Minimax Private Estimation of Smooth Optimal-Transport Maps

Clément Lalanne
ANITI, Univ Toulouse,
INSA Toulouse, CNRS, IMT
Toulouse, France
clement.lalanne@math.univ-toulouse.fr

David Rodríguez-Vítores
Universidad de Valladolid, IMUVA
Valladolid, Spain

Franck Iutzeler
ANITI, Univ Toulouse,
INSA Toulouse, CNRS, IMT
Toulouse, France

Jean-Michel Loubes
ANITI Université Toulouse
Equipe REGALIA INRIA Bordeaux
France

Abstract

We study the problem of estimating smooth optimal transport (OT) maps between two probability distributions under differential privacy (DP) constraints. Leveraging wavelet-based density estimators and recent stability bounds for smooth OT maps, we propose differentially private estimators that apply to both central and local DP models. Our main estimator achieves near-minimax optimal rates in dimension $d \geq 2$, and we complement it with a quantile-based estimator that attains minimax optimal rates in dimension $d = 1$ under central DP. We further establish matching minimax lower bounds, confirming the near-optimality of our approach. To the best of our knowledge, this constitutes the first differentially private procedure for OT map estimation with minimax optimality guarantees.

1 Introduction

Working with real-world data introduces challenges that go beyond solving the problem at hand; in particular, one may wish to protect it against privacy attacks [82, 7, 46, 35, 57, 78, 83, 104, 111, 105]. To this end, Differential Privacy (DP) [38] has become the gold standard and is now widely employed by prominent institutions [1, 40, 107, 33].

Formally, differential privacy [39] comes in two main flavors: *central* and *local*. Let $\epsilon > 0$ be a *privacy budget*. In the *central model*, a randomized *mechanism* $\mathcal{M}$ acting on the full dataset D is said to be ϵ-differentially private if, for any pair of neighboring datasets $D \sim D'$ (where the neighboring relation is left general at this point) and any measurable set A, $\mathbb{P}(\mathcal{M}(D) \in A) \leq e^{\epsilon}\,\mathbb{P}(\mathcal{M}(D') \in A)$. In contrast, in the *local model*, each data point X_i is privatized *before* aggregation through a randomized mapping $Q(\cdot \mid X_i)$, often referred to as a *channel*, satisfying $Q(A|x,z) \leq e^{\epsilon} Q(A|x',z)$ for all x, x', A, z, where z is a variable used to model the *context*, which may for instance contain information about the privatized data of previous users (allowing, for instance, for interactive mechanisms). We refer the reader to [16] or [36] for more details. While central DP requires a trusted aggregator, local DP enforces privacy at the level of individual samples, typically at the cost of a more stringent privacy-utility tradeoff. Private mechanisms (or channels) enjoy many desirable properties regarding composition and post-processing. We refer the reader to textbooks such as [39] for further details.

A flourishing line of work has studied the problem of performing statistics under differential privacy constraints [32, 66, 14, 15]. A central finding across this literature is that privacy is usually not

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

free [60, 12, 64, 2, 3, 13, 17, 61, 100, 62, 63, 21, 10]: enforcing differential privacy constraints degrades statistical accuracy, and one of the main goals of the field is to precisely quantify this cost, typically by characterizing the minimax estimation rate under privacy constraints and comparing it to its unconstrained counterpart. In particular, in private nonparametric statistics [112, 9, 73, 72, 41, 67, 36, 37, 16, 70, 95, 54, 11, 53], the privacy overhead can take many different rates based on the underlying *smoothness* of the problem. The present paper contributes to this litterature by addressing a problem that has so far remained largely open: the private estimation of smooth optimal transport maps.

Optimal Transport (OT) has emerged as both a powerful analytical framework [110] and a practical tool, finding applications across a wide range of data-driven tasks in Computer Science [42, 76, 101, 102], Machine Learning [4, 5, 6, 18, 50, 44, 47, 51, 59, 80, 96, 103, 77] and Statistics [29, 30, 19, 31, 69, 71, 85, 90, 91, 98, 106, 113, 115]. In many of these applications, OT maps are not only a theoretical construct but a concrete computational object that is explicitly or implicitly estimated and deployed, for instance in domain adaptation [27], generative modeling [93], and fairness-aware learning [49]. As these applications routinely involve sensitive personal data, the question of how to estimate OT maps while providing formal privacy guarantees arises naturally, and motivates the present work.

Given two probability distributions P_X and P_Y on $\mathbb{R}^d$, the *Wasserstein distance* between them is defined, for any $p \geq 1$, by

$$W_p^p(P_X, P_Y) := \inf_{\pi \in \Pi(P_X, P_Y)} \int \|x - y\|^p \, d\pi(x, y), \tag{1}$$

where $\Pi(P_X, P_Y)$ denotes the set of couplings between P_X and P_Y, that is, the set of joint probability distributions with marginals P_X and P_Y. This distance naturally arises from the problem of optimally *transporting* mass from P_X to P_Y via the so-called *OT maps*, defined as

$$T_0 \in \operatorname*{arg\,min}_{T:\, T\#P_X = P_Y} \int \|T(x) - x\|^p \, dP_X(x), \tag{2}$$

where $T\#P$ denotes the *push-forward* of P by T. In particular, a central result in optimal transport theory (often referred to as Brenier's theorem [94, Sec. 1.3.1]) ensures that when $p = 2$, and both P_X and P_Y are absolutely continuous with respect to the Lebesgue measure and have finite second moments, the optimization problems Equation (1) and Equation (2) are equivalent. Moreover, the optimal coupling is uniquely defined almost everywhere, its support coincides almost everywhere with the graph of T_0, and T_0 is the gradient of a convex function. In this case, the map T_0 exists and is unique almost everywhere. Whenever these conditions hold, we will therefore refer to the optimal transport map between the distributions under consideration without further justifying its existence. We refer the reader to the textbooks [110, 94] for general references on optimal transport.

In this article, we are interested in estimating the OT map T_0 between two distributions P_X and P_Y known only through samples, a problem commonly referred to as *statistical OT*. This problem has attracted considerable attention, in particular in connection with minimax estimation under smoothness assumptions [20, 28, 58, 88, 81, 52, 79, 34, 8]. Indeed, without smoothness, the problem suffers from the curse of dimensionality, whereas smoothness helps alleviate this issue. As we shall see later in this article, privacy only amplifies this dependence, which further motivates the use of smoothness assumptions to control the privacy noise.

1.1 Related work

Statistical optimal transport has a rich literature [45, 86, 99, 23, 24, 25]. The most closely related line of work concerns the estimation of OT maps from samples [58, 79, 88].

The non-private minimax rate for estimating smooth OT maps was first characterized by [58], which established that, under appropriate assumptions, the $L^2(P_X)$ error of a transport map estimator is controlled by the suboptimality of its associated potential in the semi-dual problem of Equation (1). Since this semi-dual objective admits a natural empirical counterpart, the authors derived minimax optimal estimators by combining tools from empirical process theory with an approximation strategy that restricts potentials to suitable finite-dimensional approximation spaces.

Another strategy for OT map estimation is to use *plug-in* estimators, which construct transport maps from other intermediate quantities built from the samples, such as empirical OT plans or density

estimators [79, 88]. Our work falls within this framework: we estimate OT maps by post-processing private quantities derived from the data. A key ingredient in our analysis is a recent stability bound from [8], which links the error in transport map estimation to that of estimating the source and target densities.

To our knowledge, the only prior work on privately estimating OT maps is [75], which proposes a semi-dual estimator building on [58]. The central observation is that the semi-dual objective of [58] has tame sensitivity [38], enabling a private selection mechanism that identifies a good Brenier potential from a finite candidate set and recovers an OT map by taking its gradient. Combined with a covering of the space of admissible potentials, this yields the first upper bound for the problem.

This approach has two significant limitations. First, it is computationally intractable: constructing the covering required by the estimator is infeasible in practice. Consequently, the authors benchmark only the *selection* step on a *simplified* problem instance where the solution is known to belong to a low-dimensional parametric family. Second, the method is suboptimal in every regime: although the paper establishes both upper and lower bounds, a systematic gap persists in the privacy term, leaving the minimax rate unresolved.

Optimal transport also appears in several works under differential privacy, either as a problem-specific tool [89, 55, 97, 108, 92] or as a component of novel privacy paradigms [87, 68, 114]. None of these works are directly related to ours.

1.2 Contributions

The main contributions of this article are as follows.

- We introduce a private plug-in method for OT-map estimation based on private density estimation, which applies under both central and local differential privacy. In the local model, to the best of our knowledge, this is the first method for private OT-map estimation.
- Using the stability bounds of [8], we show that this method is (near) minimax optimal for $d \geq 2$ under both privacy models. In particular, this closes the gap between upper and lower bounds left by [75].
- For $d = 1$, we reduce OT-map estimation to quantiles estimation, yielding a general plug-in approach based on private quantiles estimators. Combined with [65], this gives a near minimax optimal mechanism under central DP, leaving only the local-DP case open.
- We provide numerical experiments illustrating the benefit of smoothness for private OT-map estimation, and highlight the computational limitations of the method.

A comparison with prior work is given in Table 1.

Table 1: Comparison with prior work. Rates of works assuming smoothness α of the map are written with $\alpha = s + 1$ (see Section 3).

Work	Privacy	(Near) Optimality	Estimation Rate
[58, 79, 8]	No Privacy	YES	$\tilde{O}\left(n^{-1} + n^{-\frac{2(s+1)}{2s+d}}\right)$
[75]	Central DP	NO	$\tilde{O}\left(n^{-1} + n^{-\frac{2(s+1)}{2s+d}} + (n\epsilon)^{-\frac{2(s+1)}{2s+d+2}}\right)$
This Work	Central & Local DP	YES Except under LDP when $d = 1$[1]	$\tilde{O}\left(n^{-1} + n^{-\frac{2(s+1)}{2s+d}} + \pi_{n,\epsilon}^{-\frac{2(s+1)}{s+d}}\right)$

Here, $\pi_{n,\epsilon} := n\epsilon$ when working with ϵ-central DP and $\pi_{n,\epsilon} := \sqrt{n}\epsilon$ when working with ϵ-local DP.

[1] Under LDP, $d = 1$, we show an estimation rate in $O\left(n^{-1} + n^{-\frac{2(s+1)}{2s+d}} + (\sqrt{n}\epsilon)^{-\frac{2(s+1)}{s+\frac{3}{2}}}\right)$.

1.3 Organization of the article

The remainder of the article is organized as follows. In Section 2, we introduce the private querying mechanisms used throughout the paper. Then, in Section 3, we present our main results on the private estimation of smooth OT maps. A discussion of optimality, together with broader implications of our results, is given in Section 4. Finally, numerical experiments are reported in Section 5.

Notations and Conventions. $\Omega = [0,1]^d$ is the support of the distributions for the entire article. When considering differential privacy, it is convenient to define $\pi_{n,\epsilon} := n\epsilon$ when working with ϵ-central DP and $\pi_{n,\epsilon} := \sqrt{n}\epsilon$ when working with ϵ-local DP. In any case, the asymptotic regimes are considered when $\min(n, \pi_{n,\epsilon}) \to +\infty$. For any $k \in \mathbb{N}$, $\mathcal{C}^k(\mathcal{S})$ denotes the set of functions on a space $\mathcal{S}$ that are k times continuously differentiable, and $\mathcal{C}^\infty(\mathcal{S})$ is defined as $\cap_{k\in\mathbb{N}}\mathcal{C}^k(\mathcal{S})$. Whenever applicable, ∇ and ∇^2 are used to refer to gradient and Hessian operators, respectively. For any set S, $|S|$ denotes its cardinality. The Hölder norm of order α on S is denoted by $\|T\|_{C^\alpha(S)}$. Explicitly, given $s > 0$, the Hölder space $\mathcal{C}^s(\Omega)$ is the set of functions f that are uniformly continuous on Ω, $\lfloor s \rfloor$ times differentiable in the interior of Ω with derivatives extending continuously to the boundary, and such that

$$\|f\|_{C^s(\Omega)} := \max_{|\alpha|\le\lfloor s\rfloor} \sup_{x\in\Omega} |\partial^\alpha f(x)| + \max_{|\alpha|=\lfloor s\rfloor} \sup_{x\neq y} \frac{|\partial^\alpha f(x) - \partial^\alpha f(y)|}{\|x-y\|^{s-\lfloor s\rfloor}} \tag{3}$$

is finite, where the maxima are taken over multi-indices $\alpha \in \mathbb{N}^d$ and where the second term is dropped when s is an integer. In particular, for an integer s, this definition agrees with the notation $\mathcal{C}^k$ above. For a measure μ, $\|\cdot\|_{L^2(\mu)}$ is the usual L^2 norm with reference measure μ, and for a vector-valued map $S : \Omega \to \mathbb{R}^d$, $\|S\|^2_{L^2(\mu)} := \int_\Omega \|S(x)\|_2^2 d\mu(x)$, as in [58, 79]. For the full construction of the functional spaces, norm and wavelet system, we refer the reader to Appendix A from [79] and we recall below the main notations and conventions for this article. Unless explicitly mentioned otherwise, Ψ refers to the boundary-corrected wavelet system on Ω built from the compactly-supported N-th Daubechies scaling and wavelet functions that are r times continuously differentiable where $r = 0.18(N-1)$ for $N \ge 2$. $\mathfrak{B}^s_{p,q}$ and $\|\cdot\|_{\mathfrak{B}^s_{p,q}}$ are respectfully used to denote the Besov space and the Besov norm and similarly whenever we work in a Hölder space of regularity α or Besov space of regularity s, we implicitly assume that r is strictly greater than $\lceil s \rceil$ or α. For a density of probability f we may abuse the notation and refer to the induced probability measure by f as well. $\mathrm{d_H}$ is the Hamming distance.

On the constants. This article will adopt, unless specifically mentioned otherwise, the usual nonparametric convention of considering as a constant any quantity that does not depend on the sample size n nor on the privacy budget ϵ.

2 Tools from Private Density Estimation

This section introduces the private density estimation primitives used in our main estimator later in the article. Throughout, f_Z denotes a probability density on Ω, assumed to belong to $L^2(\Omega)$, from which we observe n i.i.d. samples $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} f_Z$. In the central model, we adopt the replacement neighboring relation, meaning that $(Z_1, \dots, Z_n) \sim (Z'_1, \dots, Z'_n)$ whenever the two datasets differ in exactly one coordinate.

Let $\Psi := \Phi \cup \bigcup_{j=j_0}^\infty \Psi_j$ be an orthonormal system of functions, where Φ and, for each j, Ψ_j are sets of functions on Ω. This notation is chosen because, later in the article, we specialize this family to a system of boundary-corrected Daubechies wavelets; we refer to [79] for additional background on this construction. A projection estimator of the density f_Z is then given by $\tilde{f}_Z := \sum_{\xi\in\Phi} \hat{\beta}_\xi \xi + \sum_{j=j_0}^J \sum_{\xi\in\Psi_j} \hat{\beta}_\xi \xi$ where, for every ξ, $\hat{\beta}_\xi := \frac{1}{n}\sum_{i=1}^n \xi(Z_i)$, and where J is a tunable maximum resolution level. Since the final estimator must be a *proper* density, we also define $\hat{f}_Z := (\tilde{f}_Z \vee 0)/(\int_\Omega (\tilde{f}_Z \vee 0))$, which is pointwise non-negative and integrates to 1. In the context of differential privacy, the collection of coefficients $\hat{\beta}_\xi$ can be viewed as a data-dependent query that must be privatized.

We impose the following assumption on the wavelet system.

Algorithm 1 Querying mechanism, central DP

Input: data $(Z_i; 1 \leq i \leq n)$, privacy budget $\epsilon > 0$, constants C_1, C_2, C_3, C_4 from Assumption 1, Depth J.
Compute Δ given in Proposition 2.1.
Query and Publish: the vector of private estimated decomposition coefficients $(\hat{\beta}_{\xi,\text{priv}}; \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j) := (\frac{1}{n}\sum_{i=1}^{n} \xi(Z_i) + \frac{\Delta}{n\epsilon}\mathcal{L}(1); \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j)$.

Algorithm 2 Querying mechanism, local DP

Input: data $(Z_i; 1 \leq i \leq n)$, privacy budget $\epsilon > 0$, constants C_1, C_2, C_3, C_4 from Assumption 1, Depth J.
Compute Δ given in Proposition 2.1.
for every individual $1 \leq i \leq n$ **do**
Query: the privatized individual contribution to the coefficients $(\hat{\beta}^{(i)}_{\xi,\text{priv}}; \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j) := (\xi(Z_i) + \frac{\Delta}{\epsilon}\mathcal{L}(1); \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j)$.
end for
Publish: the vector of private estimated decomposition coefficients $(\hat{\beta}_{\xi,\text{priv}}; \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j) := (\frac{1}{n}\sum_{i=1}^{n} \hat{\beta}^{(i)}_{\xi,\text{priv}}; \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j)$.

Assumption 1 (Wavelet scaling and support partition)**.** There exist constants $C_1, C_2, C_3, C_4, C_5 \geq 0$ such that (i) $|\Phi| \leq C_1$ and $\sup_{\xi\in\Phi} \|\xi\|_\infty \leq C_2$; (ii) for all $j \geq j_0$ and all $\xi \in \Psi_j$, there exists a rectangle $I_\xi \subseteq \Omega$ that contains the support of ξ; (iii) for all $j \geq j_0$, $\|\sum_{\xi\in\Psi_j} \mathbb{1}(\cdot \in I_\xi)\|_\infty \leq C_3$; (iv) $\sup_{j\geq j_0} \sup_{\xi\in\Psi_j} 2^{-jd/2}\|\xi\|_\infty \leq C_4$; (v) for all $j \geq j_0$, $|\Psi_j| \leq C_5 2^{jd}$.

Under this assumption, we obtain the following sensitivity bound.

Proposition 2.1 (Sensitivity)**.** *If Ψ satisfies Assumption 1, the query $q : z \mapsto (\xi(z); \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j)$ satisfies* $\sup_{z,z'\in\Omega} \|q(z) - q(z')\|_1 \leq \underbrace{2C_1C_2 + 6C_3C_4 2^{(J+1)d/2}}_{=:\Delta}$.

Proof. See Section B.1. □

On Assumption 1. As detailed in [79], boundary-corrected Daubechies wavelets satisfy Assumption 1. In this article, we therefore fix Ψ to be this system. Under this choice, Assumption 1 is not restrictive, since we explicitly work with a family of functions known to satisfy it. The assumption is included mainly for modularity, as it allows the privacy argument to be stated for alternative approximation spaces satisfying the same structural properties. The corresponding utility guarantees, however, may depend on the specific system under consideration. In other words, the privacy guarantees of Theorem 2.1 hold for any system satisfying Assumption 1 and for any data distribution, while the utility guarantees of Theorems A.1 and 3.1 use both Assumption 1 and properties that are specific to the boundary-corrected Daubechies system, and are only claimed for this system.

The estimator $\hat{f}_Z$ is not differentially private, and must therefore be privatized. This is done in Algorithm 1 and Algorithm 2 through the definition of the private coefficients $\beta_{\xi,\text{priv}}$.

As above, we also define the corresponding possibly improper and proper density estimators $\tilde{f}_{Z,\text{priv}} := \sum_{\xi\in\Phi} \hat{\beta}_{\xi,\text{priv}}\xi + \sum_{j=j_0}^{J}\sum_{\xi\in\Psi_j} \hat{\beta}_{\xi,\text{priv}}\xi$ and $\hat{f}_{Z,\text{priv}} := (\tilde{f}_{Z,\text{priv}} \vee 0)/(\int_\Omega(\tilde{f}_{Z,\text{priv}} \vee 0))$.

Their privacy guarantees are stated in Theorem 2.1.

Theorem 2.1 (Privacy of Algorithm 1 and Algorithm 2)**.** *If the wavelet system Ψ satisfies Assumption 1, then Algorithm 1 (resp. Algorithm 2) achieves ϵ-central differential privacy (resp. ϵ-local differential privacy). Furthermore, by the post-processing property of differential privacy, the objects $\tilde{f}_{Z,J,priv}$ and $\hat{f}_{Z,J,priv}$ satisfy the same privacy guarantees.*

Proof. See Section B.2. □

3 Improved Rates for Differentially Private Smooth Optimal Transport Map Estimation

This section introduces and analyzes our two main estimation procedures, namely a plug-in estimator based on private projection density estimation and an estimator based on private quantiles estimation.

Let P_X and P_Y be two probability distributions on a common domain Ω. We assume access to independent samples $X_1, \dots, X_{n_X} \sim P_X$ and $Y_1, \dots, Y_{n_Y} \sim P_Y$, where the X-samples and Y-samples are mutually independent.

Under *central differential privacy*, we define the neighboring relation $\big((X_1, \dots, X_{n_X}), (Y_1, \dots, Y_{n_Y})\big) \sim \big((X'_1, \dots, X'_{n_X}), (Y'_1, \dots, Y'_{n_Y})\big)$ whenever either $(X_1, \dots, X_{n_X}) \sim (X'_1, \dots, X'_{n_X})$ or $(Y_1, \dots, Y_{n_Y}) \sim (Y'_1, \dots, Y'_{n_Y})$, but not both, where the relation $\sim$ on each sample is as defined in the previous section. As in [75], this formulation implicitly assumes that the two samples are obtained through separately identified channels. Likewise, in the *local differential privacy* setting, we assume that the observations are partitioned according to their source distribution before the local privatization channels are applied.

The main estimator. Our first estimator is obtained by plugging private density estimators into the optimal transport problem. Concretely, we construct private density estimators $\hat{f}_{X,J,\text{priv}}$ and $\hat{f}_{Y,J,\text{priv}}$ for the marginals P_X and P_Y using the procedure of Section 2. In particular, the decomposition coefficients are privatized via Algorithm 1 in the central model and via Algorithm 2 in the local model. We then define $\hat{T}_{\text{priv}} := \text{OT-MAP}(\hat{f}_{X,\text{priv}}, \hat{f}_{Y,\text{priv}})$, where $\text{OT-MAP}(\hat{f}_{X,\text{priv}}, \hat{f}_{Y,\text{priv}})$ denotes the optimal transport map between the probability distributions on Ω with densities $\hat{f}_{X,\text{priv}}$ and $\hat{f}_{Y,\text{priv}}$. Since this map is computed solely from the privatized density estimators, its privacy is an immediate consequence of the post-processing property of differential privacy.

Upper-bound. As we seek to leverage the smoothness of the densities in the estimation of transport maps, we work under the following standard assumption from the literature on smooth OT map estimation.

Assumption 2. There exist $s > 0$, $M > 0$ and $\gamma > 1$ such that $f_X, f_Y \in \{f \in \mathcal{C}^s(\Omega) : \|f\|_{C^s(\Omega)} \leq M, \gamma > f > \gamma^{-1}\}$ and there exist $0 < L_1 \leq L_2$ and $\phi : \Omega \to \mathbb{R}$ that is twice continuously differentiable, such that $\nabla\phi(\Omega) \subseteq \Omega$, $L_1 I_d \preceq \nabla^2\phi \preceq L_2 I_d$ pointwise and $\nabla\phi \# P_X = P_Y$.

We can now state our main utility guarantee for the above mechanism.

Theorem 3.1. *If P_X and P_Y satisfy Assumption 2 and $n_X = n_Y = n$, then for a privacy budget ϵ, tuning $2^J \asymp \min\Big(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{\max(d;3/2)+s}}\Big)$ leads to the error*

$$\mathbb{E}\left[\|\hat{T}_{priv} - T\|^2_{L^2(P_X)}\right] \lesssim \begin{cases} \max\Big(n^{-\frac{2(s+1)}{2s+d}}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\Big) & \text{for } d \geq 3 \\ \max\Big(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+2}}\Big) \times \textit{PolyLog}(n,\epsilon) & \text{for } d = 2 \\ \max\Big(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+3/2}}\Big) & \text{for } d = 1 \end{cases}$$

where T is the optimal transport map between P_X and P_Y, and where the constant hidden in $\lesssim$ only depends on d, on the constants s, M, γ, L_1, L_2 of Assumption 2, on Ψ and on the constants hidden in the choice of 2^J. We recall that $\pi_{n,\epsilon} := n\epsilon$ when working under central DP and $\pi_{n,\epsilon} := \sqrt{n}\epsilon$ when working under local DP.

Proof. See Section C.1 □

Idea of the proof. We combine Theorem A.1 with the stability bound of [8, Theorem 3], whose hypotheses only bear on the Brenier potential of (P_X, P_Y) and are thus guaranteed by Assumption 2 alone. The truncation and normalization of the private density estimators are handled on a good event of high probability.

On Assumption 2. Let us discuss the assumptions gathered in Assumption 2. The first is the Hölder regularity of the source and target densities. This is precisely how smoothness is quantified here and should be viewed as a structural feature of the problem rather than as a technical artifact: it is what allows one to mitigate the curse of dimensionality in the rates. The lower and upper bounds on the densities, as well as the bounds on the Hessian of the transport potential, are standard in the literature, both under differential privacy [75] and in the non-private setting [58, 79]. Moreover, it is known that in statistical optimal transport, even without privacy, the rates can differ substantially depending on whether a lower bound on the density is imposed [84]. Removing this assumption makes the problem strictly harder and lies outside the scope of the present paper. In particular, these assumptions are not induced by privacy, but are the standard structural conditions already used in prior work on smooth OT map estimation. Note that the existence of the *transport potential* ϕ is not an assumption, since Brenier's theorem [94, Sec. 1.3.1] ensures that the OT map from P_X to P_Y is the gradient of a convex function ϕ. What Assumption 2 imposes is the curvature bound $L_1 I_d \preceq \nabla^2\phi \preceq L_2 I_d$, which makes the target map well-posed and stable, and which is required by the stability bound of [8]. Finally, Assumption 2 is only used for the statistical accuracy, and the privacy of $\hat{T}_{\text{priv}}$ holds for any distributions on Ω.

Comparison to prior work. A direct comparison with [75] requires some care, since that work assumes smoothness α of the transport map itself, whereas we assume smoothness s of the underlying densities. By Caffarelli-type regularity theory, one may heuristically relate these quantities through $\alpha \approx s+1$, since for densities bounded above and below and lying in $\mathcal{C}^s$, the Brenier potential lies in $\mathcal{C}^{s+2}$ (see [79, Theorem 3]). Under this correspondence, our estimator achieves substantially faster rates than those of [75], namely $n^{-1} + n^{-\frac{2\alpha}{2\alpha-2+d}} + (n\epsilon)^{-\frac{2\alpha}{2\alpha+d}}$ (see also Table 1), since ours read, up to polylogarithmic factors, $n^{-1} + n^{-\frac{2\alpha}{2\alpha-2+d}} + (n\epsilon)^{-\frac{2\alpha}{\alpha-1+d}}$, with the same non-private terms and a faster privacy term. A theorem-level comparison, however, appears difficult at present. Indeed, the available Caffarelli-type regularity results are generally not uniform, and our domain has a non-smooth boundary; see, for instance, the discussion in [79]. More precisely, the global version of Caffarelli's theorem requires a $\mathcal{C}^2$ uniformly convex boundary, which $[0,1]^d$ does not have, its interior version only holds on subdomains compactly contained in Ω whereas our risk is measured up to the boundary, and even on an admissible domain, the global statement yields no bound on the potential that is uniform over a $\mathcal{C}^s$ ball of densities, as a minimax comparison would require. Hence, neither class contains the other. For this reason, and as is common in the smooth OT literature, we only invoke the usual heuristic correspondence between density smoothness and transport-map smoothness rather than claim a formal equivalence. What can be compared is what each work proves inside its own class. The upper and lower bounds of [75] do not match in the privacy term, so that the minimax rate over their class remains open, whereas ours settle it over ours (up to polylogarithmic factors, except under local DP when $d=1$). In general, neither modeling choice strikes us as better, since smoothness of the densities bears on the data-generating distributions, which is the usual starting point of nonparametric statistics, while smoothness of the map bears on the object to estimate, which is more natural when prior knowledge concerns the transformation itself. Assuming smoothness of the densities, however, makes optimal procedures available, and here the optimal procedure is also implementable.

The special case $d=1$. When $d=1$, the rate achieved by the previous estimator deteriorates. At present, we do not know whether this reflects a genuine limitation of the method in one dimension or a suboptimality of our analysis. Fortunately, the one-dimensional setting admits a different approach based on private quantile estimation. Indeed, under mild assumptions, the optimal transport map between f_X and f_Y is given by $T(\cdot) = F_{f_Y}^{-1}(F_{f_X}(\cdot))$, where F_{f_X} is the cumulative distribution function of f_X and $F_{f_Y}^{-1}$ is the quantile function of f_Y. Fix a tuning parameter $m \geq 1$, and define $\mathbf{p} := (1/m, 2/m, \dots, (m-1)/m)$. Suppose that we are given estimators $\mathbf{q}_X$ and $\mathbf{q}_Y$, both of length $m-1$, for the quantiles of orders $\mathbf{p}$ of f_X and f_Y, respectively. We then define the estimator

$$\hat{T}(x) = \begin{cases} \mathbf{q}_Y[1] & \text{if } x \in [0, \mathbf{q}_X[1]], \\ \mathbf{q}_Y[2] & \text{if } x \in (\mathbf{q}_X[1], \mathbf{q}_X[2]], \\ \dots & \\ \mathbf{q}_Y[m-1] & \text{if } x \in (\mathbf{q}_X[m-2], \mathbf{q}_X[m-1]], \\ 1 & \text{if } x \in (\mathbf{q}_X[m-1], 1]. \end{cases} \tag{4}$$

By construction, if $\mathbf{q}_X$ and $\mathbf{q}_Y$ are obtained through an ϵ-private mechanism, then $\hat{T}$ satisfies the same privacy guarantee by post-processing.

For example, by combining this construction with the private quantile mechanism of [65] and tuning m appropriately, we obtain the following result.

Theorem 3.2. *If there exist $0 < a < b$ such that $a \leq f_X, f_Y \leq b$ on Ω, then there exists an ϵ-central DP estimator $\hat{T}$ such that, if asymptotically $\epsilon = \Omega(n^{-1+\kappa})$ for some $\kappa > 0$, then this mechanism has error $\mathbb{E}\left[\|\hat{T} - T\|^2_{L^2(P_X)}\right] \lesssim$ PolyLog$(n, \epsilon) \times \max\left(\frac{1}{n}, \frac{1}{n^2\epsilon^2}\right)$, where the constant hidden in $\lesssim$ only depends on a, b and κ.*

Proof. See Section C.2 □

On the assumptions. The main additional assumption compared to the previous setting is that $\epsilon = \Omega(n^{-1+\kappa})$. We believe this condition is an artifact of our analysis rather than a genuine limitation of the method. Moreover, it is not especially restrictive: by the group privacy property of differential privacy [39], one already expects the signal to be essentially lost when $\epsilon = O(1/n)$. Thus, our condition simply requires staying at a polynomial distance from this degenerate regime. Note that κ is unrelated to the density bounds, and that it is fixed in advance and may be taken as small as one wishes, so that this regime is never empty. Furthermore, apart from this condition, Theorem 3.2 only requires the densities to be bounded above and below, and neither their smoothness nor a curvature bound on the transport potential. Lastly, the privacy of $\hat{T}$ holds for any distributions on Ω, and the conditions of Theorem 3.2 are only used for its statistical accuracy.

4 Optimality and Broader Discussion

The first lower bound for this problem was established in [75]. As discussed above, however, that work assumes smoothness of the transport map itself, whereas our analysis is carried out under smoothness assumptions on the source and target densities. Although these two viewpoints are heuristically related, no formal equivalence is currently available. We therefore state below a lower bound tailored to the hypothesis class considered in this paper. In addition, our result also covers the local-DP setting, which was not addressed in prior work.

Theorem 4.1. *For a fixed set of constants s, M, γ, L_1, L_2 we denote by $\mathcal{M}$ the set of pairs of distributions that satisfy Assumption 2 for these constants. If ϵ is small enough, we have, as soon as $\min(n, \pi_{n,\epsilon})$ is larger than a constant, $\inf_{\hat{T}} \sup_{(P_X,P_Y)\in\mathcal{M}} \mathbb{E}\left[\|\hat{T} - T\|^2_{L^2(P_X)}\right] \gtrsim n^{-1} + n^{-\frac{2(s+1)}{2s+d}} + \pi_{n,\epsilon}^{-\frac{2(s+1)}{s+d}}$, where $\hat{T}$ is taken among the set of ϵ-DP mechanisms (central or local with the adjusted value for $\pi_{n,\epsilon}$) and T is the optimal transport map between P_X and P_Y, and where the constant hidden in $\gtrsim$ only depends on d and on s, M, γ, L_1, L_2.*

Proof. See Section C.3 □

In particular, in the central-DP setting, the rate highlighted by this lower bound matches that of [75] under the heuristic correspondence discussed earlier.

General optimality. The first main consequence of Theorem 4.1 is that, under both central and local differential privacy, the main mechanism of Section 3 is minimax optimal as soon as $d \geq 3$, and near minimax optimal up to polylogarithmic factors when $d = 2$. Moreover, the reduction to quantile estimation introduced in Section 3 is near minimax optimal in dimension 1 under central differential privacy. In all of these regimes, our results provide, to the best of our knowledge, the first (nearly) matching upper and lower bounds for the problem.

The local-DP case when $d = 1$. Under local differential privacy in dimension 1, our results do not allow us to determine whether the upper bound or the lower bound is tight, which remains a limitation of the present work. If we were to speculate, the fact that the lower bound appears robust across privacy models, whereas the upper bound requires a separate and more delicate treatment in dimension 1, suggests that the gap may come from the upper bound rather than from the lower

bound. On the positive side, the central-DP result in dimension 1 relies on recent advances in private estimation of many quantiles [48, 65], and our reduction is modular enough that any future progress on the corresponding problem under local differential privacy could be translated directly into an improved upper bound for private OT map estimation.

The cost of privacy. For simplicity, we focus on central DP, but the discussion can be adapted to local DP for $d \geq 2$ by replacing $n\epsilon$ with $\pi_{n,\epsilon} = \sqrt{n}\epsilon$. Our bounds show that the cost of privacy appears through the additional term $(n\epsilon)^{-\frac{2(s+1)}{s+d}}$, which is added to the usual non-private rate $n^{-1} + n^{-\frac{2(s+1)}{2s+d}}$. In this sense, privacy is *free* whenever the privacy term is asymptotically no larger than the dominant non-private term, and it has to be *paid* once it becomes the leading contribution to the error. For $d \geq 3$, this boundary is obtained by equating the two terms, which yields the threshold $\epsilon \asymp n^{-\frac{s}{2s+d}}$. Thus, if $\epsilon \gg n^{-\frac{s}{2s+d}}$, privacy does not affect the minimax rate, whereas if $\epsilon \ll n^{-\frac{s}{2s+d}}$, the rate becomes privacy-dominated. In dimension $d = 2$, the same interpretation holds up to polylogarithmic factors, with threshold $\epsilon \asymp n^{-\frac{s}{2(s+1)}}$.

This also highlights the role of smoothness. As s increases, the privacy term decays faster and the threshold between the privacy-free and privacy-limited regimes shifts toward smaller values of ϵ. In other words, smoother densities allow one to sustain stronger privacy while still preserving the non-private statistical rate. Conversely, when smoothness is low, the privacy-free regime becomes much smaller, and the statistical cost of privacy is felt earlier.

The dimension jump. A striking consequence of our rates is the following dimension-dependent transition. In the non-private setting, the estimation problem only departs from the parametric rate once $d \geq 3$. Under central differential privacy, however, the corresponding private parametric rate $\frac{1}{n} + \frac{1}{n^2\epsilon^2}$ already ceases to hold as soon as $d \geq 2$. In other words, privacy makes the effect of dimension appear one dimension earlier, a phenomenon which, to the best of our knowledge, has not been documented for related estimation problems. This observation further supports the view that, under privacy constraints, it is already crucial in ambient dimension 2 to use procedures that explicitly exploit smoothness, rather than relying on mechanisms that ignore it.

5 Experiments

This section explains how to turn the estimator of Section 3 into an implementable procedure and provides numerical evidence of its performance.

Algorithmic considerations. The estimator of Section 3 relies on two steps: first, it privately constructs surrogate densities for the two underlying distributions; second, it computes the OT map between these surrogate densities. The first step is directly implementable, since private querying mechanisms were introduced in Section 2. Moreover, any subsequent post-processing of these privatized objects preserves their differential privacy guarantees. The main algorithmic difficulty lies in the second step. Although the OT map between the surrogate densities is well defined at the population level, computing it directly from the analytic form of the private density estimators is not always straightforward in practice. To address this issue, we adopt a resampling approach. More precisely, we first draw an arbitrary number of synthetic samples from each surrogate density, in our implementation using rejection sampling. By post-processing, these synthetic datasets remain differentially private with respect to the original data. We then apply *any off-the-shelf non-private OT map estimator* to the resulting synthetic samples in order to approximate the target transport map. In all of our experiments, we use what is, to the best of our knowledge, the empirically most efficient approach, namely the entropically regularized barycentric projector of [88] (see Eq. (14) therein for a formal definition). Additional implementation details are provided in Section D.

Datasets. To make the error fully observable, we require access to both the true transport map and the underlying source and target distributions, so that we can measure suboptimality without approximation bias from the numerical OT solver. We therefore work with synthetic benchmarks for which sampling is straightforward and OT maps are easily computed. The precise specification of these source–target pairs is given in Section D.

Baselines. Since the method of [75] is computationally intractable, we cannot include it as an experimental baseline. Instead, we compare our method against two non-private baselines in order to assess the empirical cost of privacy: the entropically regularized transport map estimator of [88] (**EOT baseline**), and the simple barycentric projection empirical OT map (**OT baseline**). For the latter baseline, the estimator is only defined on the support of the training samples. As a consequence, we can report only its *training* error. This gives it a slight advantage over the other methods, including ours, and this should be kept in mind when interpreting the results.

Results. Full experimental results are given in Section D; in the main text, we only report Figure 1, which already captures the main behavior of our estimator. Two conclusions stand out. First, the problem is statistically difficult: at small sample sizes, privacy induces a substantial utility loss across all settings. Second, as the sample size grows, smoothness becomes beneficial: methods exploiting smoothness exhibit a steeper slope than the empirical OT baseline, indicating faster convergence, as predicted by the theory. In the $d = 2$ case of Figure 1, we even observe that private smooth estimators can outperform the naive non-private baseline.

Limitations. The main empirical bottleneck of our method is the resampling step. While we do not currently see how to avoid it, improving this component appears to be an important direction for future work.

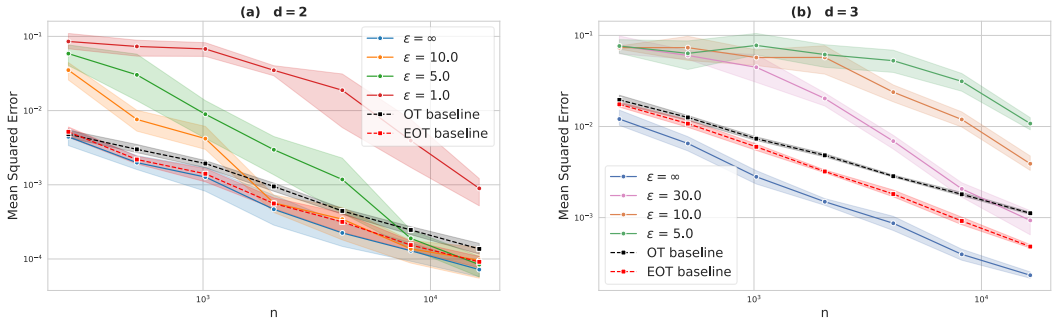

Figure 1: **unif-unif, N=2, J=2, for d=2 (a) and d=3 (b).** Mean squared error vs. sample size n for varying central privacy budgets ϵ (log-log scale), averaged over 5 independent runs. Solid lines represent the ϵ-DP wavelet estimator; black and red dashed lines denote the OT and EOT baselines, respectively.

6 Conclusion

In this work, we studied the private estimation of smooth optimal transport maps and proposed a plug-in approach based on private density estimation. Under standard assumptions, this yields near-minimax guarantees for $d \geq 2$ in both the central and local privacy models, together with a separate one-dimensional central-DP result based on private quantiles and matching lower bounds.

A notable structural feature of the problem is its dimension-dependent transition: in the non-private setting, smooth OT-map estimation remains essentially parametric up to $d = 2$ and loses this behavior from $d \geq 3$, whereas under differential privacy it already departs from the usual private-parametric benchmark $1/n + 1/\pi^2_{n,\epsilon}$ in dimension $d = 2$.

Beyond the statistical guarantees, our approach also suggests a concrete computational route: once private density estimates are released, downstream OT-map computation is protected by the post-processing property of differential privacy. Our current implementation should be viewed as a first proof of concept, and an important direction for future work is to design more efficient procedures that build on similar private surrogates while avoiding the current resampling bottleneck.

Acknowledgments and Disclosure of Funding

This paper has been partially funded by the Agence Nationale de la Recherche under grants ANR-23-CE23-0029 Regul-IA and ANR-24-CE23-1529 MAD. The authors also acknowledge the support of the AI Cluster ANITI (ANR-23-IACL-0002). The second author was partially supported by grant PID2024-162240NB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER, and by Department of Education of the Junta de Castilla y León and FEDER Funds through grant CLU-2025-1-02- IMUVA MCIN/AEI/10. As part of the Madlearning project, this work received funding from the French government, managed by the National Research Agency (ANR) under the France 2030 program, with reference number ANR-25-PEIA-0002.

References

[1] J. M. Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2867–2867, 2018.

[2] J. Acharya, Z. Sun, and H. Zhang. Differentially private Assouad, Fano, and Le Cam. In V. Feldman, K. Ligett, and S. Sabato, editors, *Algorithmic Learning Theory, 16-19 March 2021, Virtual Conference, Worldwide*, volume 132 of *Proceedings of Machine Learning Research*, pages 48–78. PMLR, 2021.

[3] I. Aden-Ali, H. Ashtiani, and G. Kamath. On the sample complexity of privately learning unbounded high-dimensional gaussians. In V. Feldman, K. Ligett, and S. Sabato, editors, *Algorithmic Learning Theory, 16-19 March 2021, Virtual Conference, Worldwide*, volume 132 of *Proceedings of Machine Learning Research*, pages 185–216. PMLR, 2021.

[4] J. Alaux, E. Grave, M. Cuturi, and A. Joulin. Unsupervised hyperalignment for multilingual word embeddings. *CoRR*, abs/1811.01124, 2018.

[5] D. Alvarez-Melis, T. S. Jaakkola, and S. Jegelka. Structured optimal transport. In A. J. Storkey and F. Pérez-Cruz, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, volume 84 of *Proceedings of Machine Learning Research*, pages 1771–1780. PMLR, 2018.

[6] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 2017.

[7] L. Backstrom, C. Dwork, and J. M. Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In C. L. Williamson, M. E. Zurko, P. F. Patel-Schneider, and P. J. Shenoy, editors, *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 181–190. ACM, 2007.

[8] S. Balakrishnan and T. Manole. Stability bounds for smooth optimal transport maps and their statistical implications, 2025.

[9] R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *CoRR*, abs/1412.4451, 2014.

[10] M. Beraha, S. Favaro, and V. Rao. Mcmc for bayesian nonparametric mixture modeling under differential privacy. *arXiv preprint arXiv:2310.09818*, 2023.

[11] T. B. Berrett, L. Györfi, and H. Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15(1):2430 – 2453, 2021.

[12] S. Biswas, Y. Dong, G. Kamath, and J. R. Ullman. Coinpress: Practical private mean and covariance estimation. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

[13] G. Brown, M. Gaboardi, A. D. Smith, J. R. Ullman, and L. Zakynthinou. Covariance-aware private mean estimation without private covariance estimation. In M. Ranzato, A. Beygelzimer, Y. N. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 7950–7964, 2021.

[14] M. Bun, G. Kamath, T. Steinke, and Z. S. Wu. Private hypothesis selection. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 156–167, 2019.

[15] M. Bun, G. Kamath, T. Steinke, and Z. S. Wu. Private hypothesis selection. *IEEE Trans. Inf. Theory*, 67(3):1981–2000, 2021.

[16] C. Butucea, A. Dubois, M. Kroll, and A. Saumard. Local differential privacy: Elbow effect in optimal density estimation and adaptation over besov ellipsoids. *CoRR*, abs/1903.01927, 2019.

[17] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *CoRR*, abs/1902.04495, 2019.

[18] G. D. Cañas and L. Rosasco. Learning probability measures with respect to optimal transport metrics. In P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pages 2501–2509, 2012.

[19] E. Cazelles, V. Seguy, J. Bigot, M. Cuturi, and N. Papadakis. Geodesic PCA versus log-pca of histograms in the wasserstein space. *SIAM J. Sci. Comput.*, 40(2), 2018.

[20] V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-kantorovich depth, quantiles, ranks, and signs, 2015.

[21] J. Chhor and F. Sentenac. Robust estimation of discrete distributions under local differential privacy. In *International Conference on Algorithmic Learning Theory*, pages 411–446. PMLR, 2023.

[22] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*, 1993.

[23] N. Courty, R. Flamary, A. Habrard, and A. Rakotomamonjy. Joint distribution optimal transportation for domain adaptation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[24] N. Courty, R. Flamary, and D. Tuia. Domain adaptation with regularized optimal transport. In T. Calders, F. Esposito, E. Hüllermeier, and R. Meo, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I*, volume 8724 of *Lecture Notes in Computer Science*, pages 274–289. Springer, 2014.

[25] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(9):1853–1865, 2017.

[26] M. Cuturi, L. Meng-Papaxanthos, Y. Tian, C. Bunne, G. Davis, and O. Teboul. Optimal transport tools (ott): A jax toolbox for all things wasserstein. *arXiv preprint arXiv:2201.12324*, 2022.

[27] B. B. Damodaran, B. Kellenberger, R. Flamary, D. Tuia, and N. Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part IV*, volume 11208 of *Lecture Notes in Computer Science*, pages 467–483. Springer, 2018.

[28] N. Deb and B. Sen. Multivariate rank-based distribution-free nonparametric testing using measure transportation, 2019.

[29] E. del Barrio, A. González-Sanz, and J.-M. Loubes. Central limit theorems for general transportation costs. In *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques*, volume 60, pages 847–873. Institut Henri Poincaré, 2024.

[30] E. Del Barrio, A. González Sanz, and J.-M. Loubes. Central limit theorems for semi-discrete wasserstein distances. *Bernoulli*, 30(1):554–580, 2024.

[31] E. del Barrio, P. Gordaliza, H. Lescornel, and J. Loubes. Central limit theorem and bootstrap procedure for wasserstein's variations with an application to structural relationships between distributions. *J. Multivar. Anal.*, 169:341–362, 2019.

[32] I. Diakonikolas, M. Hardt, and L. Schmidt. Differentially private learning of structured discrete distributions. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 2566–2574, 2015.

[33] B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 3571–3580, 2017.

[34] Y. Ding, R. Li, and L. Xue. Statistical convergence rates of optimal transport map estimation between general distributions, 2024.

[35] I. Dinur and K. Nissim. Revealing information while preserving privacy. In F. Neven, C. Beeri, and T. Milo, editors, *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*, pages 202–210. ACM, 2003.

[36] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013, Allerton Park & Retreat Center, Monticello, IL, USA, October 2-4, 2013*, page 1592. IEEE, 2013.

[37] J. C. Duchi, M. J. Wainwright, and M. I. Jordan. Minimax optimal procedures for locally private estimation. *CoRR*, abs/1604.02390, 2016.

[38] C. Dwork, F. McSherry, K. Nissim, and A. D. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.

[39] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[40] Ú. Erlingsson, V. Pihur, and A. Korolova. RAPPOR: randomized aggregatable privacy-preserving ordinal response. In G. Ahn, M. Yung, and N. Li, editors, *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pages 1054–1067. ACM, 2014.

[41] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '03, page 211–222, New York, NY, USA, 2003. Association for Computing Machinery.

[42] J. Feydy, B. Charlier, F. Vialard, and G. Peyré. Optimal transport for diffeomorphic registration. In M. Descoteaux, L. Maier-Hein, A. M. Franz, P. Jannin, D. L. Collins, and S. Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I*, volume 10433 of *Lecture Notes in Computer Science*, pages 291–299. Springer, 2017.

[43] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

[44] R. Flamary, M. Cuturi, N. Courty, and A. Rakotomamonjy. Wasserstein discriminant analysis. *Mach. Learn.*, 107(12):1923–1945, 2018.

[45] A. Forrow, J. Hütter, M. Nitzan, P. Rigollet, G. Schiebinger, and J. Weed. Statistical optimal transport via factored couplings. In K. Chaudhuri and M. Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 2454–2465. PMLR, 2019.

[46] M. Fredrikson, S. Jha, and T. Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In I. Ray, N. Li, and C. Kruegel, editors, *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-16, 2015*, pages 1322–1333. ACM, 2015.

[47] A. Genevay, G. Peyré, and M. Cuturi. Learning generative models with sinkhorn divergences. In A. J. Storkey and F. Pérez-Cruz, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, volume 84 of *Proceedings of Machine Learning Research*, pages 1608–1617. PMLR, 2018.

[48] J. Gillenwater, M. Joseph, and A. Kulesza. Differentially private quantiles. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 3713–3722. PMLR, 2021.

[49] P. Gordaliza, E. Del Barrio, G. Fabrice, and J.-M. Loubes. Obtaining fairness using optimal transport theory. In *International conference on machine learning*, pages 2357–2365. PMLR, 2019.

[50] P. Gordaliza, E. del Barrio, F. Gamboa, and J. Loubes. Obtaining fairness using optimal transport theory. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 2357–2365. PMLR, 2019.

[51] E. Grave, A. Joulin, and Q. Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. In K. Chaudhuri and M. Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 1880–1890. PMLR, 2019.

[52] F. F. Gunsilius. On the convergence rate of potentials of brenier maps. *Econometric Theory*, 38(2):381–417, 2022.

[53] L. Györfi and M. Kroll. On rate optimal private regression under local differential privacy. *arXiv preprint arXiv:2206.00114*, 2022.

[54] L. Györfi and M. Kroll. Multivariate density estimation from privatised data: universal consistency and minimax rates. *Journal of Nonparametric Statistics*, 0(0):1–23, 2023.

[55] F. Harder, K. Adamczewski, and M. Park. Dp-merf: Differentially private mean embeddings with randomfeatures for practical privacy-preserving data generation. In *International conference on artificial intelligence and statistics*, pages 1819–1827. PMLR, 2021.

[56] J. Holm, T. Arildsen, M. Nielsen, and S. L. Nielsen. Orthonormal, moment preserving boundary wavelet scaling functions in python. *SN Applied Sciences*, 2(12):2032, 2020.

[57] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genet*, 4(8):e1000167, 2008.

[58] J.-C. Hütter and P. Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166 – 1194, 2021.

[59] H. Janati, M. Cuturi, and A. Gramfort. Wasserstein regularization for sparse multi-task regression. In K. Chaudhuri and M. Sugiyama, editors, *The 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019, 16-18 April 2019, Naha, Okinawa, Japan*, volume 89 of *Proceedings of Machine Learning Research*, pages 1407–1416. PMLR, 2019.

[60] G. Kamath, J. Li, V. Singhal, and J. R. Ullman. Privately learning high-dimensional distributions. In A. Beygelzimer and D. Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 1853–1902. PMLR, 2019.

[61] G. Kamath, X. Liu, and H. Zhang. Improved rates for differentially private stochastic convex optimization with heavy-tailed data. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, and S. Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 10633–10660. PMLR, 2022.

[62] G. Kamath, A. Mouzakis, M. Regehr, V. Singhal, T. Steinke, and J. R. Ullman. A bias-variance-privacy trilemma for statistical estimation. *CoRR*, abs/2301.13334, 2023.

[63] G. Kamath, A. Mouzakis, and V. Singhal. New lower bounds for private estimation and a generalized fingerprinting lemma. In *NeurIPS*, 2022.

[64] G. Kamath, V. Singhal, and J. R. Ullman. Private mean estimation of heavy-tailed distributions. In J. D. Abernethy and S. Agarwal, editors, *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125 of *Proceedings of Machine Learning Research*, pages 2204–2235. PMLR, 2020.

[65] H. Kaplan, S. Schnapp, and U. Stemmer. Differentially private approximate quantiles. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvári, G. Niu, and S. Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 10751–10761. PMLR, 2022.

[66] V. Karwa and S. P. Vadhan. Finite sample differentially private confidence intervals. In A. R. Karlin, editor, *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, volume 94 of *LIPIcs*, pages 44:1–44:9. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[67] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 531–540, 2008.

[68] Y. Kawamoto and T. Murakami. Local obfuscation mechanisms for hiding probability distributions. In K. Sako, S. A. Schneider, and P. Y. A. Ryan, editors, *Computer Security - ESORICS 2019 - 24th European Symposium on Research in Computer Security, Luxembourg, September 23-27, 2019, Proceedings, Part I*, volume 11735 of *Lecture Notes in Computer Science*, pages 128–148. Springer, 2019.

[69] M. Klatt, C. Tameling, and A. Munk. Empirical regularized optimal transport: Statistical theory and applications. *SIAM J. Math. Data Sci.*, 2(2):419–443, 2020.

[70] M. Kroll. On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics*, 15(1):1783 – 1813, 2021.

[71] A. Kroshnin, V. G. Spokoiny, and A. L. Suvorikova. Statistical inference for bures–wasserstein barycenters. *The Annals of Applied Probability*, 2019.

[72] C. Lalanne and S. Gadat. Privately Learning Smooth Distributions on the Hypercube by Projections. In *ICML 2024 - 41st International Conference on Machine Learning*, page 39 p., Vienna, Austria, July 2024.

[73] C. Lalanne, A. Garivier, and R. Gribonval. About the cost of central privacy in density estimation. *Transactions on Machine Learning Research*, 2023.

[74] C. Lalanne, A. Garivier, and R. Gribonval. Private Statistical Estimation of Many Quantiles. In *ICML 2023 - 40th International Conference on Machine Learning*, Honolulu, United States, July 2023.

[75] C. Lalanne, F. Iutzeler, J.-M. Loubes, and J. Chhor. On the Private Estimation of Smooth Transport Maps. In *ICML 2025 - 42nd International Conference on Machine Learning*, page 30 p., Vancouver (BC), Canada, July 2025.

[76] H. Lavenant, S. Claici, E. Chien, and J. Solomon. Dynamical optimal transport on discrete surfaces. *ACM Trans. Graph.*, 37(6):250, 2018.

[77] T. Le, K. Nguyen, S. Sun, K. Han, N. Ho, and X. Xie. Diffeomorphic mesh deformation via efficient optimal transport for cortical surface reconstruction. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024.

[78] G. Loukides, J. C. Denny, and B. A. Malin. The disclosure of diagnosis codes can breach research participants' privacy. *J. Am. Medical Informatics Assoc.*, 17(3):322–327, 2010.

[79] T. Manole, S. Balakrishnan, J. Niles-Weed, and L. A. Wasserman. Plugin estimation of smooth optimal transport maps. *The Annals of Statistics*, 2024.

[80] G. Montavon, K. Müller, and M. Cuturi. Wasserstein training of restricted boltzmann machines. In D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3711–3719, 2016.

[81] B. Muzellec, A. Vacher, F. Bach, F.-X. Vialard, and A. Rudi. Near-optimal estimation of smooth transport maps with kernel sums-of-squares, 2021.

[82] A. Narayanan and V. Shmatikov. How to break anonymity of the netflix prize dataset. *CoRR*, abs/cs/0610105, 2006.

[83] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*, pages 111–125. IEEE Computer Society, 2008.

[84] J. Niles-Weed and Q. Berthet. Minimax estimation of smooth densities in wasserstein distance. *The Annals of Statistics*, 2022.

[85] V. M. Panaretos and Y. Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6(Volume 6, 2019):405–431, 2019.

[86] M. Perrot, N. Courty, R. Flamary, and A. Habrard. Mapping estimation for discrete optimal transport. In D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 4197–4205, 2016.

[87] C. Pierquin, A. Bellet, M. Tommasi, and M. Boussard. Rényi pufferfish privacy: General additive noise mechanisms and privacy amplification by iteration via shift reduction lemmas. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024.

[88] A.-A. Pooladian and J. Niles-Weed. Entropic estimation of optimal transport maps. *arXiv preprint arXiv:2109.12004*, 2021.

[89] A. Rakotomamonjy and L. Ralaivola. Differentially private sliced wasserstein distance. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 8810–8820. PMLR, 2021.

[90] A. Ramdas, N. G. Trillos, and M. Cuturi. On wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2):47, 2017.

[91] P. Rigollet and J. Weed. Entropic optimal transport is maximum-likelihood deconvolution. *Comptes Rendus Mathematique*, 356(11):1228–1235, 2018.

[92] D. Rodríguez-Vítores, C. Lalanne, and J.-M. Loubes. Learning with Differentially Private (Sliced) Wasserstein Gradients. working paper or preprint, May 2025.

[93] L. Rout, A. Korotin, and E. Burnaev. Generative modeling with optimal transport maps. In *International Conference on Learning Representations*, 2022.

[94] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Birkhäuser Cham/Springer, 2016.

[95] S. Schluttenhofer and J. Johannes. Adaptive pointwise density estimation under local differential privacy, 2022.

[96] M. A. Schmitz, M. Heitz, N. Bonneel, F. M. N. Mboula, D. Coeurjolly, M. Cuturi, G. Peyré, and J. Starck. Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM J. Imaging Sci.*, 11(1):643–678, 2018.

[97] I. Sebag, M. S. Pydi, J. Franceschi, A. Rakotomamonjy, M. Gartrell, J. Atif, and A. Allauzen. Differentially private gradient flow based on the sliced wasserstein distance for non-parametric generative modeling. *CoRR*, abs/2312.08227, 2023.

[98] V. Seguy and M. Cuturi. Principal geodesic analysis for probability measures under the optimal transport metric. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 3312–3320, 2015.

[99] V. Seguy, B. B. Damodaran, R. Flamary, N. Courty, A. Rolet, and M. Blondel. Large scale optimal transport and mapping estimation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.

[100] V. Singhal. A polynomial time, pure differentially private estimator for binary product distributions. *CoRR*, abs/2304.06787, 2023.

[101] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. J. Guibas. Convolutional wasserstein distances: efficient optimal transportation on geometric domains. *ACM Trans. Graph.*, 34(4):66:1–66:11, 2015.

[102] J. Solomon, G. Peyré, V. G. Kim, and S. Sra. Entropic metric alignment for correspondence problems. *ACM Trans. Graph.*, 35(4):72:1–72:13, 2016.

[103] M. Staib, S. Claici, J. Solomon, and S. Jegelka. Parallel streaming wasserstein barycenters. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 2647–2658, 2017.

[104] L. Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000):1–34, 2000.

[105] L. Sweeney. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, 10(5):557–570, 2002.

[106] C. Tameling and A. Munk. Computational strategies for statistical inference based on empirical optimal transport. In *2018 IEEE Data Science Workshop, DSW 2018, Lausanne, Switzerland, June 4-6, 2018*, pages 175–179. IEEE, 2018.

[107] A. G. Thakurta, A. H. Vyrros, U. S. Vaishampayan, G. Kapoor, J. Freudiger, V. R. Sridhar, and D. Davidson. Learning new words. *Granted US Patents*, 9594741, 2017.

[108] N. L. Tien, A. Habrard, and M. Sebban. Differentially private optimal transport: Application to domain adaptation. In S. Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 2852–2858. ijcai.org, 2019.

[109] T. van Erven and P. Harremoës. Rényi divergence and kullback-leibler divergence. *IEEE Trans. Inf. Theory*, 60(7):3797–3820, 2014.

[110] C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[111] I. Wagner and D. Eckhoff. Technical privacy metrics: A systematic survey. *ACM Comput. Surv.*, 51(3):57:1–57:38, 2018.

[112] L. A. Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

[113] J. Weed and Q. Berthet. Estimation of smooth densities in wasserstein distance. In A. Beygelzimer and D. Hsu, editors, *Conference on Learning Theory, COLT 2019, 25-28 June 2019, Phoenix, AZ, USA*, volume 99 of *Proceedings of Machine Learning Research*, pages 3118–3119. PMLR, 2019.

[114] C. Yang, J. Qi, and A. Zhou. Wasserstein differential privacy. In M. J. Wooldridge, J. G. Dy, and S. Natarajan, editors, *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 16299–16307. AAAI Press, 2024.

[115] Y. Zemel and V. M. Panaretos. Fréchet means and Procrustes analysis in Wasserstein space. *Bernoulli*, 25(2):932 – 976, 2019.

A Complements on Section 2

Theorem A.1. *Under the same notations as in Section 2, if there exists $s > 0$ such that $f_Z \in \mathfrak{B}^s_{\infty,\infty}(\Omega)$ and $\gamma > 1$ such that $\gamma > f_Z > \gamma^{-1}$ on Ω, then the estimator $\hat{f}_{Z,priv}$ using*

$$2^J \asymp \min\left(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{\max(d;3/2)+s}}\right)$$

satisfies

$$\mathbb{E}\left[W_2^2(f_Z, \hat{f}_{Z,J,priv})\right] \lesssim \begin{cases} \max\left(n^{-\frac{2(s+1)}{2s+d}}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\right) & \text{for } d \geq 3 \\ \max\left(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+2}}\right) \times \textit{PolyLog}(n,\epsilon) & \text{for } d = 2 \\ \max\left(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+3/2}}\right) & \text{for } d = 1 \end{cases}$$

Proof. See Section B.3. □

We recall that $\pi_{n\epsilon} := n\epsilon$ when working under central DP and $\pi_{n\epsilon} := \sqrt{n}\epsilon$ when working under local DP.

B Proofs of Section 2

B.1 Proof of Proposition 2.1

Let $z, z' \in \Omega$, we can first notice that

$$\begin{aligned} \|q(z) - q(z')\|_1 &= \|(\xi(z); \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j) - (\xi(z'); \xi \in \Phi \cup \bigcup_{j=j_0}^{J} \Psi_j)\|_1 \\ &= \|(\xi(z); \xi \in \Phi) - (\xi(z'); \xi \in \Phi)\|_1 + \sum_{j=j_0}^{J} \|(\xi(z); \xi \in \Psi_j) - (\xi(z'); \xi \in \Psi_j)\|_1 \,. \end{aligned} \tag{5}$$

Then, for any $j_0 \leq j \leq J$, we can exploit the restricted support property of Assumption 1 which yields

$$\begin{aligned} \|(\xi(z); \xi \in \Psi_j) - (\xi(z'); \xi \in \Psi_j)\|_1 &\leq \| \sum_{\xi \in \Psi_j} \mathbb{1}(\cdot \in I_\xi)\|_\infty 2 \sup_{\xi \in \Psi_j} \|\xi\|_\infty \\ &\leq 2C_3 \sup_{\xi \in \Psi_j} \|\xi\|_\infty \,, \end{aligned} \tag{6}$$

and the scaling bound of Assumption 1 yields

$$\|(\xi(z); \xi \in \Psi_j) - (\xi(z'); \xi \in \Psi_j)\|_1 \leq 2C_3C_4 2^{jd/2} \,. \tag{7}$$

Similarly, $\|(\xi(z); \xi \in \Phi) - (\xi(z'); \xi \in \Phi)\|_1$ is controlled as

$$\begin{aligned} \|(\xi(z); \xi \in \Phi) - (\xi(z'); \xi \in \Phi)\|_1 &\leq 2|\Phi| \sup_{\xi \in \Phi} \|\xi\|_\infty \\ &\leq 2C_1C_2 \end{aligned} \tag{8}$$

Finally,

$$\begin{aligned}\|q(z)-q(z')\|_1 &\leq 2C_1C_2 + \sum_{j=j_0}^{J} 2C_3C_4 2^{jd/2} \\ &= 2C_1C_2 + 2C_3C_4 \sum_{j=j_0}^{J} 2^{jd/2} \\ &= 2C_1C_2 + 2C_3C_4 \frac{2^{(J+1)d/2} - 2^{j_0 d/2}}{2^{d/2}-1} \\ &\leq 2C_1C_2 + 6C_3C_4 2^{(J+1)d/2}\end{aligned} \tag{9}$$

where the last inequality comes from the fact that $(2^{d/2}-1)^{-1} \leq (\sqrt{2}-1)^{-1} \leq 3$.

B.2 Proof of Theorem 2.1

The claim about local DP is the simplest, as $\Delta = 2C_1C_2 + 6C_3C_4 2^{(J+1)d/2}$ is an upper bound on the l_1 sensitivity of the querying mechanism by Proposition 2.1. The claimed privacy follows from a direct application of the Laplace mechanism [38].

For the central DP case, we have to study the l_1 sensitivity of the query

$$(Z_1, \ldots, Z_n) \mapsto \frac{1}{n} \sum_{i=1}^{n} q(Z_i) \tag{10}$$

which is obtained by noticing that, without loss of generality (by symmetry of the expression),

$$\begin{aligned}\left| \frac{1}{n}\left(\sum_{i=1}^{n-1} q(Z_i) + q(Z_n) \right) - \frac{1}{n}\left(\sum_{i=1}^{n-1} q(Z_i) + q(Z'_n) \right) \right| &= \frac{1}{n} |q(Z_n) - q(Z'_n)| \\ &\leq \frac{\Delta}{n}\end{aligned} \tag{11}$$

where the last inequality follows from Proposition 2.1.

Thus, the l_1 sensitivity of the query of (10) is upper bounded by $\frac{\Delta}{n}$. Then, the claimed DP guarantees in the central DP case follows from a simple application of the Laplace mechanism [38, 39] to this deterministic query.

B.3 Proof of Theorem A.1

Let us define the event $A = \left\{ \tilde{f}_{Z,J,\text{priv}} \geq \gamma^{-1}/2 \text{ pointwise and } 1 - a \leq 1/\int_\Omega \tilde{f}_{Z,J,\text{priv}} \leq 1 + a \right\}$ where a is a fixed positive constant strictly smaller than 1.

Risk decomposition. We can write

$$\mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z) \right] = \mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z) \mathbb{1}_A \right] + \mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z) \mathbb{1}_{A^c} \right] . \tag{12}$$

Since Ω is compact, the second term is easily tractable as

$$\mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z) \mathbb{1}_{A^c} \right] \lesssim \mathbb{P}(A^c) \tag{13}$$

and $\mathbb{P}(A^c)$ is negligible compared to the rate we show in this proof (in local or central DP) by Lemma B.5.

The first term can be controlled by first using Theorem 4 from [84] which gives

$$\mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z) \mathbb{1}_A \right] \lesssim \mathbb{E}\left[\|\hat{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathcal{B}^{-1}_{2,1}(\Omega)} \mathbb{1}_A \right] , \tag{14}$$

which in turn can be controlled by using the event A as

$$\begin{aligned}
&\mathbb{E}\left[W_2^2(\hat{f}_{Z,J,\text{priv}}, f_Z)\mathbb{1}_A\right] \\
&\lesssim \mathbb{E}\left[\|\frac{1}{\int_\Omega \tilde{f}_{Z,J,\text{priv}}}\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\mathbb{1}_A\right] \\
&= \mathbb{E}\left[\left(\|\frac{1}{\int_\Omega \tilde{f}_{Z,J,\text{priv}}}\left(\tilde{f}_{Z,J,\text{priv}} - f_Z\right) - \left(1 - \frac{1}{\int_\Omega \tilde{f}_{Z,J,\text{priv}}}\right)f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right)\mathbb{1}_A\right] \\
&\lesssim \mathbb{E}\left[\left((1+a)^2\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right)\mathbb{1}_A\right] + \\
&\qquad\qquad \mathbb{E}\left[(1+a)^2\left(\left(\int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1\right)^2\|f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right)\mathbb{1}_A\right] \\
&\lesssim \mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] + \mathbb{E}\left[\left(\int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1\right)^2\right].
\end{aligned} \tag{15}$$

The term $\mathbb{E}\left[\left(\int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1\right)^2\right]$ is controlled using Lemma B.3, and by the rest of the analysis is always smaller than $\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right]$ up to a multiplicative constant.

For the term $\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right]$, we can use the following decomposition.

$$\begin{aligned}
&\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)} \\
&\lesssim \underbrace{\|\tilde{f}_{Z,J,\text{priv}} - \tilde{f}_{Z,J}\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}}_{\text{Privacy noise}} + \underbrace{\|\tilde{f}_{Z,J} - f_{Z,J}\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}}_{\text{Sampling noise}} + \underbrace{\|f_{Z,J} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}}_{\text{Bias}}
\end{aligned} \tag{16}$$

Control of the bias. By Lemma 33 in [79], since f_Z satisfies $\frac{1}{\gamma} \leq f_Z \leq \gamma$ on Ω and since $f_Z \in \mathfrak{B}^s_{\infty,\infty}(\Omega)$,

$$\|f_{Z,J} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)} \lesssim 2^{-2J(s+1)} \tag{17}$$

Control of the sampling noise. The sampling noise is controlled as in [79]. We rewrite the analysis adapted to the notations of the article below for completeness. For any $\eta \in \mathbb{R}$,

$$\begin{aligned}
&\|\tilde{f}_{Z,J} - f_{Z,J}\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)} \\
&= \|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Phi)\|_2^2 + \left(\sum_{j=j_0}^J 2^{-j}\|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Psi_j)\|_2\right)^2 \\
&\overset{\text{Cauchy-Schwarz}}{\leq} \|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Phi)\|_2^2 + \left(\sum_{j=j_0}^J 2^{2(\eta-1)j}\|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Psi_j)\|_2^2\right)\left(\sum_{j=j_0}^J 2^{-2\eta j}\right).
\end{aligned} \tag{18}$$

Then, we use again Lemma 33 in [79] which first guarantees that

$$\mathbb{E}\left[\|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Phi)\|_2^2\right] \lesssim \frac{1}{n} \tag{19}$$

and Proposition 4 from [84], or Lemma 33 from [79] guarantee that

$$\mathbb{E}\left[\|(\hat{\beta}_\xi - \beta_\xi; \xi \in \Psi_j)\|_2^2\right] \lesssim (2^{jd/2}/\sqrt{n})^2 \tag{20}$$

for any $j \geq j_0$.

Overall, this analysis yields that for any $\eta \in \mathbb{R}$,

$$\mathbb{E}\left[\|\tilde{f}_{Z,J} - f_{Z,J}\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] \leq \frac{1}{n}\left(1 + \left(\sum_{j=j_0}^{J} 2^{2(\eta+\frac{d}{2}-1)j}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right)\right). \tag{21}$$

Control of the privacy noise, central DP case. In the central DP case, a similar decomposition yields that for any $\eta \in \mathbb{R}$

$$\begin{aligned}
&\|\tilde{f}_{Z,\text{priv}} - \tilde{f}_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)} \\
&= \|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2 + \left(\sum_{j=j_0}^{J} 2^{-j}\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_j)\|_2\right)^2 \\
&\overset{\text{Cauchy-Schwarz}}{\leq} \|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2 + \left(\sum_{j=j_0}^{J} 2^{-2(\eta-1)j}\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_j)\|_2^2\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right).
\end{aligned} \tag{22}$$

We recall that in the central DP case, the $\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi$'s are i.i.d with distribution $\frac{\Delta}{n\epsilon}\mathcal{L}(1)$.

As a consequence, one has

$$\begin{aligned}
\mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2\right] &= \sum_{\xi\in\Phi} \mathbb{E}\left[(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2\right] \\
&\lesssim |\Phi|\frac{\Delta^2}{n^2\epsilon^2} \\
&\lesssim \frac{\Delta^2}{n^2\epsilon^2}
\end{aligned} \tag{23}$$

where the last line follows from Assumption 1, and for any $j_0, \leq j \leq J$,

$$\begin{aligned}
\mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_i)\|_2^2\right] &= \sum_{\xi\in\Psi_j} \mathbb{E}\left[(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2\right] \\
&\lesssim |\Psi_j|\frac{\Delta^2}{n^2\epsilon^2} \\
&\lesssim 2^{jd}\frac{\Delta^2}{n^2\epsilon^2}
\end{aligned} \tag{24}$$

where the last line again follows from Assumption 1.

Hence,

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{f}_{Z,\text{priv}} - \tilde{f}_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] \\
&\leq \mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2\right] + \left(\sum_{j=j_0}^{J} 2^{2(\eta-1)j}\mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_j)\|_2^2\right]\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right) \\
&\lesssim \frac{\Delta^2}{n^2\epsilon^2} + \left(\sum_{j=j_0}^{J} 2^{2(\eta-1)j}2^{jd}\frac{\Delta^2}{n^2\epsilon^2}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right) \\
&\lesssim \frac{2^{(J+1)d}}{n^2\epsilon^2} + \left(\sum_{j=j_0}^{J} 2^{2(\eta-1)j}2^{jd}\frac{2^{(J+1)d}}{n^2\epsilon^2}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right) \\
&= \frac{2^{(J+1)d}}{n^2\epsilon^2}\left(1 + \left(\sum_{j=j_0}^{J} 2^{2(\eta+\frac{d}{2}-1)j}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right)\right).
\end{aligned} \tag{25}$$

where the upper-bound on Δ follows from Proposition 2.1.

Control of the privacy noise, local DP case. Again, in the central DP case, the following decomposition still holds for any $\eta \in \mathbb{R}$

$$\begin{aligned}
&\|\tilde{f}_{Z,\text{priv}} - \tilde{f}_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)} \\
&\leq \|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2 + \left(\sum_{j=j_0}^{J} 2^{-2(\eta-1)j} \|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_j)\|_2^2\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right) .
\end{aligned} \tag{26}$$

We recall that in the local DP case, the $\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi$'s are i.i.d with distribution $\frac{1}{n}\sum_{i=1}^n \frac{\Delta}{\epsilon}\mathcal{L}(1)$.

As a consequence, one has

$$\begin{aligned}
\mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Phi)\|_2^2\right] &= \sum_{\xi \in \Phi} \mathbb{E}\left[(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2\right] \\
&\lesssim |\Phi| \frac{\Delta^2}{n\epsilon^2} \\
&\lesssim \frac{\Delta^2}{n\epsilon^2}
\end{aligned} \tag{27}$$

where the last line follows from Assumption 1, and for any $j_0, \leq j \leq J$,

$$\begin{aligned}
\mathbb{E}\left[\|(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi; \xi \in \Psi_i)\|_2^2\right] &= \sum_{\xi \in \Psi_j} \mathbb{E}\left[(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2\right] \\
&\lesssim |\Psi_j| \frac{\Delta^2}{n\epsilon^2} \\
&\lesssim 2^{jd} \frac{\Delta^2}{n\epsilon^2}
\end{aligned} \tag{28}$$

where the last line again follows from Assumption 1.

Hence,

$$\mathbb{E}\left[\|\tilde{f}_{Z,\text{priv}} - \tilde{f}_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] \lesssim \frac{2^{(J+1)d}}{n\epsilon^2}\left(1 + \left(\sum_{j=j_0}^{J} 2^{2(\eta+\frac{d}{2}-1)j}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right)\right) . \tag{29}$$

where the upper-bound on Δ follows from Proposition 2.1.

Conclusion of the proof. Putting the pieces together, we get that (both with central and local DP)

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] \\
&\lesssim \left(\frac{1}{n} + \frac{2^{Jd}}{\pi^2_{n,\epsilon}}\right)\left(1 + \left(\sum_{j=j_0}^{J} 2^{2(\eta+\frac{d}{2}-1)j}\right)\left(\sum_{j=j_0}^{J} 2^{-2\eta j}\right)\right) + 2^{-2J(s+1)} .
\end{aligned} \tag{30}$$

When $d \geq 3$, choosing $1 - \frac{d}{2} < \eta < 0$ yields

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] &\lesssim \left(\frac{1}{n} + \frac{2^{Jd}}{\pi^2_{n,\epsilon}}\right)2^{(d-2)J} + 2^{-2J(s+1)} \\
&\lesssim \max\left(n^{-\frac{2(s+1)}{2s+d}}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\right) ,
\end{aligned} \tag{31}$$

since $2^J \asymp \min\left(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{d+s}}\right)$.

When $d = 2$, we can choose $\eta = 0$ which yields

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] &\lesssim \left(\frac{1}{n} + \frac{2^{Jd}}{\pi^2_{n,\epsilon}}\right)J^2 + 2^{-2J(s+1)} \\
&\lesssim \text{polylog}(n, \epsilon) \max\left(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\right) .
\end{aligned} \tag{32}$$

When $d = 1$, we take $0 < \eta < 1/2$. In this case,

$$\begin{aligned}\mathbb{E}\left[\|\tilde{f}_{Z,J,\text{priv}} - f_Z\|^2_{\mathfrak{B}^{-1}_{2,1}(\Omega)}\right] &\lesssim \frac{1}{n} + \frac{2^{Jd}}{\pi^2_{n,\epsilon}} + 2^{-2J(s+1)} \\ &\lesssim \max\left(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+3/2}}\right) .\end{aligned} \tag{33}$$

since $2^J \asymp \min\left(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{2}{2s+3}}\right)$.

B.4 Auxiliary results for Section 2

Lemma B.1. *Let $L_1, \ldots, L_M$ be M independent random variables with Laplace distribution (that is, with density $t \mapsto \frac{1}{2}e^{-|t|}$ w.r.t. Lebesgue's measure). Then*

$$\mathbb{P}\left(\max_{i=1,\ldots,M} |L_i| > \delta\right) = 1 - (1 - e^{-\delta})^M \leq \min(1, Me^{-\delta}) . \tag{34}$$

Proof. Let $\delta \geq 0$, by independence,

$$\begin{aligned}\mathbb{P}\left(\max_{i=1,\ldots,M} |L_i| \leq \delta\right) &= \mathbb{P}\left(\bigcap_{i=1,\ldots,M} (|L_i| \leq \delta)\right) \\ &= \prod_{i=1}^{M} \mathbb{P}\left((|L_i| \leq \delta)\right) \\ &= (1 - e^{-\delta})^M .\end{aligned} \tag{35}$$

Thus,

$$\begin{aligned}\mathbb{P}\left(\max_{i=1,\ldots,M} |L_i| > \delta\right) &= 1 - (1 - e^{-\delta})^M \\ &\leq 1 - e^{Me^{-\delta}} \\ &\leq \min(1, Me^{-\delta}) .\end{aligned} \tag{36}$$

Between the first and second line, we used that $1 - x \leq e^{-x}$ and between the second and last line, we used that $1 - e^{-x} \leq x$. □

Lemma B.2. *Let $v = (v_1, \ldots, v_n)$ be a vector of independent, centered Laplace random variables. Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ be a deterministic vector and define*

$$S := \langle a, v\rangle = \sum_{i=1}^{n} a_i v_i. \tag{37}$$

Then, for all $t \geq 0$,

$$\mathbb{P}(|S| \geq t) \leq 2\exp(-\frac{t^2}{4\|a\|_2^2 + 2\|a\|_\infty t}) . \tag{38}$$

Consequently,

$$\mathbb{P}(|S| \geq t) \leq 2\exp\left(-\min\left\{\frac{t^2}{8\|a\|_2^2}, \frac{t}{4\|a\|_\infty}\right\}\right) . \tag{39}$$

Proof. Since the v_i's are independent and symmetric, it suffices to control $\mathbb{P}(S \geq t)$.

For $|\lambda| < 1$, the Laplace distribution satisfies

$$\mathbb{E}[e^{\lambda v_i}] = \int_{\mathbb{R}} e^{\lambda x}\frac{1}{2}e^{-|x|}dx = \frac{1}{1-\lambda^2} \tag{40}$$

Therefore, for any λ such that $|\lambda a_i| < 1$ for all i,

$$\mathbb{E}[e^{\lambda S}] = \prod_{i=1}^{n} \mathbb{E}\left[e^{\lambda a_i v_i}\right] = \prod_{i=1}^{n} \frac{1}{1-\lambda^2 a_i^2}. \tag{41}$$

Using the inequality $-\log(1-u) \leq \frac{u}{1-u}$ for $u \in [0,1)$, we obtain

$$\log \mathbb{E}[e^{\lambda S}] \leq \sum_{i=1}^{n} \frac{\lambda^2 a_i^2}{1-\lambda^2 a_i^2} \leq \frac{\lambda^2 \|a\|_2^2}{1-\lambda \|a\|_\infty}, \tag{42}$$

for any $0 \leq \lambda < \|a\|_\infty^{-1}$.

For such λ, Markov's inequality yields

$$\mathbb{P}(S \geq t) \leq \exp\left(-\lambda t + \frac{\lambda^2 \|a\|_2^2}{1-\lambda \|a\|_\infty}\right). \tag{43}$$

Choose

$$\lambda^\star = \frac{t}{2\|a\|_2^2 + \|a\|_\infty t} < \|a\|_\infty^{-1}. \tag{44}$$

A direct computation yields

$$-\lambda^\star t + \frac{\lambda^{\star 2} \|a\|_2^2}{1-\lambda^\star \|a\|_\infty} = -\frac{t^2}{4\|a\|_2^2 + 2\|a\|_\infty t}. \tag{45}$$

Hence,

$$\mathbb{P}(S \geq t) \leq \exp\left(-\frac{t^2}{4\|a\|_2^2 + 2\|a\|_\infty t}\right). \tag{46}$$

Since S is symmetric,

$$\mathbb{P}(|S| \geq t) \leq 2\mathbb{P}(S \geq t), \tag{47}$$

which gives the first inequality.

If $t \leq 2\|a\|_2^2/\|a\|_\infty$, then

$$4\|a\|_2^2 + 2\|a\|_\infty t \leq 8\|a\|_2^2. \tag{48}$$

If $t \geq 2\|a\|_2^2/\|a\|_\infty$, then

$$4\|a\|_2^2 + 2\|a\|_\infty t \leq 4\|a\|_\infty t. \tag{49}$$

This yields the final result. □

Lemma B.3. *If Φ spanss constant functions and if Ψ is an orthonormal family of $L^2(\Omega)$, then for any $\delta \geq 0$, the estimator built with Algorithm 1 satisfies*

$$\mathbb{P}\left(\left|\int_\Omega \tilde{f}_{Z,J,priv} - 1\right| \geq \delta\right) \lesssim_\Psi e^{-C\frac{n\epsilon}{\Delta}\delta}, \tag{50}$$

and

$$\mathbb{E}\left[\left|\int_\Omega \tilde{f}_{Z,J,priv} - 1\right|^2\right] \lesssim_\Psi \frac{\Delta^2}{(n\epsilon)^2}, \tag{51}$$

and the estimator built with Algorithm 2 satisfies

$$\mathbb{P}\left(\left|\int_\Omega \tilde{f}_{Z,J,priv} - 1\right| \geq \delta\right) \lesssim_\Psi e^{-C\min\left(\frac{n\epsilon^2\delta^2}{\Delta^2}, \frac{n\epsilon\delta}{\Delta}\right)}, \tag{52}$$

and

$$\mathbb{E}\left[\left|\int_\Omega \tilde{f}_{Z,J,priv} - 1\right|^2\right] \lesssim_\Psi \frac{\Delta^2}{n\epsilon^2}, \tag{53}$$

for a positive constant C depending only on Ψ. In particular, this result applies if Ψ is the boundary corrected Daubechies Wavelet system thanks to Lemma 26 of [79].

Proof. In order to keep the notations consistent with [79], let us note

$$\beta'_\xi := \int_\Omega \xi \ . \tag{54}$$

Since Φ spans the function constant to 1, one obtains by orthonormality that

$$1 = \sum_{\xi \in \Phi} \beta'_\xi \xi \ . \tag{55}$$

Then, exploiting orthonormality again yields

$$\begin{aligned}
\int_\Omega \tilde{f}_{Z,J,\text{priv}} &= \int_\Omega 1 \times \left(\sum_{\xi \in \Phi} \hat{\beta}_{\xi,\text{priv}} \xi + \sum_{j=j_0}^{J} \sum_{\xi \in \Psi_j} \hat{\beta}_{\xi,\text{priv}} \xi \right) \\
&= \int_\Omega \left(\sum_{\xi \in \Phi} \beta'_\xi \xi \right) \times \left(\sum_{\xi \in \Phi} \hat{\beta}_{\xi,\text{priv}} \xi + \sum_{j=j_0}^{J} \sum_{\xi \in \Psi_j} \hat{\beta}_{\xi,\text{priv}} \xi \right) \\
&= \sum_{\xi \in \Phi} \beta'_\xi \hat{\beta}_{\xi,\text{priv}} \\
&= \sum_{\xi \in \Phi} \beta'_\xi \Big(\hat{\beta}_\xi + (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \Big) \\
&= \sum_{\xi \in \Phi} \beta'_\xi \hat{\beta}_\xi + \sum_{\xi \in \Phi} \beta'_\xi (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \\
&= \int_\Omega \left(\sum_{\xi \in \Phi} \beta'_\xi \xi \right) dZ_n + \sum_{\xi \in \Phi} \beta'_\xi (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \\
&= 1 + \sum_{\xi \in \Phi} \beta'_\xi (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)
\end{aligned} \tag{56}$$

In other words, the normalization factor is equal to 1 plus a residual term that depends only on the privacy noise.

In the following, we treat the β'_ξ's as non-zero without loss of generality, as the possible null values can be excluded without changing the result.

Central DP case. In the central DP case, $(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \sim \frac{\Delta}{n\epsilon} \mathcal{L}(1)$ and are independent when varying the ξ.

Thus, by Lemma B.1,

$$\mathbb{P}\left(\left| \int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1 \right| \geq \delta \right) \lesssim_\Psi e^{-C \frac{n\epsilon}{\Delta} \delta} \ , \tag{57}$$

where C is a constant depending only on Φ. Furthermore,

$$\begin{aligned}
\mathbb{E}\left[\left| \int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1 \right|^2 \right] &= \mathbb{E}\left[\left(\sum_{\xi \in \Phi} \beta'_\xi (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \right)^2 \right] \\
&= \mathbb{E}\left[\sum_{\xi \in \Phi} \beta'^2_\xi (\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2 \right] \\
&= \mathbb{E}\left[\sum_{\xi \in \Phi} \beta'^2_\xi (\frac{\Delta}{n\epsilon} \mathcal{L}(1))^2 \right] \\
&\lesssim_\Psi |\Phi| \frac{\Delta^2}{(n\epsilon)^2} \\
&\lesssim_\Psi \frac{\Delta^2}{(n\epsilon)^2} \ .
\end{aligned} \tag{58}$$

Local DP case. In the central DP case, $(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi) \sim \frac{1}{n}\sum_{i=1}^n \frac{\Delta}{\epsilon}\mathcal{L}(1)$ and are independent when varying the ξ.

Because of Lemma B.2, we have

$$\mathbb{P}\left(\left|\int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1\right| \geq \delta\right) \lesssim_\Psi e^{-C\min\left(\frac{n\epsilon^2\delta^2}{\Delta^2}, \frac{n\epsilon\delta}{\Delta}\right)}, \tag{59}$$

where C is again a constant depending only on Φ. Furthermore,

$$\begin{aligned}
\mathbb{E}\left[\left|\int_\Omega \tilde{f}_{Z,J,\text{priv}} - 1\right|^2\right] &= \mathbb{E}\left[\left(\sum_{\xi\in\Phi}\beta'_\xi(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{\xi\in\Phi}\beta'^2_\xi(\hat{\beta}_{\xi,\text{priv}} - \hat{\beta}_\xi)^2\right] \\
&= \mathbb{E}\left[\sum_{\xi\in\Phi}\beta'^2_\xi(\frac{1}{n}\sum_{i=1}^n \frac{\Delta}{\epsilon}\mathcal{L}(1))^2\right] \\
&\lesssim_\Psi |\Phi|\frac{\Delta^2}{n\epsilon^2} \\
&\lesssim_\Psi \frac{\Delta^2}{n\epsilon^2}.
\end{aligned} \tag{60}$$

□

Lemma B.4. *If $N \geq 2$, $f_Z \in \mathfrak{B}^s_{\infty,\infty}$ for a $s > 0$, then there exist two constants a, b depending only on Ψ such that for any $J \geq j_0$ and any $\delta > 0$,*

- *The result of Algorithm 1 satisfies*

$$\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,priv} - \beta_\xi\right| \geq \delta\right) \lesssim_\Psi e^{-\frac{n\delta^2}{4a^2}} + e^{-\frac{n\epsilon}{2\Delta}\delta}, \tag{61}$$

and for any $j_0 \leq j \leq J$,

$$\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,priv} - \beta_\xi\right| \geq \delta\right) \lesssim_\Psi 2^{jd}\left(e^{-\frac{n\delta^2/4}{b+2^{jd/2}b\delta/2}} + e^{-\frac{n\epsilon}{2\Delta}\delta}\right). \tag{62}$$

Furthermore, if $2^J = c_0\min\left(n^{\frac{1}{d+2s}}, (n\epsilon)^{\frac{1}{d+s}}\right)$ when $d \geq 2$ and $2^J = c_0\min\left(n^{\frac{1}{1+2s}}, (n\epsilon)^{\frac{1}{s+3/2}}\right)$ when $d = 1$

$$\mathbb{P}\left(\|\hat{f}_{Z,J,priv} - f_{Z,J}\|_\infty \geq \delta\right) \leq CJ2^{Jd(d+3)}\exp\left(-C\min\left(\frac{n\delta^2}{J^2 2^{Jd}}, \frac{n\epsilon}{J2^{Jd}}\delta\right)\right). \tag{63}$$

- *The result of Algorithm 2 satisfies*

$$\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,priv} - \beta_\xi\right| \geq \delta\right) \lesssim_\Psi e^{-\frac{n\delta^2}{4a^2}} + e^{-\frac{(\delta/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(\delta/2)}}, \tag{64}$$

and for any $j_0 \leq j \leq J$,

$$\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,priv} - \beta_\xi\right| \geq \delta\right) \lesssim_\Psi 2^{jd}\left(e^{-\frac{n\delta^2/4}{b+2^{jd/2}b\delta/2}} + e^{-\frac{(\delta/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(\delta/2)}}\right). \tag{65}$$

Furthermore, if $2^J = c_0\min\left(n^{\frac{1}{d+2s}}, (\sqrt{n}\epsilon)^{\frac{1}{d+s}}\right)$ when $d \geq 2$ and $2^J = c_0\min\left(n^{\frac{1}{1+2s}}, (\sqrt{n}\epsilon)^{\frac{1}{s+3/2}}\right)$ when $d = 1$

$$\mathbb{P}\left(\|\hat{f}_{Z,J,priv} - f_{Z,J}\|_\infty \geq \delta\right) \leq CJ2^{Jd(d+3)}\exp\left(-C\min\left(\frac{n\delta^2}{J^2 2^{Jd}}, \frac{n\epsilon^2}{J^2 2^{2Jd}}\delta^2, \frac{n\epsilon}{J2^{Jd}}\delta\right)\right). \tag{66}$$

Proof. Let us start with the central DP (Algorithm 1) case. Let $\delta > 0$.

$$\begin{aligned}\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\delta\right) &\leq \mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\hat{\beta}_\xi\right|\geq\delta/2\right)+\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_\xi-\beta_\xi\right|\geq\delta/2\right)\ , \\ &\lesssim_\Psi e^{-\frac{n\epsilon}{2\Delta}\delta}+e^{-\frac{n\delta^2}{4a^2}}\ ,\end{aligned} \tag{67}$$

where the second term in the sum is controlled using Lemma 30 in [79], and where the first term is controlled with a union-bound and Lemma B.1.

Similarly,

$$\begin{aligned}\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\delta\right) &\leq \mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\hat{\beta}_\xi\right|\geq\delta/2\right)+\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_\xi-\beta_\xi\right|\geq\delta/2\right) \\ &\lesssim_\Psi 2^{jd}\left(e^{-\frac{n\epsilon}{2\Delta}\delta}+e^{-\frac{n\delta^2/4}{b+2^{jd/2}b\delta/2}}\right),\end{aligned} \tag{68}$$

where the second term in the sum is again controlled using Lemma 30 in [79], and where the first term is again controlled with a union-bound and Lemma B.1.

For the concentration in infinite norm, following Lipschitz bounds and the covering from the proof of Lemma 30 in [79], we can say that for any $2^{-J}\leq\delta\leq 1$,

$$\begin{aligned}&\mathbb{P}\left(\|\hat{f}_{Z,J,\text{priv}}-f_{Z,J}\|_\infty\geq\delta\right) \\ &\leq K\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right)+K\mathbb{P}\left(J2^{Jd/2}\sup_{j_0\leq j\leq J}\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right)\end{aligned} \tag{69}$$

where $K=O(2^{Jd(d+2)})$ and where c is a positive constant.

Combining with the previous results, it follows that

$$\begin{aligned}&\mathbb{P}\left(\|\hat{f}_{Z,J,\text{priv}}-f_{Z,J}\|_\infty\geq\delta\right) \\ &\leq K\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right)+K\mathbb{P}\left(J2^{Jd/2}\sup_{j_0\leq j\leq J}\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right) \\ &\lesssim K\left(e^{-\frac{n(c\delta)^2}{4a^2}}+e^{-\frac{n\epsilon}{2\Delta}c\delta}\right)+K\sum_{j=j_0}^{J}\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\frac{c\delta}{J2^{Jd/2}}\right) \\ &\lesssim K\left(e^{-c_1n\delta^2}+e^{-c_2\frac{n\epsilon}{\Delta}\delta}\right) \\ &\qquad+K\sum_{j=j_0}^{J}2^{jd}\left(\exp\left(-\frac{n\left(\frac{c\delta}{J2^{Jd/2}}\right)^2/4}{b+2^{jd/2}b\left(\frac{c\delta}{J2^{Jd/2}}\right)/2}\right)+\exp\left(-\frac{n\epsilon}{2\Delta}\,\frac{c\delta}{J2^{Jd/2}}\right)\right) \\ &\lesssim K\left(e^{-c_1n\delta^2}+e^{-c_2\frac{n\epsilon}{\Delta}\delta}\right) \\ &\qquad+K\,J2^{Jd}\exp\left(-c_3\frac{n\delta^2}{J^22^{Jd}}\right)+KJ2^{Jd}\exp\left(-c_4\frac{n\epsilon}{\Delta}\,\frac{\delta}{J2^{Jd/2}}\right),\end{aligned} \tag{70}$$

for some positive constants c_1, c_2, c_3, c_4.

In particular, since $\Delta\asymp_\Psi 2^{Jd/2}$, the last term becomes

$$K\,J2^{Jd}\exp\Big\{-c_5\frac{n\epsilon}{J2^{Jd}}\delta\Big\}\ , \tag{71}$$

for a positive constant c_5 and thus

$$\mathbb{P}\left(\|\hat{f}_{Z,J,\text{priv}}-f_{Z,J}\|_\infty\geq\delta\right)\lesssim KJ2^{Jd}\exp\left(-c\min\left(\frac{n\delta^2}{J^22^{Jd}},\frac{n\epsilon}{J2^{Jd}}\delta\right)\right). \tag{72}$$

The claimed result follows.

Now for the local DP (Algorithm 2) case, let $\delta > 0$.

$$\begin{aligned}\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\delta\right) &\leq \mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\hat{\beta}_\xi\right|\geq\delta/2\right)+\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_\xi-\beta_\xi\right|\geq\delta/2\right)\,, \\ &\lesssim e^{-\frac{(\delta/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(\delta/2)}}+e^{-\frac{n\delta^2}{4a^2}}\,,\end{aligned} \tag{73}$$

where the second term in the sum is controlled using Lemma 30 in [79], and where the first term is controlled with a union-bound and Lemma B.2.

Similarly,

$$\begin{aligned}\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\delta\right) &\leq \mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\hat{\beta}_\xi\right|\geq\delta/2\right)+\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_\xi-\beta_\xi\right|\geq\delta/2\right) \\ &\lesssim 2^{jd}\left(e^{-\frac{(\delta/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(\delta/2)}}+e^{-\frac{n\delta^2/4}{b+2^{jd/2}b\delta/2}}\right),\end{aligned} \tag{74}$$

where the second term in the sum is again controlled using Lemma 30 in [79], and where the first term is again controlled with a union-bound and Lemma B.2.

For the concentration in infinite norm, the same arguments as in the previous case give the existence of $K = O(2^{Jd(d+2)})$ and of a positive constant c such that

$$\begin{aligned}&\mathbb{P}\left(\|\hat{f}_{Z,J,\text{priv}}-f_{Z,J}\|_\infty\geq\delta\right) \\ &\leq K\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right)+K\mathbb{P}\left(J2^{Jd/2}\sup_{j_0\leq j\leq J}\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right) \\ &\lesssim K\mathbb{P}\left(\sup_{\xi\in\Phi}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq c\delta\right)+K\sum_{j=j_0}^{J}\mathbb{P}\left(\sup_{\xi\in\Psi_j}\left|\hat{\beta}_{\xi,\text{priv}}-\beta_\xi\right|\geq\frac{c\delta}{J2^{Jd/2}}\right) \\ &\lesssim K\left(e^{-\frac{nc^2\delta^2}{4a^2}}+e^{-\frac{(c\delta/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(c\delta/2)}}\right) \\ &\qquad+K\sum_{j=j_0}^{J}\left(2^{jd}\left(e^{-\frac{n\frac{c\delta}{J2^{Jd/2}}^2/4}{b+2^{jd/2}b\frac{c\delta}{J2^{Jd/2}}/2}}+e^{-\frac{(\frac{c\delta}{J2^{Jd/2}}/2)^2}{4\frac{\Delta^2}{n\epsilon^2}+2\frac{\Delta}{n\epsilon}(\frac{c\delta}{J2^{Jd/2}}/2)}}\right)\right),\end{aligned} \tag{75}$$

In particular, since $\Delta \asymp 2^{Jd/2}$, the last term becomes

$$KJ2^{Jd}\exp\left(-c\min\left(\frac{n\epsilon^2}{J^2 2^{2Jd}}\delta^2,\frac{n\epsilon}{J2^{Jd}}\delta\right)\right), \tag{76}$$

for a positive constant c and thus

$$\mathbb{P}\left(\|\hat{f}_{Z,J,\text{priv}}-f_{Z,J}\|_\infty\geq\delta\right)\lesssim KJ2^{Jd}\exp\left(-c\min\left(\frac{n\delta^2}{J^2 2^{Jd}},\frac{n\epsilon^2}{J^2 2^{2Jd}}\delta^2,\frac{n\epsilon}{J2^{Jd}}\delta\right)\right) \tag{77}$$

for a possibly different positive constant c. The claimed result follows. □

Lemma B.5 (Convergence in L^∞ and normalization)**.** *If $N \geq 2$, $f_Z \in \mathfrak{B}^s_{\infty,\infty}$ for a $s > 0$, $\gamma \geq f_Z \geq \gamma^{-1}$ on Ω, if $2^J \asymp \min\left(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{d+s}}\right)$ when $d \geq 2$ and $2^J \asymp \min\left(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{s+3/2}}\right)$ when $d = 1$, the event $A = \left\{\tilde{f}_{Z,J,priv} \geq \gamma^{-1}/2 \text{ pointwise and } 1-a \leq 1/\int_\Omega \tilde{f}_{Z,J,priv} \leq 1+a\right\}$ where a is a fixed positive constant strictly smaller than 1 satisfies*

$$\mathbb{P}(A^c) \lesssim \frac{1}{n^2}+\frac{1}{\pi^4_{n,\epsilon}} \tag{78}$$

Proof. In this proof, we bound the probability of A by the probability of $\{ \|f_Z - f_{Z,J}\|_{L^\infty} \leq \gamma^{-1}/4$ and $\|\tilde{f}_{Z,J} - f_{Z,J}\|_{L^\infty} \leq \gamma^{-1}/4$ and $1 - a \leq 1/\int_\Omega \tilde{f}_{Z,J,\text{priv}} \leq 1 + a \}$.

We first bound the shift in infinite norm between f_Z and $f_{Z,J}$ as in the proof of Lemma 31 in [79] by

$$\begin{aligned}\|f_Z - f_{Z,J}\|_{L^\infty} &\leq \sum_{j \geq J+1} 2^{jd/2} \|(\beta_\xi ; \xi \in \Psi_j)\|_\infty \\ &\leq \|f_Z\|_{\mathfrak{B}^s_{\infty\infty}(\Omega)} \sum_{j \geq J+1} 2^{jd/2 - j(d/2+s)} \\ &\lesssim \|f_Z\|_{\mathfrak{B}^s_{\infty\infty}(\Omega)} 2^{-Js} \, .\end{aligned} \tag{79}$$

In particular, as soon as J (or equivalently, $\min(n, \pi_{n,\epsilon})$) is large enough (larger than a constant), this term is smaller than $\gamma^{-1}/4$.

The second part involving $\tilde{f}_{Z,J,\text{priv}}$ is controlled respectively by Lemma B.4 (for the infinite norm part) and Lemma B.3 (for the integral part).Because $\Delta \asymp 2^{Jd/2}$, we get that, when $\min(n, \pi_{n,\epsilon})$ is bigger than a constant, A^c is of probability at most

$$Ce^{-C\frac{n\epsilon}{2^{Jd/2}}} + CJ2^{Jd(d+3)} \exp\Big(-C \min\Big(\frac{n}{J^2 2^{Jd}}, \frac{n\epsilon}{J2^{Jd}}\Big)\Big) \tag{80}$$

under central DP for a positive constant C, and

$$Ce^{-C\min\left(\frac{n\epsilon^2}{\Delta^2}, \frac{n\epsilon}{\Delta}\right)} + CJ2^{Jd} \exp\left(-c \min\left(\frac{n}{J^2 2^{Jd(d+3)}}, \frac{n\epsilon^2}{J^2 2^{2Jd(d+3)}}, \frac{n\epsilon}{J2^{Jd}}\right)\right) \tag{81}$$

with again $\Delta \asymp 2^{Jd/2}$.

under local DP where C is again a positive constant.

Since we fix $2^J \asymp \min\Big(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{d+s}}\Big)$ when $d \geq 2$ and $2^J \asymp \min\Big(n^{\frac{1}{d+2s}}, (\pi_{n,\epsilon})^{\frac{1}{s+3/2}}\Big)$ when $d = 1$, a gross majoration yields, by super-polynomial decay,

$$\mathbb{P}(A^C) \lesssim \frac{1}{n^2} + \frac{1}{n^4\epsilon^4} \, . \tag{82}$$

under central DP, and

$$\mathbb{P}(A^C) \lesssim \frac{1}{n^2} + \frac{1}{n^2\epsilon^4} \, . \tag{83}$$

under local DP. □

C Proofs of Section 3

C.1 Proof of Theorem 3.1

This proof leverages the stability bounds of [8].

Before going into the details, let us explain how the hypotheses of these stability bounds are verified. Theorem 3 of [8] only requires the regularity and the curvature bounds $L_1 I_d \preceq \nabla^2 \phi \preceq L_2 I_d$ of the Brenier potential ϕ of the population pair (P_X, P_Y). These are guaranteed by Assumption 2 alone, and privatization does not affect them. The positive-part truncation and the normalization of the private density estimators are handled by a good event A, defined below, on which these two operations degrade the error by at most a multiplicative constant, and outside of which the error is negligible.

As in the proof of Theorem A.1 (in Section B.3), let us define the event $A = \{ \tilde{f}_{Z,J,\text{priv}} \geq \gamma^{-1}/2$ pointwise and $1 - a \leq 1/\int_\Omega \tilde{f}_{Z,J,\text{priv}} \leq 1 + a \}$ with $Z = X$, where a is a fixed positive constant strictly smaller than 1.This event consists of a pointwise lower bound on the privatized estimator together with a control of the normalization factor, both obtained from the two-sided bound $\gamma^{-1} < f_X < \gamma$ of Assumption 2 (see Lemmas B.4 and B.5) and from the concentration of $\int_\Omega \tilde{f}_{X,J,\text{priv}}$ around 1 (see Lemma B.3). By Lemma B.5, $\mathbb{P}(A^c) \lesssim \frac{1}{n^2} + \frac{1}{\pi^4_{n,\epsilon}}$.

On A, the positive-part truncation is inactive, since $\tilde{f}_{X,J,\text{priv}} \geq \gamma^{-1}/2 > 0$ pointwise, and the normalization multiplies $\tilde{f}_{X,J,\text{priv}}$ by a factor in $[1-a, 1+a]$, so that $\hat{f}_{X,J,\text{priv}} \geq (1-a)\gamma^{-1}/2$ pointwise. Also, under A, $\frac{df_X}{d\hat{f}_{X,J,\text{priv}}}$ (used to refer the Radon-Nikodym derivative of the distribution of probability P_X on Ω of density f_X by the one of density $\hat{f}_{X,J,\text{priv}}$) exists and is upper-bounded almost surely by $2\gamma^2/(1-a)$, since $f_X \leq \gamma$. This is what allows the change of reference measure from P_X to $\hat{f}_{X,J,\text{priv}}$ in the first inequality of Equation (84) below.

Under Assumption 2, the transport map T between P_X and P_Y is unique (see e.g. [94, Th. 1.17]). Let $\hat{T}_{\text{priv}}(x)$ be an optimal coupling of $\hat{f}_{X,J,\text{priv}}$ and $\hat{f}_{Y,J,\text{priv}}$. We have

$$\begin{aligned}
&\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(P_X)}\right] \\
&= \mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(P_X)}\mathbb{1}_A\right] + \mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(P_X)}\mathbb{1}_{A^c}\right] \\
&\lesssim \mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\mathbb{1}_A\right] + \mathbb{P}\left(\mathbb{1}_{A^c}\right) \\
&\leq \mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\right] + \mathbb{P}\left(\mathbb{1}_{A^c}\right).
\end{aligned} \tag{84}$$

In the first inequality, we also used that off A, since Ω is compact and both $\hat{T}_{\text{priv}}$ and T are Ω-valued, the error $\|\hat{T}_{\text{priv}} - T\|^2_{L^2(P_X)}$ is bounded by the constant $\text{diam}(\Omega)^2 = d$. As previously, $\mathbb{P}\left(\mathbb{1}_{A^c}\right)$ is negligible compared to the expected rate, in both privacy models, since $\frac{1}{n^2} + \frac{1}{\pi^4_{n,\epsilon}}$ is negligible against the rates of Theorem 3.1. Let us now study $\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,\text{priv}})}\right]$.

Since the Brenier potential ϕ of (P_X, P_Y) satisfies the curvature bounds of Assumption 2, Theorem 3 of [8] (combined with Young's inequality) gives

$$\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\right] \lesssim W_2^2(f_X, \hat{f}_{X,J,\text{priv}}) + W_2^2(f_Y, \hat{f}_{Y,J,\text{priv}}), \tag{85}$$

and we can now apply Theorem A.1 (the bound in Holder norm implies the finite Besov norm assumption by Lemma 27 in [79]) which gives that

$$\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\right] \lesssim \max\left(n^{-\frac{2(s+1)}{2s+d}}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\right) \tag{86}$$

when $d \geq 3$,

$$\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\right] \lesssim \text{PolyLog}(n,\epsilon) \times \max\left(n^{-\frac{2(s+1)}{2s+d}}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+d}}\right) \tag{87}$$

when $d = 2$, and

$$\mathbb{E}\left[\|\hat{T}_{\text{priv}}(x) - T(x)\|^2_{L^2(\hat{f}_{X,J,\text{priv}})}\right] \lesssim \max\left(n^{-1}, (\pi_{n,\epsilon})^{-\frac{2(s+1)}{s+3/2}}\right) \tag{88}$$

when $d = 1$. Combining with Equation (84) yields the stated result.

C.2 Proof of Theorem 3.2

We introduce the auxiliary map

$$\tilde{T}(x) = \begin{cases} F^{-1}_{f_Y}(1/m) \text{ if } x \in [0, F^{-1}_{f_X}(1/m)] \\ F^{-1}_{f_Y}(2/m) \text{ if } x \in (F^{-1}_{f_X}(1/m), F^{-1}_{f_X}(2/m)] \\ \dots \\ F^{-1}_{f_Y}((m-1)/m) \text{ if } x \in (F^{-1}_{f_X}((m-2)/m), F^{-1}_{f_X}((m-1)/m)] \\ 1 \text{ if } x \in (F^{-1}_{f_X}((m-1)/m), 1] \end{cases} \tag{89}$$

where $F^{-1}_{f_X}$ exists and is single value thanks to our assumption on the density of X.

Then we may decompose the error as

$$\begin{aligned}
\mathbb{E}\left[\|\hat{T} - T\|^2_{L^2(P_X)}\right] &= \mathbb{E}\left[\|(\hat{T} - \tilde{T}) - (T - \tilde{T})\|^2_{L^2(P_X)}\right] \\
&\lesssim \underbrace{\mathbb{E}\left[\|\hat{T} - \tilde{T}\|^2_{L^2(P_X)}\right]}_{\text{Sampling + Privacy noise}} + \underbrace{\mathbb{E}\left[\|T - \tilde{T}\|^2_{L^2(P_X)}\right]}_{\text{Bias}}
\end{aligned} \tag{90}$$

Bias. We control the bias as

$$\begin{aligned}\mathbb{E}\left[\|T-\tilde{T}\|^2_{L^2(P_X)}\right] &= \int_\Omega |T(x)-\tilde{T}(x)|^2 f_X(x)dx \\ &= \int_{[0,1]} |T(F_{f_X}^{-1}(u))-\tilde{T}(F_{f_X}^{-1}(u))|^2 du\end{aligned} \tag{91}$$

and we can notice that under our assumptions, for any $u \in [0,1]$, $|T(F_{f_X}^{-1}(u))-\tilde{T}(F_{f_X}^{-1}(u))|^2 \lesssim \frac{1}{m^2}$, and thus

$$\mathbb{E}\left[\|T-\tilde{T}\|^2_{L^2(P_X)}\right] \lesssim \frac{1}{m^2}\ . \tag{92}$$

Sampling and privacy noise. First, let us notice that since f_X is upper-bounded,

$$\|\hat{T}-\tilde{T}\|^2_{L^2(P_X)} \lesssim \int_\Omega \|\hat{T}(x)-\tilde{T}(x)\|^2 dx\ . \tag{93}$$

Then, we can notice that both $\hat{T}$ and $\tilde{T}$ are piecewise constant, and if both the quantiles of f_X and f_Y are estimated with a low error, they should jump to almost the same levels (depending on the error on the estimation of the quantiles of f_Y), and almost at the same time (depending on the error on the estimation of the quantiles of f_X). We formalize this reasoning below.

First, let us notice that since the densities f_X and f_Y are uniformly upper and lower bounded, there exist $0 < \alpha \leq \beta$ such that, for any i,

$$\frac{\alpha}{m} \leq F_{f_X}^{-1}((i+1)/m) - F_{f_X}^{-1}(i/m) \leq \frac{\beta}{m} \tag{94}$$

$$\frac{\alpha}{m} \leq F_{f_Y}^{-1}((i+1)/m) - F_{f_Y}^{-1}(i/m) \leq \frac{\beta}{m}\ . \tag{95}$$

Thus, if we note δ the maximal error on the estimation on the quantiles of f_X and of f_Y by $\mathbf{q}_X$ and $\mathbf{q}_Y$, we have for $\delta \leq \frac{\alpha}{2m}$

$$\begin{aligned}\int_\Omega \|\hat{T}(x)-\tilde{T}(x)\|^2 dx &\leq \underbrace{m}_{\text{Number of jumps}} \underbrace{\underbrace{(2\delta)}_{\text{Base}} \underbrace{(\frac{\beta}{m}+2\delta)^2}_{\text{Squared height}}}_{\text{Maximal error per jump}} + \underbrace{m}_{\text{number of constant parts}} \underbrace{\underbrace{(2\delta)^2}_{\text{Squared height}} \underbrace{\frac{\beta}{m}}_{\text{Base}}}_{\text{Maximal error per constant part}} \\ &\lesssim \frac{\delta}{m} + m\delta^3 + \delta^2\ .\end{aligned} \tag{96}$$

Thus, recalling that $\delta \leq \frac{\alpha}{2m}$ and denoting by B the event $\{$ "the maximal error on the estimation on the quantiles of f_X and of f_Y by $\mathbf{q}_X$ and $\mathbf{q}_Y$ is smaller δ " $\}$, we get

$$\begin{aligned}\mathbb{E}\left[\|\hat{T}-\tilde{T}\|^2_{L^2(P_X)}\right] &= \mathbb{E}\left[\|\hat{T}-\tilde{T}\|^2_{L^2(P_X)}\mathbb{1}_B\right] + \mathbb{E}\left[\|\hat{T}-\tilde{T}\|^2_{L^2(P_X)}\mathbb{1}_{B^c}\right] \\ &\lesssim \frac{\delta}{m} + m\delta^3 + \delta^2 + \mathbb{P}\left(B^c\right)\ ,\end{aligned} \tag{97}$$

which combined to the bias yields

$$\mathbb{E}\left[\|\hat{T}-T\|^2_{L^2(P_X)}\right] \lesssim \frac{1}{m^2} + \frac{\delta}{m} + m\delta^3 + \delta^2 + \mathbb{P}\left(B^c\right)\ . \tag{98}$$

Obtaining rates now boils down to tuning m and δ depending on the algorithm that is used to estimate the quantiles.

Upper bound under central DP. Under central DP, we use the algorithm of [65] to privately estimate the m quantiles (they use the add/remove neighboring relation but it remains private with replacement if one is willing to lose a factor 2 on the privacy budget). Its statistical convergence analysis was done in [74], and we can readily apply their Theorem 3.5 to control $\mathbb{P}\left(B^c\right)$.

We take $m \asymp \min(\sqrt{n}, n\epsilon)/(\log(\min(\sqrt{n}, n\epsilon)))^{2M}$ and $\delta \asymp \frac{\log_2(m)^2 \log(\min(\sqrt{n},n\epsilon))^M}{\min(\sqrt{n},n\epsilon)}$ for a fixed $M > 1$.If $\min(\sqrt{n}, n\epsilon)$ is big enough, all of the hypothesis of Theorem 3.5 in [74] are satisfied, and

$$\mathbb{P}(B^c) \lesssim C\left(n\sqrt{\min(\sqrt{n}, n\epsilon)}e^{-C\log(\min(\sqrt{n},n\epsilon))^M} + \min(\sqrt{n}, n\epsilon)e^{-C\log(\min(\sqrt{n},n\epsilon))^{2M}}\right) \quad (99)$$

for a constant $C > 0$ that depends on the problem parametters only.

Finally, since $\epsilon = \Omega(n^{-1+\kappa})$ for the constant $\kappa > 0$ of Theorem 3.2, this expression can be made as $O\big(\frac{1}{n^2} + \frac{1}{n^4\epsilon^4}\big)$ asymptotically.

Thus, by Equation (98),

$$\mathbb{E}\left[\|\hat{T} - T\|^2_{L^2(P_X)}\right] \lesssim \text{PolyLog}(n, \epsilon) \times \max\left(\frac{1}{n}, \frac{1}{n^2\epsilon^2}\right). \quad (100)$$

C.3 Proof of Theorem 4.1

The proof of the lower-bound heavily uses elements from the original lower-bound proof without differential privacy [58], and from its adaptation with central differential privacy [75]. Those two previous articles used packing-type arguments to obtain their stated lower-bounds, but they operated under slightly modified assumptions (by imposing smoothness on the transport map itself rather than on the source and target distributions). This proof thus does two things :

- It first checks that the packings used in [58] and [75] are still compatible with our modified set of hypothesis (which directly allows to obtain the lower-bound without privacy and under central DP).
- It then extends the proof to cover the local DP case.

Packing from the literature. We recall the construction of [75], the one from [58] being the same except on the placement of the bump functions. In particular, checking that the construction presented here yields a packing that is compatible with our modified hypothesis straightforwardly yields the same result for the construction of [58] when $d \geq 2$ and also for their simpler packing when $d = 1$.

Let m be an integer and $N = m^d$. [75] considers the grid $\left(\left(\frac{k_1}{m+1}, \dots, \frac{k_d}{m+1}\right)\right)_{1 \leq k_1, \dots, k_d \leq m}$ and enumerates its elements as $(p_1, \dots, p_N) \in (\mathbb{R}^d)^N$ with the property that $\forall i, j \in \{1, \dots, N\}, i \neq j \implies \|p_i - p_j\|_\infty \geq \frac{1}{m+1}$. For a C^∞ function B defined over $\mathbb{R}$ and supported on $[-1, 1]$ taking positive values over $(-1, 1)$ they define for some sufficiently small constant $a > 0$ the function $\psi(x_1, \dots, x_d) = a\prod_{i=1}^d B\big(\frac{x_i}{2}\big)$, and for some $h \in (0, 1]$ and any $\theta \in \{0, 1\}^N$ they define the potential $\phi_\theta : \Omega \to \mathbb{R}$ as $\phi_\theta(\cdot) := \frac{1}{2}\|\cdot\|^2 + h^{s+2}\sum_{i=1}^N \theta_i \psi(\frac{\cdot - p_i}{h})$. To ensure that the functions $\psi\big(\frac{\cdot - p_i}{h}\big)$ have disjoint supports for $i \in \{1, \dots, N\}$ they take $h < \frac{1}{2(m+1)} \lesssim \frac{1}{m}$.

The ϕ_θ's have a few interesting properties (all of them formally proved in [75]) : If a is taken small enough, the eigenvalues of $\nabla^2\phi_\theta(x)$ can be made arbitrarily close to 1 uniformly in x and in θ (which for this article means that we can tune a to get the required strong convexity and smoothness of the potentials), and if again a is small enough, for any θ, $\nabla\phi_\theta$ is a C^∞ diffeomorphism from Ω to Ω such that $\forall i, \nabla\phi_\theta(B_\infty(p_i, h)) = B_\infty(p_i, h)$ and that is the identity outside of those balls.

Then they define $P_X := \text{Unif}([0, 1]^d)$ and $P_{Y,\theta} := \nabla\phi_\theta \# P_X$ for any θ, and the family $((P_X, P_{Y,\theta}))_\theta$ defines the packing.

Checking the new hypothesis. We must check that this packing satisfies Assumption 2. As we already have discussed, the assumptions on the convexity of the potential are already satisfied if a is small enough. Furthermore, since the source distribution is always the uniform distribution, the condition on the Holder norm and uniform lower-bound on the density are trivially satisfied. We need to check that the same conditions hold for the target distribution.

We note f_X the density of P_X and $f_{Y,\theta}$ the density of $P_{Y,\theta}$ for any θ (that exist and can be taken continuous by the change of variable formula). Let us fix θ, the change of variable formula gives that,

for any $y \in \Omega$,

$$f_Y(y) = \frac{1}{\det(\nabla^2 \phi_\theta(\nabla \phi_\theta^{-1}(y)))} f_X(\nabla \phi_\theta^{-1}(y)) \,. \tag{101}$$

Furthermore, for any $y \in \Omega$, $f_X(\nabla \phi_\theta^{-1}(y)) = 1$, and thus $\forall y \in \Omega$, $f_Y(y) = \frac{1}{\det(\nabla^2 \phi_\theta(\nabla \phi_\theta^{-1}(y)))}$.
We now seek to apply Lemma C.1 with

$$u(x) := \frac{1}{\det(\nabla^2 \phi_\theta(x))} - 1, \qquad v(y) := \nabla \phi_\theta^{-1}(y) \tag{102}$$

and proceed that checking its hypotheses.

We start by studying v. We note that $\nabla \phi_\theta$ is a C^∞ diffeomorphism of Ω and that

$$\nabla \phi_\theta(x) = x + a h^{s+1} \sum_{i=1}^N \theta_i \nabla \psi \left(\frac{x - p_i}{h} \right). \tag{103}$$

Since the supports of the functions $\psi\big(\frac{\cdot - p_i}{h}\big)$ are disjoint and $\psi \in C_c^\infty$, we have

$$\|\nabla \phi_\theta - \mathrm{Id}\|_{C^{s+1}(\Omega)} \lesssim a, \tag{104}$$

uniformly in θ and $h \leq 1$. Thus, for a small enough, by Lemma C.2

$$\|\nabla \phi_\theta^{-1}\|_{C^{s+1}(\Omega)} \leq C, \tag{105}$$

where C does not depend on θ or h.

We now turn to the study of u. By the definition of ϕ_θ, we have

$$\nabla^2 \phi_\theta(x) = I + a h^s \sum_{i=1}^N \theta_i \nabla^2 \psi \left(\frac{x - p_i}{h} \right). \tag{106}$$

Since the supports of the bumps $\psi\big(\frac{\cdot - p_i}{h}\big)$ are disjoint, for every $x \in \Omega$ there exists at most one index $i(x) \in \{1, \dots, N\}$ such that $\nabla^2 \psi\big(\frac{x - p_{i(x)}}{h}\big) \neq 0$. Hence, for all $x \in \Omega$,

$$\nabla^2 \phi_\theta(x) = I + E_\theta(x), \qquad E_\theta(x) := \theta_{i(x)} a h^s \, \nabla^2 \psi \left(\frac{x - p_{i(x)}}{h} \right), \tag{107}$$

with the convention $E_\theta(x) = 0$ if x is outside the union of the bump supports.

We define the scalar function $F(M) := \frac{1}{\det(I+M)} - 1$ and thus have $u(x) = F\big(E_\theta(x)\big)$

We first bound E_θ in C^s. Let $m := \lfloor s \rfloor$ and $r := s - m \in [0, 1)$. For any multi-index α with $|\alpha| \leq m$ and any x such that $E_\theta(x) \neq 0$ (i.e. x belongs to one bump),

$$\partial^\alpha E_\theta(x) = a h^s \, h^{-|\alpha|} \, (\partial^\alpha \nabla^2 \psi) \left(\frac{x - p_{i(x)}}{h} \right), \tag{108}$$

hence, since $h \leq 1$ and $|\alpha| \leq m \leq s$,

$$\|\partial^\alpha E_\theta\|_{L^\infty(\Omega)} \lesssim a \, h^{s - |\alpha|} \leq a. \tag{109}$$

Moreover, if $r > 0$ and $|\alpha| = m$, for x, y inside the same bump support,

$$\begin{aligned} &\frac{\|\partial^\alpha E_\theta(x) - \partial^\alpha E_\theta(y)\|}{\|x - y\|^r} \\ &= a h^{s-m} \, \frac{\|(\partial^\alpha \nabla^2 \psi)((x - p_i)/h) - (\partial^\alpha \nabla^2 \psi)((y - p_i)/h)\|}{\|x - y\|^r} \lesssim a h^{s-m-r} = a, \end{aligned} \tag{110}$$

since $s - m = r$. If x and y lie in two different bumps, then we may swap any of them by another value in the same bump in the other one to keep the numerator of the previous ratio constant such that the denominator is smaller, granted that $h < \frac{1}{C(m+1)}$ where C is a constant that depend on the dimension only (which is possible to take by construction). C is selected so that the diameter of

the support of each bump is bigger than the minimal distance between any pair of different bump supports. Therefore,

$$\|E_\theta\|_{C^s(\Omega)} \lesssim a, \tag{111}$$

uniformly in θ and h as long as $h < \frac{1}{4(m+1)}$.

We now control $u = F \circ E_\theta$ in C^s by a direct Faà di Bruno argument. Note that F is C^∞ on $\{M : \|M\| \leq 1/2\}$, and for a small enough Equation (111) implies $\|E_\theta\|_{L^\infty} \leq 1/2$, so $F(E_\theta(x))$ is well-defined and smooth for all $x \in \Omega$. Furthermore, since $F(0) = 0$, it suffices to bound $\|F \circ E_\theta\|_{C^s}$.

Let α be a multi-index with $|\alpha| \leq m$. By the multivariate Faà di Bruno formula, $\partial^\alpha(F \circ E_\theta)$ is a finite sum (with coefficients depending only on d and $|\alpha|$) of terms of the form

$$D^k F(E_\theta(x))\Big[\partial^{\gamma_1} E_\theta(x), \ldots, \partial^{\gamma_k} E_\theta(x)\Big], \quad 1 \leq k \leq |\alpha|, \quad \sum_{j=1}^{k} |\gamma_j| = |\alpha|. \tag{112}$$

Since $\|E_\theta\|_{L^\infty} \leq 1/2$, we have $\|D^k F(E_\theta)\|_{L^\infty} \leq C_k$ for constants depending only on k and d. Using Equation (109) for all $|\gamma_j| \leq |\alpha| \leq m$, we obtain

$$\|\partial^\alpha u\|_{L^\infty} \lesssim \sum_{k=1}^{|\alpha|} \prod_{j=1}^{k} \|\partial^{\gamma_j} E_\theta\|_{L^\infty} \lesssim \sum_{k=1}^{|\alpha|} a^k \lesssim a, \tag{113}$$

uniformly in θ and h for a small enough.

If $r = 0$ we are done. Otherwise, fix $|\alpha| = m$. Using again the representation Equation (112), each term is a product of bounded coefficients $D^k F(E_\theta(\cdot))$ with multilinear forms in the derivatives $\partial^{\gamma_j} E_\theta(\cdot)$. Using iteratively the same product rule as in Equation (131) and the bounds

$$\|D^k F(E_\theta)\|_{C^r} \lesssim 1 + \|E_\theta\|_{C^r} \lesssim 1, \qquad \|\partial^\gamma E_\theta\|_{C^r} \lesssim a \quad (|\gamma| \leq m),$$

(which follow from Equation (111)), we obtain

$$\|\partial^\alpha u\|_{C^r} \lesssim \sum_{k=1}^{m} a^k \lesssim a, \tag{114}$$

uniformly in θ and $h \leq 1$.

Combining Equation (113) and Equation (114) yields

$$\left\| \frac{1}{\det(\nabla^2 \phi_\theta)} - 1 \right\|_{C^s(\Omega)} = \|u\|_{C^s(\Omega)} \lesssim a, \tag{115}$$

uniformly in θ and h for a small enough.

Applying Lemma C.1, we conclude that for a small enough,

$$\|f_{Y,\theta} - C\|_{C^s(\Omega)} = C\big\|u \circ v\big\|_{C^s(\Omega)} \lesssim \|u\|_{C^s(\Omega)}\big(1 + \|v\|^s_{C^{s+1}(\Omega)}\big) \lesssim a. \tag{116}$$

In particular, for any $\delta > 0$, choosing $a > 0$ small enough ensures

$$\sup_{\theta \in \{0,1\}^N} \|f_{Y,\theta} - C\|_{C^s(\Omega)} \leq \delta, \tag{117}$$

which shows that the packing satisfies Assumption 2.

Lower-bound with central DP. The previous reasoning shows that the packing is a valid packing for our set of hypotheses (up to a rescaling of h and by a constant and taking a small enough). The rest of the proof of [75] (i.e. their paragraph "Conclusion of the proof.") can be applied directly to prove the stated lower-bound under central DP.

Lower-bound with local DP. With local differential privacy we use Assouad's lemma (see lemma F.3 from [75] for the formal version that is an adaptation of the one from [2]).

Indeed, since by Lemma F.1 from [75] we have

$$\|\nabla\phi_\theta - \nabla\phi_{\theta'}\|^2_{L^2(P_X)} \gtrsim \mathrm{d_H}(\theta_1,\theta_2) h^{2(s+1)+d}\,, \tag{118}$$

Assouad's lemma reads

$$\begin{aligned}\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{X_1,\dots,X_n\overset{\text{iid}}{\sim}P_X, Y_1,\dots,Y_n\overset{\text{iid}}{\sim}P_{Y,\theta}}[\|\nabla\phi_\theta - \hat{T}\|^2_{L^2(P_X)}] \\ \gtrsim h^{2(s+1)+d}\sum_{i=1}^N\Big(\mathbb{P}_{\theta_{-i}}\left(\hat{\theta}^i\neq 0\right) + \mathbb{P}_{\theta_{+i}}\left(\hat{\theta}^i\neq 1\right)\Big),\end{aligned} \tag{119}$$

where $\mathbb{P}_{\theta_{+i}}$ and $\mathbb{P}_{\theta_{-i}}$ are the mixture distributions

$$\mathbb{P}_{\theta_{+i}} := \frac{1}{2^{N-1}}\sum_{\theta:\theta^{(i)}=1} P_X^{\otimes n}\otimes P_{Y,\theta}^{\otimes n}, \quad \text{and} \quad \mathbb{P}_{\theta_{-i}} := \frac{1}{2^{N-1}}\sum_{\theta:\theta^{(i)}=0} P_X^{\otimes n}\otimes P_{Y,\theta}^{\otimes n},\,, \tag{120}$$

and where

$$\hat{\theta} = \operatorname*{arg\,min}_{\theta\in\{0,1\}^N} \|\nabla\phi_\theta - \hat{T}\|^2_{L^2(P_X)} \tag{121}$$

with arbitrary choice in the argmin in case of ties, and where the upper-case i refers to the i-th component of the referenced vector.

In particular, if $\hat{T}$ is produced by an ϵ-locally differentially private mechanism, then $\hat{\theta}$, being a deterministic post-processing of $\hat{T}$, is itself the output of an ϵ-local DP channel. We will rely on the contraction result of [36]. To simplify notation, we assume in the remainder of the proof that $\hat{T}$ is obtained by applying a deterministic mapping to the output of an ϵ-local DP mechanism.

For a fixed i, we build a coupling $\mathcal{C}$ between $\mathbb{P}_{\theta_{+i}}$ and $\mathbb{P}_{\theta_{-i}}$ as follows. We first sample uniformly at random all all the components of a θ except for the i-th component, and then we make two copies of θ, on with a 1 in the i-th component (named θ), and one with a 0 (named θ'). Then we construct

$$X_1,\dots,X_n,Y_1,\dots,Y_n \overset{\text{iid}}{\sim} \underbrace{P_X^{\otimes n}\otimes P_{Y,\theta}^{\otimes n}}_{=:\mathbb{P}_\theta}, \quad X'_1,\dots,X'_n,Y'_1,\dots,Y'_n \overset{\text{iid}}{\sim} \underbrace{P_X^{\otimes n}\otimes P_{Y,\theta'}^{\otimes n}}_{=:\mathbb{P}_{\theta'}}, \tag{122}$$

this coupling doesn't matter at this stage.

Then,

$$\begin{aligned}&\mathbb{P}_{\theta_{-i}}\left(\hat{\theta}^i\neq 0\right) + \mathbb{P}_{\theta_{+i}}\left(\hat{\theta}^i\neq 1\right) \\ &= \mathbb{E}_{\theta,\theta'}\left[\mathbb{P}_\theta\left(\hat{\theta}^i\neq 0\right) + \mathbb{P}_{\theta'}\left(\hat{\theta}^i\neq 1\right)|\theta,\theta'\right]\,.\end{aligned} \tag{123}$$

Let Z denote the (possibly interactive) transcript produced by the ϵ-locally differentially private mechanism applied to the data, and let $\hat{\theta}^i = f_i(Z)$ be the i-th coordinate of the estimator, viewed as a deterministic function of Z. For $\theta\in\{0,1\}^N$, we denote by $\mathbb{P}_\theta(Z)$ the distribution of Z when the data are distributed according to $\mathbb{P}_\theta$.

Then, for any pair (θ,θ') in the previous coupling, Le Cam's lemma and Pinsker's inequality yield

$$\begin{aligned}\mathbb{P}_\theta(\hat{\theta}^i\neq 0) + \mathbb{P}_{\theta'}(\hat{\theta}^i\neq 1) &\geq 1 - \mathrm{TV}\left(f_i\#\mathbb{P}_\theta(Z),\, f_i\#\mathbb{P}_{\theta'}(Z)\right) \\ &\geq 1 - \sqrt{\mathrm{KL}\left(f_i\#\mathbb{P}_\theta(Z)||f_i\#\mathbb{P}_{\theta'}(Z)\right)/2}.\end{aligned} \tag{124}$$

Furthermore, by Corollary 1 in [36] and the data processing inequality [109] for KL,

$$\mathrm{KL}\left(f_i\#\mathbb{P}_\theta(Z)||f_i\#\mathbb{P}_{\theta'}(Z)\right) \leq 4(e^\epsilon-1)^2\big(n\,\mathrm{TV}(P_X,P_X)^2 + n\,\mathrm{TV}(P_{Y,\theta},P_{Y,\theta'})^2\big). \tag{125}$$

Finally, by Lemma F.3 from [75], we know that $\mathrm{TV}(P_{Y,\theta},P_{Y,\theta'}) \lesssim \mathrm{d_H}(\theta,\theta')h^{s+d}$, and the previous coupling we always have $\mathrm{d_H}(\theta,\theta')=1$. Thus, is ϵ is smaller than a constant, and if $h \asymp (\sqrt{n}\epsilon)^{-\frac{1}{(s+d)}}$ for a small enough multiplicative constant, we get

$$\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{X_1,\dots,X_n\overset{\text{iid}}{\sim}P_X, Y_1,\dots,Y_n\overset{\text{iid}}{\sim}P_{Y,\theta}}[\|\nabla\phi_\theta - \hat{T}\|^2_{L^2(P_X)}] \gtrsim h^{2(s+1)+d}N, \tag{126}$$

and since $N \asymp h^{-d}$, we get that

$$\sup_{\theta\in\{0,1\}^N} \mathbb{E}_{X_1,\dots,X_n \overset{\text{iid}}{\sim} P_X, Y_1,\dots,Y_n \overset{\text{iid}}{\sim} P_{Y,\theta}} [\|\nabla\phi_\theta - \hat{T}\|^2_{L^2(P_X)}] \gtrsim (\sqrt{n}\epsilon)^{-\frac{2(s+1)}{(s+d)}}, \tag{127}$$

which concludes the proof of the lower-bound under local DP

Lemmas for the proof.

Lemma C.1. *If u and v are in $C^\infty(\Omega)$ with $v(\Omega) \subseteq \Omega$ and $s \geq 0$, then*

$$\|u \circ v\|_{C^s(\Omega)} \lesssim \|u\|_{C^s(\Omega)}(1 + \max(\|v\|^s_{C^s(\Omega)}, \|v\|^s_{C^1(\Omega)})) \,. \tag{128}$$

Proof. Let $m := \lfloor s \rfloor$. Let α be a multi index of size at most m. By the Faà di Bruno formula, for any $x \in \Omega$, $\partial^\alpha(u \circ v)(x)$ is a finite sum (with coefficients and size independent of u and v) of terms of the form $\partial^k u(v(x))\Pi_{j=1}^k \partial^{\gamma_j} v(x)$ where $1 \leq k \leq m$ and where $\sum_j |\gamma_j| = m$, which gives

$$\|\partial^\alpha(u \circ v)\|_{L^\infty} \lesssim \sum_{k=1}^m \|\partial^k(u)\|_{L^\infty} \|v\|^k_{C^m} \,, \tag{129}$$

and thus

$$\|\partial^\alpha(u \circ v)\|_{L^\infty} \lesssim \|u\|_{C^s} \|v\|^m_{C^m} \leq \|u\|_{C^s}(1 + \|v\|^s_{C^s}) \,, \tag{130}$$

which in turn yields the upper-bound for the "integer" part of the Holder norm. We now look at the non-integer remainder. We define $r := s - \lfloor s \rfloor$. If $r = 0$, the proof is over. Otherwise, we fix a multi index α of size exactly $\lfloor s \rfloor$. Again, the Faà di Bruno formula says that for any $x \in \Omega$, $\partial^\alpha(u \circ v)$ is a finite sum (with coefficients and size independent of u and v) of terms of the form $\partial^k u(v(x))\Pi_{j=1}^k \partial^{\gamma_j} v(x)$ where $1 \leq k \leq m$ and where $\sum_j |\gamma_j| = m$. Furthermore, if we look at a fixed term in this sum,

$$\begin{aligned} \|x \mapsto \partial^k u(v(x))\Pi_{j=1}^k \partial^{\gamma_j} v(x)\|_{C^r} \leq{}& \|x \mapsto \partial^k u(v(x))\|_{L^\infty} \|x \mapsto \Pi_{j=1}^k \partial^{\gamma_j} v(x)\|_{C^r} \\ &+ \|x \mapsto \partial^k u(v(x))\|_{C^r} \|x \mapsto \Pi_{j=1}^k \partial^{\gamma_j} v(x)\|_{L^\infty} \,. \end{aligned} \tag{131}$$

Indeed, if f and g are two real-valued functions and $x \neq y$,

$$\begin{aligned} \frac{|(fg)(x) - (fg)(y)|}{|x-y|^r} &= \frac{|f(x)(g(x) - g(y)) + (f(x) - f(y))g(y)|}{|x-y|^r} \\ &\leq |f(x)| \frac{|g(x) - g(y)|}{|x-y|^r} + |g(y)| \frac{|f(x) - f(y)|}{|x-y|^r} \,. \end{aligned} \tag{132}$$

Furthermore,

$$\|x \mapsto \partial^k u(v(x))\|_{L^\infty} \leq \|u\|_{C^{\lfloor s \rfloor}} \,, \tag{133}$$

$$\begin{aligned} \|x \mapsto \partial^k u(v(x))\|_{C^r} &\leq \|\partial^k u\|_{C^r} \mathrm{Lip}(v)^r \\ &\lesssim \|u\|_{C^s} \|v\|^r_{C^1} \,, \end{aligned} \tag{134}$$

$$\|x \mapsto \Pi_{j=1}^k \partial^{\gamma_j} v(x)\|_{L^\infty} \leq \|v\|^{\lfloor s \rfloor}_{C^{\lfloor s \rfloor}} \,, \tag{135}$$

and

$$\|x \mapsto \Pi_{j=1}^k \partial^{\gamma_j} v(x)\|_{C^r} \lesssim \|v\|^{\lfloor s \rfloor}_{C^s} \tag{136}$$

where we used iteratively the same product rule as in Equation (131). Combining this into Equation (131) and exploiting the fact that there are only a finite number of terms in de Faà di Bruno decomposition yields

$$\|u \circ v\|_{C^s(\Omega)} \lesssim \|u\|_{C^s(\Omega)}(1 + \max(\|v\|^s_{C^s(\Omega)}, \|v\|^s_{C^1(\Omega)})) \,. \tag{137}$$

This concludes the proof. □

Lemma C.2. *Using the same notations as in the proof, if $\nabla\phi_\theta$ satisfies*

$$\|\nabla\phi_\theta - \mathrm{Id}\|_{C^{s+1}(\Omega)} \leq \epsilon, \tag{138}$$

for a small enough ϵ (depending on s, d and Ω), then

$$\|\nabla\phi_\theta^{-1}\|_{C^{s+1}(\Omega)} \leq C \tag{139}$$

where C is a constant depending only on d, s and Ω.

Proof. Write

$$\nabla\phi_\theta(x) = x + w_\theta(x), \qquad w_\theta := \nabla\phi_\theta - \mathrm{Id}. \tag{140}$$

By assumption,

$$\|w_\theta\|_{C^{s+1}(\Omega)} \le \epsilon. \tag{141}$$

Since $D(\nabla\phi_\theta)(x) = I + Dw_\theta(x)$, we have $\|Dw_\theta\|_{L^\infty(\Omega)} \le \|w_\theta\|_{C^{s+1}(\Omega)} \le \epsilon$ by definition of the Holder norms.Assume $\epsilon \le 1/2$. Then for all $x \in \Omega$,

$$\|D(\nabla\phi_\theta)(x) - I\| \le \frac{1}{2}, \tag{142}$$

hence $D(\nabla\phi_\theta)(x)$ is invertible and

$$\big\|D(\nabla\phi_\theta)(x)^{-1}\big\| \le \frac{1}{1 - \|D(\nabla\phi_\theta)(x) - I\|} \le 2. \tag{143}$$

In particular, $\nabla\phi_\theta$ is a C^1 diffeomorphism of Ω onto itself and, by the inverse function theorem, $\nabla\phi_\theta^{-1}$ is C^1. Moreover, for all $y \in \Omega$,

$$D(\nabla\phi_\theta^{-1})(y) = \Big(D(\nabla\phi_\theta)\big(\nabla\phi_\theta^{-1}(y)\big)\Big)^{-1}, \tag{144}$$

so by (143),

$$\|D(\nabla\phi_\theta^{-1})\|_{L^\infty(\Omega)} \le 2, \tag{145}$$

i.e. $\nabla\phi_\theta^{-1}$ is uniformly Lipschitz.

We will work from the implicit equation that defines the inverse. For all $y \in \Omega$,

$$\nabla\phi_\theta\big(\nabla\phi_\theta^{-1}(y)\big) = y. \tag{146}$$

Using $\nabla\phi_\theta = \mathrm{Id} + w_\theta$, this rewrites as

$$\nabla\phi_\theta^{-1}(y) + w_\theta\big(\nabla\phi_\theta^{-1}(y)\big) = y. \tag{147}$$

We now look at integer derivatives up to order $\lfloor s+1 \rfloor$. Let $m := \lfloor s+1 \rfloor$. Let α be a multi-index with $1 \le |\alpha| \le m$. Differentiating (147) yields

$$\partial^\alpha \nabla\phi_\theta^{-1}(y) + \partial^\alpha\Big(w_\theta \circ \nabla\phi_\theta^{-1}\Big)(y) = 0. \tag{148}$$

By the Faà di Bruno formula, $\partial^\alpha(w_\theta \circ \nabla\phi_\theta^{-1})$ is a finite sum of terms

$$\partial^k w_\theta\big(\nabla\phi_\theta^{-1}(y)\big) \prod_{j=1}^{k} \partial^{\gamma_j} \nabla\phi_\theta^{-1}(y), \qquad 1 \le k \le |\alpha|, \qquad \sum_{j=1}^{k} |\gamma_j| = |\alpha|. \tag{149}$$

Among these terms, the ones with $k = 1$ and $|\gamma_1| = |\alpha|$ equal $Dw_\theta(\nabla\phi_\theta^{-1}(y))\,\partial^\alpha\nabla\phi_\theta^{-1}(y)$. Thus (148) can be rewritten as

$$\Big(I + Dw_\theta\big(\nabla\phi_\theta^{-1}(y)\big)\Big)\,\partial^\alpha\nabla\phi_\theta^{-1}(y) = -R_\alpha(y), \tag{150}$$

where $R_\alpha(y)$ is a finite sum of terms of the form (149) with $k \ge 2$, hence involving only derivatives of $\nabla\phi_\theta^{-1}$ of order $< |\alpha|$.

Since $\|Dw_\theta\|_{L^\infty} \le \epsilon \le 1/2$, we have

$$\sup_{y\in\Omega} \left\| \Big(I + Dw_\theta(\nabla\phi_\theta^{-1}(y))\Big)^{-1} \right\| \le 2. \tag{151}$$

Taking L^∞ norms in (150) gives

$$\|\partial^\alpha\nabla\phi_\theta^{-1}\|_{L^\infty(\Omega)} \le 2\,\|R_\alpha\|_{L^\infty(\Omega)}. \tag{152}$$

We now argue by induction on $l := |\alpha|$. The case $l = 2$ follows from (145). Assume that for some $l \in \{2, \dots, m\}$,

$$\max_{|\gamma|\le l-1} \|\partial^\gamma\nabla\phi_\theta^{-1}\|_{L^\infty(\Omega)} \le C_{l-1}, \tag{153}$$

and let $|\alpha| = l$. We then obtain

$$\|R_\alpha\|_{L^\infty(\Omega)} \lesssim \sum_{k=2}^{l} \|\partial^k w_\theta\|_{L^\infty(\Omega)} \, C_{l-1}^k \lesssim \|w_\theta\|_{C^l(\Omega)} \, C_{l-1}^l \leq \epsilon \, C_{l-1}^l. \tag{154}$$

Plugging this into (152) yields

$$\max_{|\alpha|=l} \|\partial^\alpha \nabla \phi_\theta^{-1}\|_{L^\infty(\Omega)} \leq C(d, l, \Omega) \, \epsilon \, C_{l-1}^l. \tag{155}$$

Choosing $\epsilon > 0$ small enough (depending only on d, s, Ω) and iterating over $l = 1, \ldots, m$ yields a uniform bound

$$\max_{|\alpha|\leq m} \|\partial^\alpha \nabla \phi_\theta^{-1}\|_{L^\infty(\Omega)} \leq C. \tag{156}$$

For the non-integer part, let $r := (s+1) - m \in [0, 1)$. If $r = 0$, we are done. Otherwise fix a multi-index α with $|\alpha| = m$. From (150) we have

$$\partial^\alpha \nabla \phi_\theta^{-1}(y) = -\Big(I + Dw_\theta(\nabla \phi_\theta^{-1}(y))\Big)^{-1} R_\alpha(y). \tag{157}$$

Using the C^r product rule (as in the proof of Lemma C.1), we obtain

$$\begin{aligned} &\|\partial^\alpha \nabla \phi_\theta^{-1}\|_{C^r(\Omega)} \\ &\lesssim \left\| \Big(I + Dw_\theta \circ \nabla \phi_\theta^{-1}\Big)^{-1} \right\|_{L^\infty(\Omega)} \|R_\alpha\|_{C^r(\Omega)} + \left\| \Big(I + Dw_\theta \circ \nabla \phi_\theta^{-1}\Big)^{-1} \right\|_{C^r(\Omega)} \|R_\alpha\|_{L^\infty(\Omega)}. \end{aligned} \tag{158}$$

The first factor is bounded by 2 by (151). We now bound the remaining terms.

First, each term in R_α is a product of factors of the form $\partial^k w_\theta \circ \nabla \phi_\theta^{-1}$ (with $2 \leq k \leq m$) and $\partial^\gamma \nabla \phi_\theta^{-1}$ with $|\gamma| \leq m - 1$. Iterating the same C^r product rule and using Lemma C.1 together with (141) and the uniform C^m bound (156) yields

$$\|R_\alpha\|_{C^r(\Omega)} + \|R_\alpha\|_{L^\infty(\Omega)} \leq C \, \epsilon. \tag{159}$$

Second, since $A \mapsto (I + A)^{-1}$ is smooth on $\{A : \|A\| \leq 1/2\}$, we have

$$\left\| \Big(I + Dw_\theta \circ \nabla \phi_\theta^{-1}\Big)^{-1} \right\|_{C^r(\Omega)} \lesssim 1 + \|Dw_\theta \circ \nabla \phi_\theta^{-1}\|_{C^r(\Omega)}. \tag{160}$$

Moreover, by Lemma C.1,

$$\|Dw_\theta \circ \nabla \phi_\theta^{-1}\|_{C^r(\Omega)} \lesssim \|Dw_\theta\|_{C^r(\Omega)} \Big(1 + \|\nabla \phi_\theta^{-1}\|_{C^1(\Omega)}^r\Big) \lesssim \|w_\theta\|_{C^{s+1}(\Omega)} \leq \epsilon, \tag{161}$$

where we used (145) to control $\|\nabla \phi_\theta^{-1}\|_{C^1}$. Plugging (159)–(161) into (158) gives

$$\max_{|\alpha|=m} \|\partial^\alpha \nabla \phi_\theta^{-1}\|_{C^r(\Omega)} \leq C. \tag{162}$$

Combining (156) with the previous result yields

$$\|\nabla \phi_\theta^{-1}\|_{C^{s+1}(\Omega)} \leq C, \tag{163}$$

where C depends only on d, s and Ω. □

D Additional details on the experiments

In this section, we provide additional details on the experiments presented in Section 5, along with new examples that offer further insight into the advantages and limitations of our private transport map estimation mechanism, as well as its dependence on the various parameters involved in the estimation, such as the privacy budget, the regularization parameter, and the scale of the wavelet system. All experiments are run under central DP.

Resources. All experiments were conducted on a local server equipped with an NVIDIA RTX A6000 GPU with 48 GB of memory. The experiments displayed in Figure 1(a) and (b) took 3.11 and 9.08 hours, respectively. The experiment for the uniform-mixture transport in Figure 6 and Figure 9 took 4.85 and 5.79 hours, respectively. Finally, the large-sample experiment in Figure 5 took 13.6 hours. Thus, the total GPU time for the reported experiments was approximately 36 hours.

D.1 Distributions used for the experiments

Throughout the experiments, we solve the problem of computing the transport map from the source distribution $P = U([0,1]^d)$, to a target distribution $Q \in \{Q_a, Q_b\}$, defined in each case as the probability in $\mathbb{R}^d$ with independent marginals with distribution:

(a) $U([0,1])$

(b) Mixture of $U([0,1])$ and $\text{Beta}(3,3)$, with weights $1/5$ and $4/5$ respectively.

The OT map between P and Q_a is trivially $T = I_d$. In the second case, by independence of the marginals, the OT map is $T(x_1, \dots, x_d) = (T_1(x_1), \dots, T_1(x_d))$, where T_1 is the marginal OT map between the uniform distribution and the mixture, i.e., the mixture quantile function. Although it does not admit a closed-form expression, it can be easily approximated.

D.2 Details on the wavelet construction

For our experiments, we use the implementation of [56], which reproduces the steps of [22] to construct the orthonormal system of boundary corrected Daubechies wavelet scaling functions of order N for $L^2([0,1])$, denoted as BC-dbN, for $N \geq 1$.

If we denote by ϕ the Daubechies scaling function with support in $[0, 2N-1]$, the implementation following [56] begins by selecting j large enough to ensure that $\text{supp}(\phi_{j,0}) \subset [0,1]$, where $\phi_{j,k}(\cdot) = 2^{j/2}\phi(2^j \cdot -k)$. Given that $\text{supp}(\phi_{j,k}) \subset [\frac{k}{2^j}, \frac{k+2N-1}{2^j}]$, the previous condition translates to $j \geq J_0 := \lceil \log_2(2N-1) \rceil$. Under this condition, it follows that $\text{supp}(\phi_{j,k}) \subset [0,1]$ if and only if $0 \leq k \leq 2^j - 2N + 1$. To construct a system with 2^j functions, the set $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$ allows for $2^j - 2N + 2$ interior candidates, leaving room for $N-1$ additional boundary functions at each side. To ensure that BC-dbN spans polynomials up to degree $N-1$, the construction in [56] omits the two outermost interior functions, leaving room for N boundary corrected functions at each side. Thus, the resulting set of wavelet scaling functions can be decomposed as:

$$\Phi_j^{BC} = \{\phi_{j,k}^{BC}\}_{k=1}^{2^j} = \{\phi_{j,k}^L\}_{k=1}^N \cup \{\phi_{j,k}\}_{k=1}^{2^j-2N} \cup \{\phi_{j,k}^R\}_{k=1}^N \tag{164}$$

where $\phi_{j,k}$ is defined as before and $\phi_{j,k}^L, \phi_{j,k}^R$, denote the left and right boundary corrected scaling functions, which have support contained in $[0, \frac{2N-1}{2^j}]$ and $[1 - \frac{2N+1}{2^j}, 1]$, respectively. Following the discussion in [22, pp. 72-73], if $V_j = \text{span}(\Phi_j^{BC})$, we obtain a sequence of nested subspaces

$$V_{J_0} \subset V_{J_0+1} \subset \ldots \subset V_j \subset \ldots \tag{165}$$

which, by their Theorem 4.4, verifies $\cup_{j \geq J_0} V_j = L_2([0,1])$. Alternatively, by their Proposition 4.2 and 4.3, for any $j \geq J_0$, one can build a wavelet basis Ψ_j^{BC} of size 2^j of the subspace $V_{j+1} \setminus V_j$, of the form

$$\Psi_j^{BC} = \{\psi_{j,k}^{BC}\}_{k=1}^{2^j} = \{\psi_{j,k}^L\}_{k=1}^N \cup \{\psi_{j,k}\}_{k=1}^{2^j-2N} \cup \{\psi_{j,k}^R\}_{k=1}^N \tag{166}$$

with the same support properties as Φ_j^{BC}. Therefore, for any $J \geq J_0$, if we denote $\Upsilon_J^{BC} = \{\upsilon_{j,k}^{BC}\}_{k=1}^{2^J} := \Phi_{J_0}^{BC} \cup \left(\bigcup_{J_0 \leq j < J} \Psi_j^{BC}\right)$, then

$$V_J = \text{span}\left(\Phi_J^{BC}\right) = \text{span}\left(\Upsilon_J^{BC}\right). \tag{167}$$

Thus, we obtain two distinct bases for the subspace $V_J \subset L^2([0,1])$. The same reasoning applies in general dimension. If we denote V_J^d the subspace of $L^2([0,1]^d)$ spanned by the products $\prod_{i=1}^d f_i(x)$,

where $f_i \in V_J$, then it is generated by both tensor basis

$$\Phi_{J,d}^{BC} = \{\phi_{d,J,k}^{BC}\}_{k=1}^{2^{Jd}} := \{\prod_{i=1}^{d} f_i(x) : f_i \in \Phi_J^{BC}\} \tag{168}$$

$$\Upsilon_{J,d}^{BC} = \{\upsilon_{d,J,k}^{BC}\}_{k=1}^{2^{Jd}} := \{\prod_{i=1}^{d} f_i(x) : f_i \in \Upsilon_J^{BC}\} \tag{169}$$

Our experiments use the scaling basis $\Phi_{J,d}^{BC}$. This choice allows us to leverage the implementation of [56] provided in the Python library `boundaryWavelets`. Constructing boundary-corrected wavelets is technically involved and, to the best of our knowledge, publicly available implementations are very limited.

Although the choice between these two bases is equivalent from the perspective of $L^2([0,1]^d)$ subspace projection, a few remarks are necessary in our specific context. For simplicity, denote $\phi_k = \phi_{d,J,k}^{BC}$ and $\upsilon_k = \upsilon_{d,J,k}^{BC}$, for $k = 1, \ldots, K := 2^{Jd}$.

Differentially private estimator based on scaling functions: As in Section 2, a projection estimator for f_Z can be built as $\tilde{f}_Z := \sum_{k=1}^{K} \hat{\beta}_k \phi_k$, where $\hat{\beta}_k := \int_\Omega \phi_k dZ_n$ for each $k = 1, \ldots, K$, followed by the normalization step $\hat{f}_Z := (\tilde{f}_Z \vee 0)/\int_\Omega (\tilde{f}_Z \vee 0)$. A private version of $\hat{f}_Z$ (under either central or local DP) can be obtained by adding noise to the coefficients $(\hat{\beta}_1, \ldots, \hat{\beta}_K)$ proportional to the ℓ_1-sensitivity of the mapping $z \mapsto q(z) := \{\phi_k(z) : 1 \le k \le K\}$. First, it is easy to show that the the maximum number of overlapping functions in $\Phi_{J,d}^{BC}$ is $(3N-2)^d$. Given that

$$\sup_{1\le k\le K} \|\phi_k\|_\infty = \sup_{1\le k\le 2^{Jd}} \|\phi_{J,d,k}^{BC}\|_\infty = \sup_{1\le k\le 2^J} \|\phi_{J,k}^{BC}\|_\infty^d =: \|\Phi_J^{BC}\|_\infty^d \,, \tag{170}$$

the sensitivity of the query q is bounded by $\Delta := 2(3N-2)^d \|\Phi_J^{BC}\|_\infty^d$. Notably, the exponential dependence of the sensitivity on the dimension d is not a consequence of an suboptimal choice of the basis Φ_J^{BC}. As shown in Proposition 2.1, this dependency appears explicitly in the term $2^{(J+1)d/2}$, and the constants C_2, C_3 are similarly affected, as high-dimensional wavelet bases are typically defined as tensor products of the one-dimensional case.

Basis dependence and private density estimation: Following the construction of [22], or simply using that both basis span V_J^d and enjoy some regularity properties, it follows that there exists an orthogonal matrix $U = (u_{i,j})_{1\le i,j\le K}$ such that

$$\phi_k(x) = \sum_{i=1}^{K} u_{k,i} \upsilon_i(x) \quad \forall\, x \in [0,1]^d \,,\ 1 \le k \le K \,. \tag{171}$$

Therefore, given a sample $(Z_1, \ldots, Z_n)$, a private density estimator using $\Phi_{J,d}^{BC}$ is

$$\tilde{f}_Z = \sum_{k=1}^{k} \Big(\frac{1}{n} \sum_{j=1}^{n} \phi_k(Z_j) + L_k \Big) \phi_k \tag{172}$$

where $L_i, i = 1, \ldots, K$, are independent random variables. Then, using (171), one can conclude that

$$\tilde{f}_Z = \sum_{k=1}^{K} \left[\frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=1}^{K} u_{k,i} \upsilon_i(Z_j) \right) + L_k \right] \left(\sum_{m=1}^{K} u_{k,m} \upsilon_m \right) \tag{173}$$

$$= \sum_{m=1}^{K} \left[\frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{K} \left(\sum_{k=1}^{K} u_{k,m} u_{k,i} \right) \upsilon_i(Z_j) + \sum_{k=1}^{K} u_{k,m} L_k \right] \upsilon_m \tag{174}$$

$$= \sum_{m=1}^{K} \Big(\frac{1}{n} \sum_{j=1}^{n} \upsilon_m(Z_j) + L'_m \Big) \upsilon_m \tag{175}$$

where the last equality follows by the orthogonality of U, and $L'_m = \sum_{k=1}^{K} u_{k,m} L_k$. This highlights that in a non-private setting ($L_i = 0$), the choice of basis has no effect on the final estimator. This invariance also holds for Gaussian noise due to its rotational symmetry. However, in the Laplace case, the estimator derived from $\Phi^{BC}_{J,d}$ is not distributionally equivalent to one built from $\Upsilon^{BC}_{J,d}$. Under the orthogonal rotation U, the noise variables L'_i lose their independence and no longer follow a Laplace distribution, despite preserving the original variance. While we do not extend the formal wavelet-based theory to this setting, we expect similar empirical performance in practice.

Exact implementation constants. Throughout the different experiments, we will work with the set of d-dimensional wavelet scaling functions $\Phi^{BC}_{N,J,d}$, where we also make explicit the dependence on the initial order N of the BC-dbN system considered. In our experiments, we will work with the following admissible values of (N, J):

- $N = 2,\ J = 2 \ \Rightarrow\ \Phi^{BC}_{N,J,d}$ contains 4^d functions, and $\Delta/2 \approx 13.88^d$
- $N = 2,\ J = 3 \ \Rightarrow\ \Phi^{BC}_{N,J,d}$ contains 8^d functions, and $\Delta/2 \approx 19.63^d$
- $N = 3,\ J = 3 \ \Rightarrow\ \Phi^{BC}_{N,J,d}$ contains 8^d functions, and $\Delta/2 \approx 35.86^d$
- $N = 4,\ J = 3 \ \Rightarrow\ \Phi^{BC}_{N,J,d}$ contains 8^d functions, and $\Delta/2 \approx 61.47^d$

where Δ denotes the upper bound on the sensitivity of the query q defined in the previous section. Before addressing the density estimation problem, we first provide a visual assessment of the approximation accuracy of $\Phi^{BC}_{N,J,d}$ to estimate the density of Q_a and Q_b. To this end, one can compute the $L^2([0,1]^d)$ projection by evaluating its inner products with the functions in $\Phi^{BC}_{N,J,d}$ for different pairs (N, J). Figure 2 illustrates this approximation in the case $d = 2$.

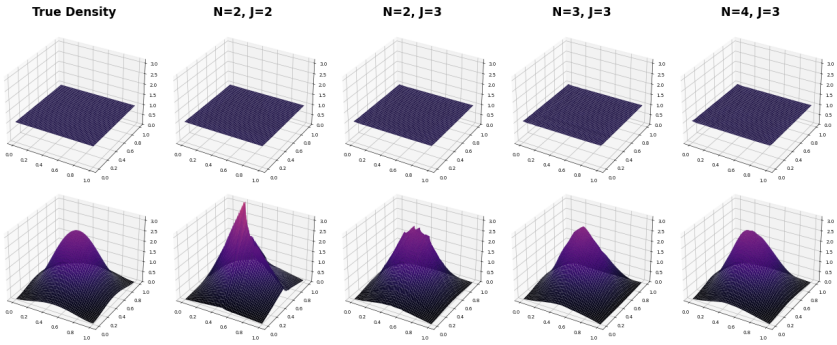

Figure 2: Projection onto the subspace of $L^2([0,1]^2)$ generated by $\Phi^{BC}_{N,J,d}$ of the densities of Q_a (top) and Q_b (bottom), for different values of (N, J).

D.2.1 Outperforming non-private empirical OT map estimation

To begin with, we consider a series of simple examples that allow us to simultaneously illustrate both the advantages and the limitations of our approach. The main takeaway is that our methodology yields a practically implementable procedure that, as the theory suggests, can even outperform the standard non-private empirical optimal transport estimator. However, the sample size required to effectively observe this improvement is typically very large, as differential privacy is strongly affected by the dimensionality and the complexity of the system under consideration.

In all examples, we consider the following setup. Let P and Q be defined as above, and let T denote the true transport map between P and Q. Consider an admissible pair (N, J). For an increasing sequence of sample sizes $0 = n_0 < n_1 < \cdots < n_K$, the comparison is carried out as follows. At step $j \in \{1, \ldots, K\}$:

- **Generation.** Draw independent samples of size $n_j - n_{j-1}$ from P and Q, and add them to the samples from the previous step to obtain samples X^{train} and Y^{train} of size n_j, respectively. In the same way, generate independent test samples X^{test} and Y^{test} of the same size.

- **OT baseline.** Compute the OT map $\hat{T}$ between X^{test} and Y^{test} without regularization (using the standard implementation of `POT` [43]) and estimate

$$\text{MSE}(\hat{T}) = \frac{1}{n_j} \sum_{i=1}^{n_j} \left\| \hat{T}(X_i^{\text{test}}) - T(X_i^{\text{test}}) \right\|^2 .$$

- **EOT baseline.** Compute the OT map $\hat{T}_{\texttt{reg}}$ between X^{train} and Y^{train} with regularization parameter $\texttt{reg} = 0.001$ (using the online batch implementation of `OTT-JAX` [26]) and estimate

$$\text{MSE}(\hat{T}_{\texttt{reg}}) = \frac{1}{n_j} \sum_{i=1}^{n_j} \left\| \hat{T}_{\texttt{reg}}(X_i^{\text{test}}) - T(X_i^{\text{test}}) \right\|^2 .$$

- **DP estimation.** For each ϵ:
  1. Compute the ϵ-DP unnormalized density estimator based on X^{train}, Y^{train} and $\Phi^{BC}_{N,J,d}$.
  2. Draw independent samples $X_\epsilon^{\text{smooth}}$ and $Y_\epsilon^{\text{smooth}}$ of size $n_j^{\text{smooth}} = 5n_j$ from the estimated densities, using rejection sampling. To prevent excessively large rejection rates, we clip the densities at 100.
  3. Compute the OT map $\hat{T}^\epsilon_{\texttt{reg}}$ between $X_\epsilon^{\text{smooth}}$ and $Y_\epsilon^{\text{smooth}}$ with regularization parameter $\texttt{reg} = 0.001$ (using `OTT-JAX` as before), and estimate

$$\text{MSE}(\hat{T}^\epsilon_{\texttt{reg}}) = \frac{1}{n_j} \sum_{i=1}^{n_j} \left\| \hat{T}^\epsilon_{\texttt{reg}}(X_i^{\text{test}}) - T(X_i^{\text{test}}) \right\|^2 .$$

Note that this comparison is not entirely fair, since the empirical map $\hat{T}$ is computed using the same points that are subsequently used to estimate the error. This follows the approach of [58] and is intended to avoid the additional error induced by using an interpolation method to define $\hat{T}$ outside the support points X^{test}.

We begin by considering the simplest two-dimensional example among all combinations in the previous section. Specifically, we take $Q = Q_a = U([0,1]^2)$ and set $N = J = 2$. Figures 1(a), 3 and 4 summarize the results of this experiment, from which several conclusions can be drawn.

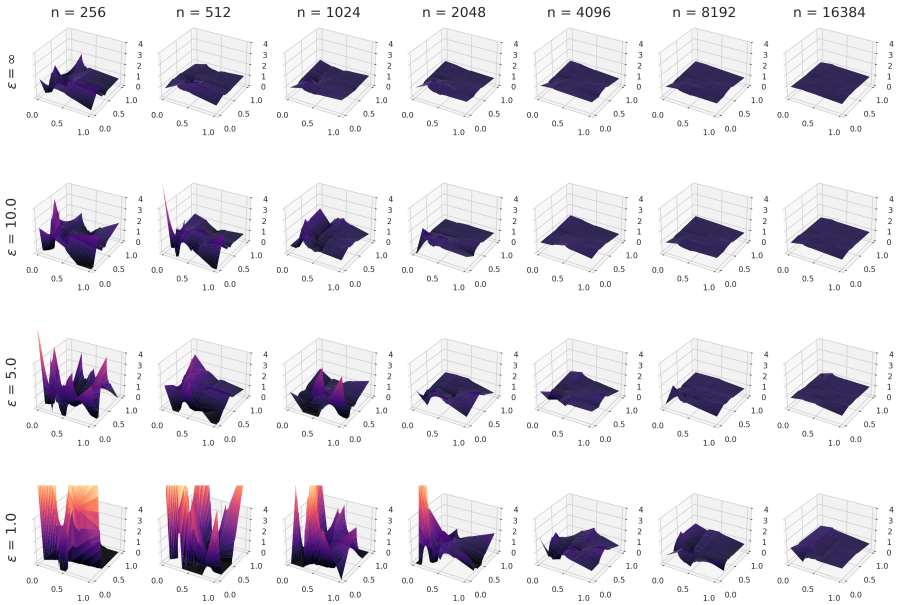

Figure 3: **d=2, uniform-uniform, N=2, J=2.** Reconstructed target densities for a single repetition across sample sizes n (columns) and central privacy budgets ϵ (rows).

Figure 1(a) shows the average mean squared error alongside the different ϵ-DP and baseline estimators. Several observations can be made. First, comparing the empirical OT map with the wavelet

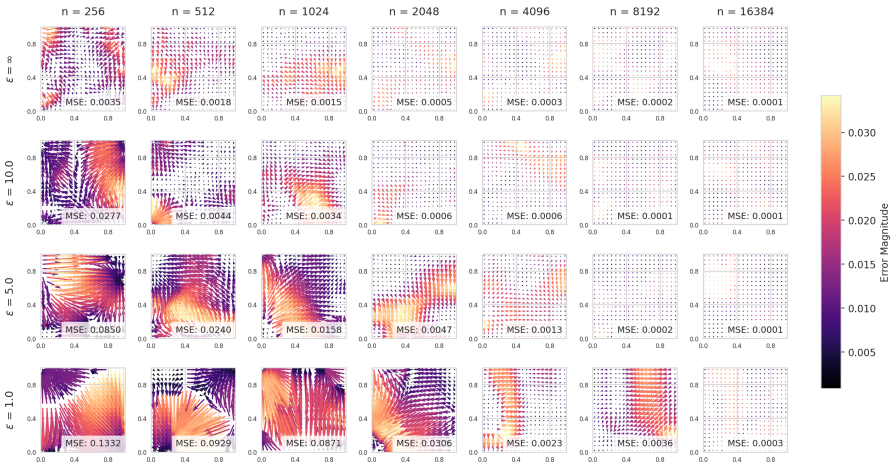

Figure 4: **d=2, uniform-uniform, N=2, J=2.** Plots of the residual $\hat{T}^{\epsilon}_{\texttt{reg}}(x) - T(x)$ across sample sizes n (columns) and central privacy budgets ϵ (rows). Vector color is proportional to the norm of the difference

estimator obtained using our procedure with $\epsilon = \infty$, it follows that the smoothing approach improves performance. Second, for small sample sizes n, the impact of privacy is severe. This degradation is a direct consequence of the large value of Δ, as discussed in the previous section. Figure 3 further illustrates this effect by showing the reconstructed target densities across different sample sizes and privacy budgets, highlighting the poor quality of the ϵ-DP density estimates for small n. Third, for sufficiently large sample sizes, the effect of privacy progressively vanishes. In this regime, the ϵ-DP estimators eventually match, and in some cases even outperform, the non-private empirical estimator, as seen in Figure 1(a).

Additional qualitative insight is provided by Figure 4, which displays the residual transport maps $\hat{T}^{\epsilon}_{\texttt{reg}}(x) - T(x)$. These plots highlight how both increasing the sample size and relaxing the privacy constraint (larger ϵ) lead to progressively more accurate transport map estimators.

The improved performance of the smoothing approach is of great interest and is consistent with the theoretical results of Section 3. However, as suggested by the figure, this improvement is not due to different estimation rates, which aligns with the fact that, in dimension $d = 2$, the empirical estimator does not suffer from the curse of dimensionality. Rather, the improvement stems from the fact that $\Phi^{BC}_{N,J,d}$ provides a good approximation of the true density; indeed, the uniform density lies in the generated subspace.

Similar behavior can be expected in more complex settings, provided that $\Phi^{BC}_{N,J,d}$ accurately captures the underlying density. We now replicate the previous experiment with $Q = Q_b$. In this case, a simple visual inspection of Figure 2 shows that the basis corresponding to $N = 2$ and $J = 2$ yields a poor approximation of the target density. This limitation directly affects the accuracy of the resulting smooth transport map estimators, as illustrated in Figures 6–8. In particular, while the effect of privacy vanishes for large sample sizes, the error induced by the density approximation persists. A natural alternative is to consider a finer basis, for instance by taking $J = 3$. However, in this case, private estimation becomes considerably more challenging, not only because of the increased value of the sensitivity bound Δ, but also because the length of the noise vector increases from 16 to 64. This substantially degrades the quality of the noisy estimates, as shown in Figures 9–11. As a result, larger sample sizes would be required to compensate for the effect of privacy, which is not compatible with the computation strategy of this section. Further analysis of the trade-off between sample size, privacy, and the choice of (N, J) is deferred to the next section.

Finally, one may consider analogous examples in dimension $d > 2$ to assess whether private estimators can still outperform the non-private baseline. However, the exponential growth of the sensitivity bound with d requires very large sample sizes, making such experiments computationally demanding. For this reason, we restrict our analysis to the case $d = 3$ and to the simplest setting in which Q is again the uniform distribution, so that it can be well approximated using the basic

scaling system with parameters $N = 2$ and $J = 2$. Even in this simple case, as noted above, we have $\Delta > 5000$. The corresponding MSE results are shown in Figure 1(b). In this experiment, we consider privacy levels $\epsilon \in \{5, 10, 30\}$. The value $\epsilon = 30$ is intentionally chosen to be very large in order to illustrate the regime in which the effect of privacy becomes negligible.

D.2.2 Large sample comparison

As discussed in the previous section, the smoothness-based approach provides an effective method to estimate the optimal transport map in the absence of privacy constraints. However, in the private setting, a practical bottleneck arises due to the large constants involved in the sensitivity bounds. The goal of this section is therefore to provide further insight into the utility gap induced by noise addition in the private framework. To this end, we consider again the mixture example, in dimension 2, with larger sample sizes. In all cases, we compare the private estimator (for fixed $\epsilon = 3$) to the non-private estimator obtained using our smoothness approach, for different values of the parameters (N, J). To enable effective computation in this large-sample regime, we do not compute baselines. In addition, we draw a test sample of fixed size $n_{\text{test}} = 10000$, as well as a smooth sample generated via rejection sampling (step 2) of fixed size $n_{\text{smooth}} = 10000$.

Fixing n_{smooth} implies that the estimator $\hat{T}^{\epsilon}_{\texttt{reg}}$, computed using a fixed number of smooth samples, will not outperform the baselines whenever $n \geq n_{\text{smooth}}$, as discussed in Section 6 in [84]. While outperforming the empirical estimator was the goal of the previous section, our objective here is different: we aim to illustrate, across several examples and for varying values of (N, J), the approximate sample size required for the effect of privacy to become negligible. For this purpose, the present experimental setting is sufficient.

Finally, in this example we also include the scaling wavelet system corresponding to $N = 1$, namely the Haar system. In the non-private setting, the resulting density estimator is exactly a histogram. In the private setting, using the scaling wavelet basis $\Phi^{BC}_{1,J,d}$ yields the standard private histogram estimator, obtained by adding independent Laplace noise to the bin weights. This is no longer the case when using the full wavelet basis $\Upsilon^{BC}_{1,J,d}$ defined in Section D.2, which includes both scaling and wavelet functions and leads to overlapping noise.

Although the theoretical results developed in this work do not apply in this case, it provides an important practical comparison. For $\Phi^{BC}_{1,J,d}$, the basis functions have disjoint supports, and it is straightforward to verify that, using standard notation, $\Delta = 2 \cdot 2^{Jd/2}$. While this quantity still exhibits exponential dependence on the dimension d, the associated constant is significantly smaller, which substantially mitigates the impact of privacy in practice. Indeed, if $\{f_1, \dots, f_K\}$ is an orthogonal system in $L^2([0,1]^d)$, then

$$\sup_{x,x' \in [0,1]^d} \|(f_1(x), \dots, f_K(x)) - (f_1(x'), \dots, f_K(x'))\|_1 \geq \tag{176}$$

$$\geq \sup_{x,x' \in [0,1]^d} \big| \|(f_1(x), \dots, f_K(x))\|_1 - \|(f_1(x'), \dots, f_K(x'))\|_1 \big| \tag{177}$$

$$\geq \sup_{x \in [0,1]^d} \|(f_1(x), \dots, f_K(x))\|_1 \tag{178}$$

$$\geq \sup_{x \in [0,1]^d} \|(f_1(x), \dots, f_K(x))\|_2 \tag{179}$$

$$= \Big(\sup_{x \in [0,1]^d} \sum_{i=1}^{K} |f_i(x)|^2 \Big)^{1/2} \tag{180}$$

$$\geq \Big(\int_{[0,1]^d} \sum_{i=1}^{K} |f_i(x)|^2 dx \Big)^{1/2} \tag{181}$$

$$= K^{1/2} \tag{182}$$

which applied to $K = 2^{Jd}$, shows that $\Phi^{BC}_{1,J,d}$ is optimal (up to a factor 2) in the sense of minimizing the noise required to obtain DP guarantees.

Figure 5 shows that, in practice, although histograms are not minimax-optimal in theory, they provide a competitive solution in the non-private setting. Among the combinations (N, J) considered, there exist values of N for which the best non-private performance is achieved by scaling wavelet systems with $N > 1$, while the Haar system remains competitive. However, in the private setting, the Haar system clearly outperforms more complex scaling wavelet systems due to the smaller constants discussed above. This behavior is also evident in Figures 12 and 13, where we display the estimated densities in the non-private and private settings, respectively.

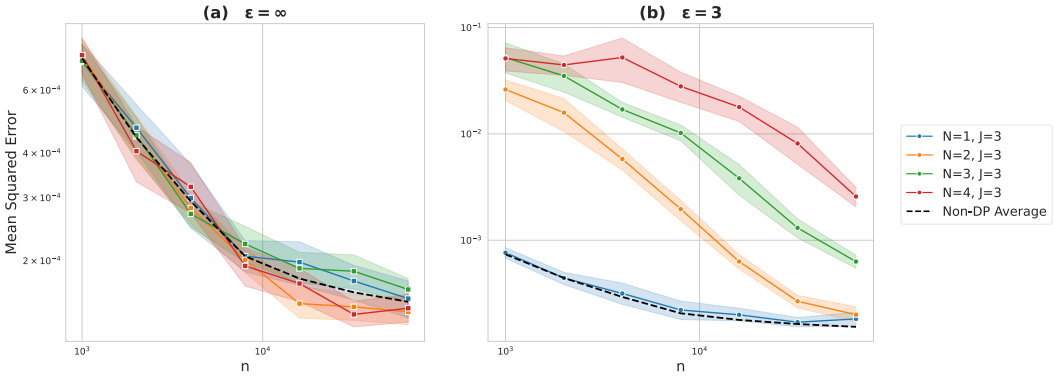

Figure 5: **d=2, uniform–mixture, varying N, J.** Mean squared error versus sample size n (log–log scale) for different parameter pairs (N, J) (colors), averaged over 10 independent runs. Results are shown for the non-private case $\epsilon = \infty$ in **(a)** and the private case $\epsilon = 3$ in **(b)**. For reference, we include a black dashed line that represents the average mean square error across all non-private models.

D.3 Additional figures

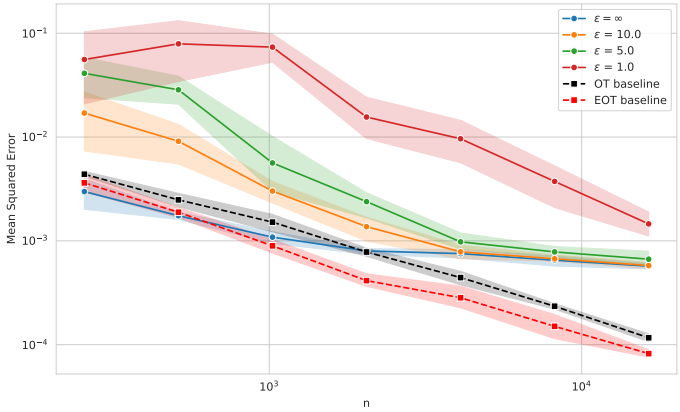

Figure 6: **d=2, uniform–mixture, N=2, J=2.** Mean squared error vs. sample size n for varying central privacy budgets ϵ (log-log scale), averaged over 5 independent runs. Solid lines represent the ϵ-DP wavelet estimator; black and red dashed lines denote the OT and EOT baselines, respectively.

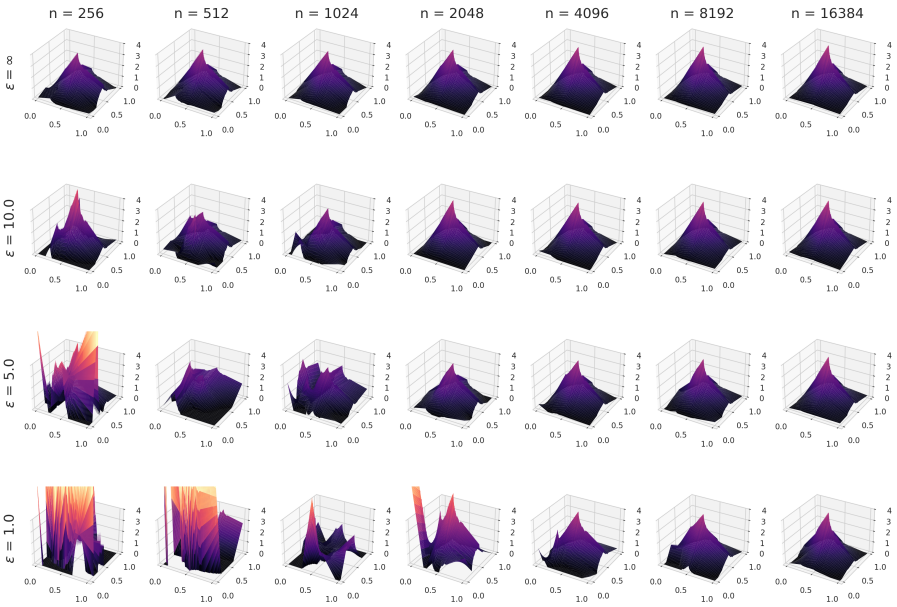

Figure 7: **d=2, uniform–mixture, N=2, J=2.** Reconstructed target densities for a single repetition across sample sizes n (columns) and central privacy budgets ϵ (rows).

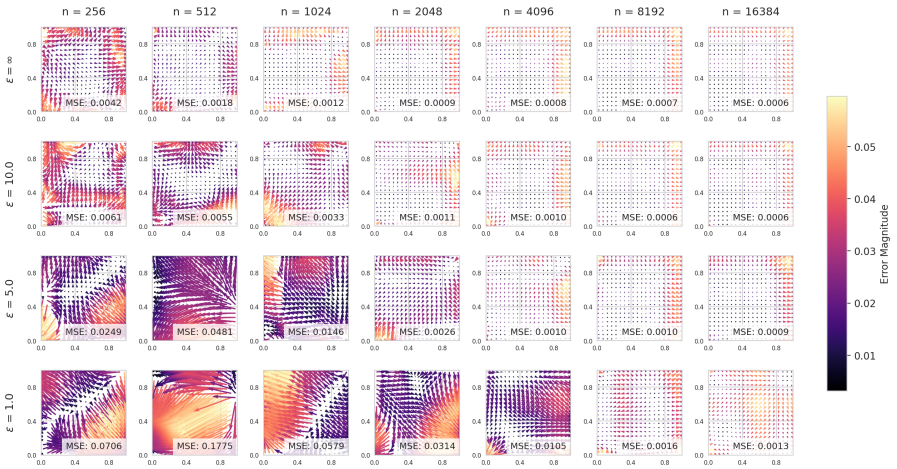

Figure 8: **d=2, uniform–mixture, N=2, J=2.** Plots of the residual $\hat{T}^{\epsilon}_{\texttt{reg}}(x) - T(x)$ across sample sizes n (columns) and central privacy budgets ϵ (rows). Vector color is proportional to the norm of the difference.

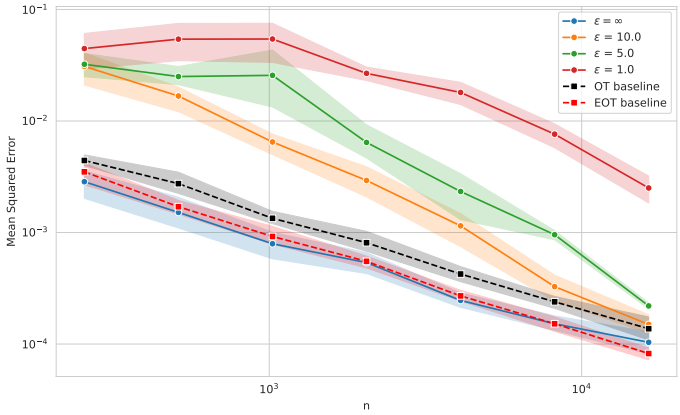

Figure 9: **d=2, uniform–mixture, N=2, J=3.** Mean squared error vs. sample size n for varying central privacy budgets ϵ (log-log scale), averaged over 5 independent runs. Solid lines represent the ϵ-DP wavelet estimator; black and red dashed lines denote the OT and EOT baselines, respectively.

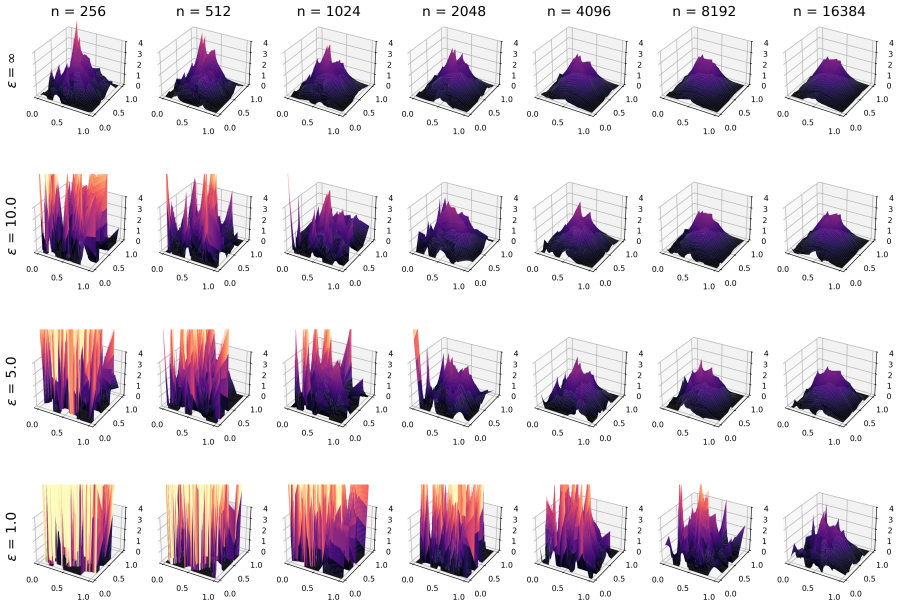

Figure 10: **d=2, uniform–mixture, N=2, J=3.** Reconstructed target densities for a single repetition across sample sizes n (columns) and central privacy budgets ϵ (rows).

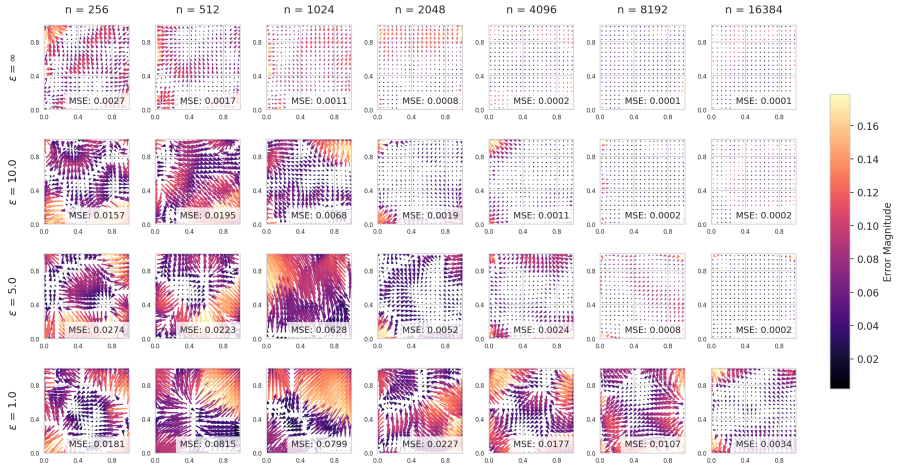

Figure 11: **d=2, uniform–mixture, N=2, J=3.** Plots of the residual $\hat{T}^{\epsilon}_{\texttt{reg}}(x) - T(x)$ across sample sizes n (columns) and central privacy budgets ϵ (rows). Vector color is proportional to the norm of the difference

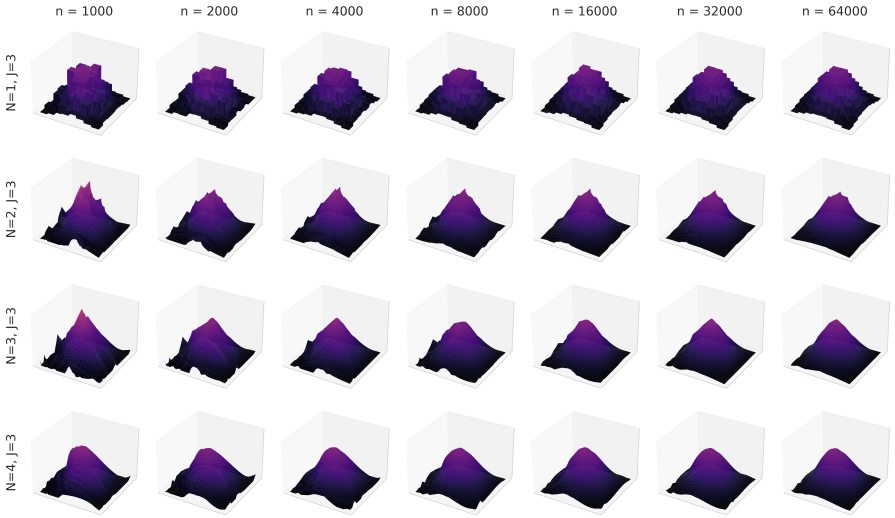

Figure 12: **d=2, uniform–mixture, varying N, J.** Reconstructed target densities for a single repetition across sample sizes n (columns) and values of (N, J) (rows), for $\epsilon = \infty$.

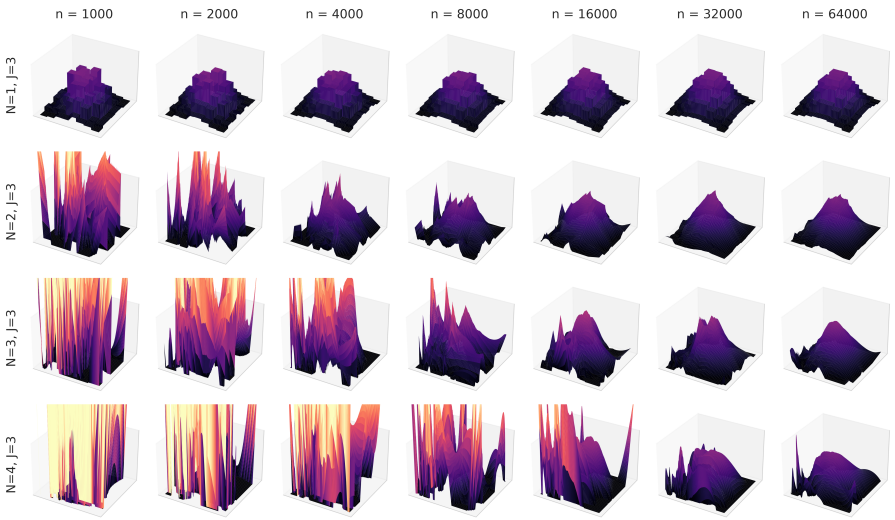

Figure 13: **d=2, uniform–mixture, varying N, J.** Reconstructed target densities for a single repetition across sample sizes n (columns) and values of (N, J) (rows), for central $\epsilon = 3$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main optimality claims made in the abstract are substantiated by Theorem 3.1, Theorem 3.2 and Theorem 4.1.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The two main limitations of this work are as follows. First, we do not establish optimality in the local-DP case when $d = 1$, as discussed in Section 4. Second, when turning our main theoretical estimator into a practical algorithm, the resampling step remains the dominant computational bottleneck, and we leave its improvement as an interesting direction for future work. Regarding the local-DP, $d = 1$ case, we emphasize that our construction is modular: any significant progress on the problem of locally private quantiles estimation can be plugged into our reduction to yield sharper guarantees in this regime.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Every theorem is stated under explicit assumptions, and we provide complete proofs for all results, or reference the adequate literature when it straightforwardly leads to the stated result.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All experimental details are provided either in the main paper or, when space is limited, in the appendix. In addition, the full code used to run our experiments is included in the supplementary material of the submission.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All experimental resources used in this work are fully disclosed. The released code includes not only the implementation of our methods and baselines, but also the routines used to generate all synthetic datasets.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All of this information is provided either in the main body of the paper or, when space constraints require it, in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Our experiments report averages over multiple runs, and all plots include confidence intervals.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The experimental section reports both the hardware configuration used for all runs and the corresponding runtimes.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification:

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

Justification: The present work advances the study of privacy-preserving learning. However, we do not believe that it raises broader societal impacts beyond those already associated with differential privacy research more generally.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This work does not introduce or release new datasets or trained models.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: Beyond standard citation practices for related work, the paper also gives explicit credit to the original authors of the algorithmic implementations used, whenever these implementations are not considered standard or ubiquitous.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.

- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release additional assets beyond the code required to reproduce the experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or research involving human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.