol. 34. Curran Associates, Inc., 2021, pp. 12 480–12 492. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/67ed94744426295f96268f4ac1881b46-Paper.pdf

[28] J. Hughes and J. Chen, "Differentially private admm-based distributed discrete optimal transport for resource allocation," 2022. [Online]. Available: https://par.nsf.gov/biblio/10387966

[29] J. Hughes, "An algorithmic foundation for fair, secure, and differentially private distributed discrete optimal transport," Ph.D. dissertation, 2021. [Online]. Available: https://www.proquest.com/dissertations-theses/algorithmic-foundation-fair-secure-differentially/docview/2599103387/se-2

[30] B. Schmitzer, "Stabilized sparse scaling algorithms for entropy regularized transport problems," *SIAM Journal on Scientific Computing*, vol. 41, no. 3, pp. A1443–A1481, 2019. [Online]. Available: https://doi.org/10.1137/16M1106018

[31] S. D. Marino and A. Gerolin, "An optimal transport approach for the schrödinger bridge problem and convergence of sinkhorn algorithm," *Journal of Scientific Computing*, vol. 85, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:208139296

[32] E. Hazan and S. Kale, "Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization," *Journal of Machine Learning Research*, vol. 15, no. 71, pp. 2489–2512, 2014. [Online]. Available: http://jmlr.org/papers/v15/hazan14a.html

[33] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, Aug. 2017. [Online]. Available: http://dx.doi.org/10.1109/CSF.2017.11