Algorithmically Effective Differentially Private Synthetic Data

Yiyun He YIYUNH4@UCI.EDU
Roman Vershynin RVERSHYN@UCI.EDU
Yizhe Zhu YIZHE.ZHU@UCI.EDU
University of California, Irvine

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

We present a highly effective algorithmic approach for generating ε-differentially private synthetic data in a bounded metric space with near-optimal utility guarantees under the 1-Wasserstein distance. In particular, for a dataset $\mathcal{X}$ in the hypercube $[0,1]^d$, our algorithm generates synthetic dataset $\mathcal{Y}$ such that the expected 1-Wasserstein distance between the empirical measure of $\mathcal{X}$ and $\mathcal{Y}$ is $O((\varepsilon n)^{-1/d})$ for $d \geq 2$, and is $O(\log^2(\varepsilon n)(\varepsilon n)^{-1})$ for $d = 1$. The accuracy guarantee is optimal up to a constant factor for $d \geq 2$, and up to a logarithmic factor for $d = 1$. Our algorithm has a fast running time of $O(\varepsilon dn)$ for all $d \geq 1$ and demonstrates improved accuracy compared to the method in (Boedihardjo et al., 2022c) for $d \geq 2$.

Keywords: differential privacy, synthetic data, Wasserstein metric

1. Introduction

Differential privacy has become the benchmark for privacy protection in scenarios where vast amounts of data need to be analyzed. The aim of differential privacy is to prevent the disclosure of information about individual participants in the dataset. In simple terms, an algorithm that has a randomized output and produces similar results when given two adjacent datasets is considered to be differentially private. This method of privacy protection is increasingly being adopted and implemented in various fields, including the 2020 US Census (Abowd et al., 2019; Hawes, 2020; Hauer and Santos-Lozada, 2021) and numerous machine learning tasks (Dwork and Roth, 2014).

A wide range of data computations can be performed in a differentially private manner, including regression (Chaudhuri and Monteleoni, 2008), clustering (Su et al., 2016), parameter estimation (Duchi et al., 2018), stochastic gradient descent (Song et al., 2013), and deep learning (Abadi et al., 2016). However, many existing works on differential privacy focus on designing algorithms for specific tasks and are restricted to queries that are predefined before use. This requires expert knowledge and often involves modifying existing algorithms.

One promising solution to this challenge is to generate a synthetic dataset similar to the original dataset with guaranteed differential privacy (Hardt et al., 2012; Blum et al., 2013; Jordon et al., 2019; Bellovin et al., 2019; Boedihardjo et al., 2022a,b,c). As any downstream tasks are based on the synthetic dataset, they can be performed without incurring additional privacy costs.

1.1. Private synthetic data

Mathematically, the problem of generating private synthetic data can be defined as follows. Let (Ω, ρ) be a metric space. Consider a dataset $\mathcal{X} = (X_1, \dots, X_n) \in \Omega^n$. Our goal is to construct an efficient randomized algorithm that outputs differentially private synthetic data $\mathcal{Y} = (Y_1, \dots, Y_m) \in \Omega^m$ such that the two empirical measures

$$\mu_{\mathcal{X}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \quad \text{and} \quad \mu_{\mathcal{Y}} = \frac{1}{m}\sum_{i=1}^{m} \delta_{Y_i}$$

are close to each other. We measure the utility of the output by $\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$, where $W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}})$ is the 1-Wasserstein distance, and the expectation is taken over the randomness of the algorithm. The Kantorovich-Rubinstein duality (see, e.g., (Villani, 2009)) gives an equivalent representation of the 1-Wasserstein distance between two measures $\nu_{\mathcal{X}}$ and $\mu_{\mathcal{Y}}$:

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) = \sup_{\mathrm{Lip}(f)\le 1} \left(\int f \mathrm{d}\mu_{\mathcal{X}} - \int f \mathrm{d}\mu_{\mathcal{Y}} \right), \tag{1.1}$$

where the supremum is taken over the set of all 1-Lipschitz functions on Ω. Since many machine learning algorithms are Lipschitz (von Luxburg and Bousquet, 2004; Kovalev, 2022; Bubeck and Sellke, 2021; Meunier et al., 2022), Equation (1.1) provides a uniform accuracy guarantee for a wide range of machine learning tasks performed on synthetic datasets whose empirical measure is close to $\mu_{\mathcal{X}}$ in the 1-Wasserstein distance.

1.2. Main results

The most straightforward way to construct differentially private synthetic data is to add independent noise to the location of each data point. However, this method can result in a significant loss of data utility as the amount of noise needed for privacy protection may be too large (Domingo-Ferrer et al., 2021). Another direct approach could be to add noise to the density function of the empirical measure of $\mathcal{X}$, by dividing Ω into small subregions and perturbing the true counts in each subregion. However, Laplacian noise may perturb the count in a certain subregion to negative, causing the output to become a signed measure. To address this issue, we introduce an algorithmically effective method called the *Private Measure Mechanism*.

Private Measure Mechanism (PMM) PMM makes the count zero if the noisy count in a subregion is negative. Instead of a single partition of Ω, we consider a collection of binary hierarchical partitions on Ω and add inhomogeneous noise to each level of the partition. However, the counts of two subregions do not always add up to the count of the region at a higher level. We develop an algorithm that enforces the consistency of counts in regions at different levels. PMM has $O(\varepsilon d n)$ running time while the running time of the approach in (Boedihardjo et al., 2022c) is polynomial in n.

The accuracy analysis of PMM uses the hierarchical partitions to estimate the 1-Wasserstein distance in terms of the multi-scale geometry of Ω and the noise magnitude in each level of the partition. In particular, when $\Omega = [0, 1]^d$, by optimizing the choice of the hierarchical partitions and noise magnitude, PMM achieves better accuracy compared to (Boedihardjo et al., 2022c) for $d \ge 2$. The accuracy is optimal rate up to a constant factor for $d \ge 2$, and up to a logarithmic factor for $d = 1$. We state it in the next theorem.

The hierarchical partitions appeared in many previous works on the approximation of distributions under Wasserstein distances in a non-private setting, including (Ba et al., 2011; Dereich et al., 2013; Weed and Bach, 2019). In the differential privacy literature, the hierarchical partitions are also closely related to the binary tree mechanism (Dwork et al., 2010; Chan et al., 2011) for differential privacy under continual observation. However, the accuracy analysis of the two mechanisms is significantly different. In addition, the TopDown algorithm in the 2020 census (Abowd et al., 2022) also has a similar hierarchical structure and enforces consistency, but the accuracy analysis of the algorithm is not provided in (Abowd et al., 2022).

Theorem 1 (PMM for data in a hypercube) *Let $\Omega = [0,1]^d$ equipped with the ℓ^∞ metric. PMM outputs an ε-differentially private synthetic dataset $\mathcal{Y}$ in time $O(\varepsilon dn)$ such that*

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \leq \begin{cases} C \log^2(\varepsilon n)(\varepsilon n)^{-1} & \text{if } d = 1, \\ C(\varepsilon n)^{-\frac{1}{d}} & \text{if } d \geq 2. \end{cases}$$

Private Signed Measure Mechanism (PSMM) In addition to PMM, we introduce an alternative method, the *Private Signed Measure Mechanism*, that achieves optimal accuracy rate on $[0,1]^d$ when $d \geq 3$ in poly(n) time. The analysis of PSMM is not restricted to 1-Wasserstein distance, and it can be generalized to provide a uniform utility guarantee of other function classes.

We first partition the domain Ω into m subregions $\Omega_1, \dots, \Omega_m$. Perturbing the counts in each subregion with i.i.d. integer Laplacian noise gives an unbiased approximation of $\mu_{\mathcal{Y}}$ with a signed measure ν. Then we find the closest probability measure $\hat{\nu}$ under the bounded Lipschitz distance by solving a linear programming problem.

In the proof of accuracy for PSMM, one ingredient is to estimate the Laplacian complexity of the Lipschitz function class on Ω and connect it to the 1-Wasserstein distance. This type of argument is similar in spirit to the optimal matching problem for two sets of random points in a metric space (Talagrand, 1992, 2005; Bobkov and Ledoux, 2021). When $\Omega = [0,1]^d$, PSMM achieves the optimal accuracy rate $O((\varepsilon n)^{-1/d})$ for $d \geq 3$. For $d = 2$, PSMM achieves a near-optimal accuracy $O(\log(\varepsilon n)(\varepsilon n)^{-1/2})$. For $d = 1$, the accuracy becomes $O((\varepsilon n)^{-1/2})$.

For the case when $d = 2$, we believe that the bound in Corollary 8 could be improved to $C\sqrt{\log(\varepsilon n)}/\sqrt{\varepsilon n}$ by replacing Dudley's chaining bound in Proposition 3 with the generic chaining bound in (Talagrand, 2005; Dirksen, 2015) involving the γ_1 and γ_2 functionals on Ω. We will not pursue this direction in this paper.

Comparison to previous results (Ullman and Vadhan, 2011) proved that it is NP-hard to generate private synthetic data on the Boolean cube which approximately preserves all two-dimensional marginals, assuming the existence of one-way functions. There exists a substantial body of work for differentially private synthetic data with guarantees limited to accuracy bounds for a finite set of specified queries (Barak et al., 2007; Thaler et al., 2012; Dwork et al., 2015; Vadhan, 2017; Liu et al., 2021; Vietri et al., 2022; Boedihardjo et al., 2022c,d, 2023).

(Wang et al., 2016) considered differentially private synthetic data in $[0,1]^d$ with guarantees for any smooth queries with bounded partial derivatives of order K, and achieved an accuracy of $O(\varepsilon^{-1} n^{-\frac{K}{2d+K}})$. Recently, (Boedihardjo et al., 2022c) introduced a method based on superregular random walks to generate differentially private synthetic data with near-optimal guarantees in general compact metric spaces. In particular, when the dataset is in $[0,1]^d$, they obtain $\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \leq C \log^{\frac{3}{2d}}(\varepsilon n)(\varepsilon n)^{-\frac{1}{d}}$. A corresponding lower bound of order $n^{-1/d}$ was also proved in (Boedihardjo et al., 2022c, Corollary 9.3). PMM matches the lower bound up to a constant factor for $d \geq 2$, and up to a logarithmic factor for $d = 1$.

In terms of computational efficiency, PMM runs in time $O(\varepsilon dn)$. This is more efficient compared to the algorithm in (Boedihardjo et al., 2022c).

Organization of the paper The rest of the paper is organized as follows. In Section 2, we introduce some background on differential privacy and distances between measures. We will first introduce and analyze the easier and more direct method PSMM before our main result. In Section 3, we describe PSMM in detail and prove its privacy and accuracy for data in a bounded metric space, and detailed results are provided for the case for the hypercube. In Section 4, we introduce PMM and analyze its privacy and accuracy. Optimizing the choices of noise parameters, we obtain the optimal accuracy on the hypercube with $O(\varepsilon dn)$ running time, which proves Theorem 1.

Additional proofs are included in Appendix A. We use a variant of Laplacian distribution, called *discrete Laplacian distribution*, in PMM and PSMM. The definition and properties of discrete Laplacian distribution are included in Appendix B.

2. Preliminaries

Differential Privacy We use the following definition from (Dwork and Roth, 2014). A randomized algorithm $\mathcal{M}$ provides ε*-differential privacy* if for any input data D, D' that differs on only one element (or D and D' are adjacent data sets) and for any measurable set $S \subseteq \operatorname{range}(\mathcal{M})$, there is

$$\frac{\mathbb{P}\left\{\mathcal{M}(D) \in S\right\}}{\mathbb{P}\left\{\mathcal{M}(D') \in S\right\}} \leq e^{\varepsilon}.$$

Here the probability is taken from the probability space of the randomness of $\mathcal{M}$.

Wasserstein distance Consider a metric space (Ω, ρ) with two probability measures μ, ν. Then the 1-Wasserstein distance (see e.g., (Villani, 2009) for more details) between them is defined as

$$W_1(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{\Omega\times\Omega} \rho(x,y)\mathrm{d}\gamma(x,y),$$

where $\Gamma(\mu, \nu)$ is the set of all couplings of μ and ν.

Bounded Lipschitz distance Let (Ω, ρ) be a bounded metric space. The *Lipschitz norm* of a function f is defined as

$$\|f\|_{\mathrm{Lip}} := \max\left\{\mathrm{Lip}(f),\ \frac{\|f\|_\infty}{\mathrm{diam}(\Omega)}\right\},$$

where $\mathrm{Lip}(f)$ is the Lipschitz constant of f. Let $\mathcal{F}$ be the set of all Lipschitz functions f on Ω with $\|f\|_{\mathrm{Lip}} \leq 1$. For signed measures μ, ν, we define the *Bounded Lipschitz distance*:

$$d_{\mathrm{BL}}(\mu,\nu) := \sup_{f\in\mathcal{F}} \left(\int f\mathrm{d}\mu - \int f\mathrm{d}\nu\right).$$

One can easily check that this is a metric. Moreover, in the special case where μ and ν are both probability measures, moving f by a constant does not change the result of $\int f\mathrm{d}\mu - \int f\mathrm{d}\nu$. Therefore, for a bounded domain Ω, we can always assume $f(x_0) = 0$ for a fixed $x_0 \in \Omega$, then $\|f\|_\infty \leq \mathrm{diam}(\Omega)$ when computing the supremum in (1.1). This implies d_{BL}-metric is equivalent to the classical W_1-metric when μ, ν are both probability measures on a bounded domain Ω:

$$W_1(\mu,\nu) = \sup_{\mathrm{Lip}(f)\leq 1} \left(\int f\mathrm{d}\mu - \int f\mathrm{d}\nu\right) = \sup_{f\in\mathcal{F}} \left(\int f\mathrm{d}\mu - \int f\mathrm{d}\nu\right) = d_{\mathrm{BL}}(\mu,\nu). \tag{2.1}$$

3. Private signed measure mechanism (PSMM)

We will first introduce PSMM, which is an easier and more intuitive approach. The procedure of PSMM is formally described in Algorithm 1. Note that in the *output* step of Algorithm 1, the size of the synthetic data m' depends on the rational approximation of the density function of $\hat{\nu}$, and we discuss the details here. Let $\hat{v}_1, \dots, \hat{v}_m$ be the weight of the probability measure $\hat{\nu}$ on $y_1, \dots, y_m$, respectively. We can choose rational numbers $r_1, \dots, r_m$ such that $\max_{i\in[m]} |r_i - \hat{\nu}_i|$ is arbitrarily small. Let m' be the least common multiple of the denominators of $r_1, \dots, r_m$, then we output the synthetic dataset $\hat{\mathcal{Y}}$ containing $m'r_i$ copies of y_i for $i = 1, \dots, m$.

Before analyzing the privacy and accuracy of PSMM, we introduce a useful complexity measure of a given function class, which quantifies the influence of the Laplacian noise on the function class.

Algorithm 1 Private Signed Measure Mechanism

Input: true data $\mathcal{X} = (x_1, \dots, x_n) \in \Omega^n$, partition $(\Omega_1, \dots, \Omega_m)$ of Ω, privacy parameter $\varepsilon > 0$.

Compute the true counts Compute the true count in each regime $n_i = \#\{x_j \in \Omega_i : j \in [n]\}$.

Create a new dataset Choose any element $y_i \in \Omega_i$ independently of $\mathcal{X}$, and let $\mathcal{Y}$ be the collection of n_i copies of y_i for each $i \in [n]$.

Add noise Perturb the empirical measure $\mu_{\mathcal{Y}}$ of $\mathcal{Y}$ and obtain a signed measure ν such that

$$\nu(\{y_i\}) := (n_i + \lambda_i)/n,$$

where $\lambda_i \sim \mathrm{Lap}_{\mathbb{Z}}(1/\varepsilon)$ are i.i.d. discrete Laplacian random variables.

Linear programming Find the closest probability measure $\hat{\nu}$ of ν in d_{BL}-metric using Algorithm 2, and generate synthetic data $\widehat{\mathcal{Y}}$ from $\hat{\nu}$.

Output: synthetic data $\widehat{\mathcal{Y}} = (y_1, \dots, y_{m'}) \in \Omega^{m'}$ for some integer m'.

3.1. Laplacian complexity

Given the Kantorovich-Rubinstein duality (1.1), to control the W_1-distance between the original measure and the private measure, we need to describe how Lipschitz functions behave under Laplacian noise. As an analog of the worst-case Rademacher complexity (Bartlett and Mendelson, 2002; Foster and Rakhlin, 2019), we consider the worst-case Laplacian complexity. Such a worst-case complexity measure appears since the original dataset is deterministic without any distribution assumption.

Definition 2 (Worst-case Laplacian complexity) *Let $\mathcal{F}$ be a function class on a metric space Ω. The worst-case Laplacian complexity of $\mathcal{F}$ is defined by*

$$L_n(\mathcal{F}) := \sup_{X_1,\dots,X_n\in\Omega} \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \lambda_i f(X_i)\right|\right], \tag{3.1}$$

where $\lambda_1, \dots, \lambda_n \sim \mathrm{Lap}(1)$ are i.i.d. random variables.

Algorithm 2 Linear Programming

Input: A discrete signed measure ν supported on $\mathcal{Y} = \{y_1, \dots, y_m\}$.

Compute the distances Compute the pairwise distances $\{\|y_i - y_j\|_\infty, i > j\}$.

Solve the linear programming Solve the linear programming problem with $2m^2$ variables and $m + 1$ constraints:

$$\begin{aligned} \min \quad & \sum_{i,j=1}^{m} \|y_i - y_j\|_\infty (u_{ij} + u'_{ij}) + 2\, v_i \\ \text{s.t.} \quad & \sum_{j=1}^{m} (u_{ij} - u'_{ij}) + v_i + \tau_i \geq \nu(\{y_i\}), && \forall i \leq m, \\ & \sum_{i=1}^{m} \tau_i = 1, \\ & u_{ij}, u'_{ij}, v_i, \tau_i \geq 0, && \forall i, j \leq m, i \neq j. \end{aligned}$$

Output: a probability measure $\hat{\nu}$ with $\hat{\nu}(\{y_i\}) = \tau_i$.

Since Laplacian random variables are sub-exponential but not sub-gaussian, its complexity measure is not equivalent to the Gaussian or Rademacher complexity, but it is related to the suprema of the mixed tail process (Dirksen, 2015) and the quadratic empirical process (Mendelson, 2010). Our next proposition bounds $L_n(\mathcal{F})$ in terms of the covering numbers of $\mathcal{F}$. Its proof is a classical application of Dudley's chaining method (see, e.g., (Vershynin, 2018)).

Proposition 3 (Bounding Laplacian complexity with Dudley's entropy integral) *Suppose that (Ω, ρ) is a metric space and $\mathcal{F}$ is a set of functions on Ω. Then*

$$L_n(\mathcal{F}) \leq C \inf_{\alpha > 0} \left(2\alpha + \frac{1}{\sqrt{n}} \int_\alpha^\infty \sqrt{\log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty)} \mathrm{d}u + \frac{1}{n} \int_\alpha^\infty \log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \mathrm{d}u \right)$$

where $\mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty)$ is the covering number of $\mathcal{F}$ and $C > 0$ is an absolute constant.

In particular, we are interested in the case where $\mathcal{F}$ is the class of all the bounded Lipschitz functions. One can find the result in (Tikhomirov, 1993) or more explicit bound in (Gottlieb et al., 2016) that for the set $\mathcal{F}$ of functions f with $\|f\|_{\mathrm{Lip}} \leq 1$, its covering number satisfies

$$\mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \leq \left(\frac{8}{u} \right)^{\mathcal{N}(\Omega, u/2, \rho)}.$$

When $\Omega = [0, 1]^d$, a better bound on the covering number for Lipschitz functions is available from (Tikhomirov, 1993; von Luxburg and Bousquet, 2004):

$$\mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \leq \Big(2 \left\lceil 2/u \right\rceil + 1 \Big) \, 2^{\mathcal{N}\left([0,1]^d, u/2, \|\cdot\|_\infty\right)},$$

which implies the following corollary.

Corollary 4 (Laplacian complexity for Lipschitz functions on the hypercube) *Let $\Omega = [0,1]^d$ with the $\|\cdot\|_\infty$ metric, and $\mathcal{F}$ be the set of all Lipschitz functions f on Ω with $\|f\|_{\mathrm{Lip}} \leq 1$. We have*

$$L_n(\mathcal{F}) \leq \begin{cases} Cn^{-1/2} & \text{if } d = 1, \\ C\log n \cdot n^{-1/2} & \text{if } d = 2, \\ Cd^{-1}n^{-1/d} & \text{if } d \geq 3. \end{cases}$$

Discrete Laplacian complexity Laplacian complexity can be useful for differential privacy algorithms based on the Laplacian mechanism (Dwork and Roth, 2014). However, since PSMM perturbs counts in each subregion, it is more convenient for us to add integer noise to the true counts. Instead, we will use the *worst-case discrete Laplacian complexity* defined below:

$$\widetilde{L}_n(\mathcal{F}) := \sup_{X_1,\ldots,X_n\in\Omega} \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \lambda_i f(X_i)\right|\right], \tag{3.2}$$

where $\lambda_1, \ldots, \lambda_n \sim \mathrm{Lap}_{\mathbb{Z}}(1)$ are i.i.d. discrete Laplacian random variables.

In particular, $\mathrm{Lap}_{\mathbb{Z}}(1)$ has a bounded sub-exponential norm, therefore the proof of Proposition 3 works for discrete Laplacian random variables as well. Consequently, Corollary 4 also holds for $\widetilde{L}_n(\mathcal{F})$, with a different absolute constant C.

3.2. Privacy and Accuracy of Algorithm 1

The privacy guarantee of Algorithm 1 can be proved by checking the definition. The essence of the proof is the same as the classical Laplacian mechanism (Dwork and Roth, 2014).

Proposition 5 (Privacy of Algorithm 1) *Algorithm 1 is ε-differentially private.*

We now turn to accuracy. The linear programming problem stated in Algorithm 2 has $(2m^2 + 2m)$ many variables and $(m+1)$ many constraints, which can be solved in polynomial time in m. We first show that Algorithm 2 indeed outputs the closest probability measure to ν in the d_{BL}-distance in the next proposition.

Proposition 6 *For a discrete signed measure ν on Ω, Algorithm 2 gives its closest probability measure in d_{BL}-distance with the same support set with a polynomial running time in m.*

Now we are ready to analyze the accuracy of Algorithm 1. In PSMM, independent Laplacian noise is added to the count of each sub-region. Therefore, the Laplacian complexity arises when considering the expected Wasserstein distance between the original empirical measure and the synthetic measure.

Theorem 7 (Accuracy of Algorithm 1) *Suppose $(\Omega_1, \ldots, \Omega_m)$ is a partition of (Ω, ρ) and $\mathcal{F}$ is the set of all functions with Lipschitz norm bounded by 1. Then the measure $\hat{\nu}$ generated from Algorithm 1 satisfies*

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \hat{\nu}) \leq \max_i \mathrm{diam}(\Omega_i) + \frac{2m}{\varepsilon n}\widetilde{L}_m(\mathcal{F}).$$

Note that $\mathrm{diam}(\Omega_i) \asymp m^{-1/d}$ can be satisfied when we take a partition of $\Omega = [0,1]^d$ where each Ω_i is a subcube of the same size. Using the formula above and the result of Laplacian complexity for the hypercube in Corollary 4, one can easily deduce the following result.

Corollary 8 (Accuracy of Algorithm 1 on the hypercube) *Take $m = \lceil \varepsilon n \rceil$ and let $(\Omega_1, \ldots, \Omega_m)$ be a partition of $\Omega = [0,1]^d$ with the norm $\|\cdot\|_\infty$. Assume that $\mathrm{diam}(\Omega_i) \asymp m^{-1/d}$. Then the measure $\hat{\nu}$ generated from Algorithm 1 satisfies*

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \hat{\nu}) \leq \begin{cases} C(\varepsilon n)^{-\frac{1}{2}} & \text{if } d = 1, \\ C\log(\varepsilon n)(\varepsilon n)^{-\frac{1}{2}} & \text{if } d = 2, \\ C(\varepsilon n)^{-\frac{1}{d}} & \text{if } d \geq 3. \end{cases}$$

4. Private measure mechanism (PMM)

4.1. Binary partition and noisy counts

A binary hierarchical partition of a set Ω of depth r is a family of subsets Ω_θ indexed by $\theta \in \{0,1\}^{\leq r}$, where

$$\{0,1\}^{\leq k} = \{0,1\}^0 \sqcup \{0,1\}^1 \sqcup \cdots \sqcup \{0,1\}^k, \quad k = 0, 1, 2 \ldots,$$

and such that Ω_θ is partitioned into $\Omega_{\theta 0}$ and $\Omega_{\theta 1}$ for every $\theta \in \{0,1\}^{\leq r-1}$. By convention, the cube $\{0,1\}^0$ consists of a single element $\emptyset$. We usually drop the subscript $\emptyset$ and write n instead of $n_\emptyset$. When $\theta \in \{0,1\}^j$, we call j the *level* of θ. We can also encode a binary hierarchical partition of Ω in a binary tree of depth r, where the root is labeled Ω and the j-th level of the tree encodes the subsets Ω_θ for θ at level j.

Let $(\Omega_\theta)_{\theta \in \{0,1\}^{\leq r}}$ be a binary partition of Ω. Given true data $(x_1, \ldots, x_n) \in \Omega^n$, the *true count* n_θ is the number of data points in the region Ω_θ, i.e.

$$n_\theta := \Big| \big\{ i \in [n] : \; x_i \in \Omega_\theta \big\} \Big| .$$

We will convert true counts into *noisy counts* n'_θ by adding Laplacian noise; all regions on the same level will receive noise of the same expected magnitude. Formally, we set

$$n'_\theta := (n_\theta + \lambda_\theta)_+ , \quad \text{where} \quad \lambda_\theta \sim \mathrm{Lap}_{\mathbb{Z}}(\sigma_j),$$

and $j \in \{0, \ldots, r\}$ is the level of θ. At this point, the magnitudes of the noise σ_j can be arbitrary.

4.2. Consistency

The true counts n_θ are non-negative and *consistent*, i.e., the counts of subregions always add up to the count of the region:

$$n_{\theta 0} + n_{\theta 1} = n_\theta \quad \text{for all } \theta \in \{0,1\}^{\leq r-1}.$$

The noisy counts n'_θ are non-negative, but not necessarily consistent. Algorithm 3 enforces consistency by adjusting the counts iteratively, from top to bottom. In the case of the deficit, when the sum of the two subregional counts is smaller than the count of the region, we increase both subregional counts. In the opposite case or surplus, we decrease both subregional counts. Apart from this requirement, we are free to distribute the deficit or surplus between the subregional counts.

It is convenient to state this requirement by considering a *product partial order* on $\mathbb{Z}^2_+$, where we declare that $(a_0, a_1) \preceq (b_0, b_1)$ if and only if $a_0 \leq b_0$ and $a_1 \leq b_1$. We call the two vectors $a, b \in \mathbb{Z}^2$ *comparable* if either $a \preceq b$ or $b \preceq a$. Furthermore, $L(a)$ denotes the line $x + y = a$ on the plane.

Algorithm 3 Consistency

Input: non-negative numbers $(n'_\theta)_{\theta \in \{0,1\}^{\leq r}}$, where n' is a nonnegative integer.
set $m := n'$.
for $j = 0, \ldots, r-1$ **do**
  for $\theta \in \{0,1\}^j$ **do**
    transform the vector $(n'_{\theta 0}, n'_{\theta 1}) \in \mathbb{Z}^2_+$ into any comparable vector $(m_{\theta 0}, m_{\theta 1}) \in \mathbb{Z}^2_+ \cap L(m_\theta)$.
  end
end
Output: non-negative integers $(m_\theta)_{\theta \in \{0,1\}^{\leq r}}$.

At each step, Algorithm 3 uses a transformation $f_\theta : \mathbb{Z}^2_+ \to \mathbb{Z}^2_+ \cap L(m_\theta)$. It can be chosen arbitrarily; the only requirement is that $f_\theta(x)$ be comparable with x. The comparability requirement is natural and non-restrictive. For example, the *uniform* transformation selects the closest point in the discrete interval $\mathbb{Z}^2_+ \cap L(m_\theta)$ in (say) the Euclidean metric. Alternatively, the *proportional* transformation selects the point in the discrete interval $\mathbb{Z}^2_+ \cap L(m_\theta)$ that is closest to the line that connects the input vector and the origin.

4.3. Synthetic data

Algorithm 3 ensures that the output counts m_θ are non-negative, integer, and consistent. They are also private since they are a function of the noisy counts n'_θ, which are private as we proved. Therefore, the counts m_θ can be used to generate *private synthetic data* by putting m_θ points in cell Ω_θ. Algorithm 4 makes this formal.

Algorithm 4 Private Measure Mechanism

Input: true data $\mathcal{X} = (x_1, \ldots, x_n) \in \Omega^n$, noise magnitudes $\sigma_0, \ldots, \sigma_r > 0$.

Compute true counts Let n_θ be the number of data points in Ω_θ.

Add noise Let $n'_\theta := (n_\theta + \lambda_\theta)_+$, where $\lambda_\theta \sim \mathrm{Lap}_{\mathbb{Z}}(\sigma_j)$ are i.i.d. random variables,

Enforce consistency Convert the noisy counts (n'_θ) to consistent counts (m_θ) using Algorithm 3.

Sample Choose any m_θ points in each cell Ω_θ, $\theta \in \{0,1\}^r$ independently of $\mathcal{X}$.

Output: the set of all these points as synthetic data $\mathcal{Y} = (y_1, \ldots, y_m) \in \Omega^m$.

4.4. Privacy and accuracy of Algorithm 4

We first prove that Algorithm 4 is differentially private. The proof idea is similar to the classic Laplacian mechanism. But now our noise is of differential scale for each level, so more delicate calculations are needed.

Theorem 9 (Privacy of Algorithm 4) *The vector of noisy counts $(n_\theta + \lambda_\theta)$ in Algorithm 4 is ε-differentially private, where*

$$\varepsilon = \sum_{j=0}^{r} \frac{1}{\sigma_j}.$$

Consequently, the synthetic data $\mathcal{Y}$ generated by Algorithm 4 is ε-differentially private.

Having analyzed the privacy of the synthetic data, we now turn to its accuracy. It is determined by the magnitudes of the noise σ_j and by the multiscale geometry of the domain Ω. The latter is captured by the diameters of the regions Ω_θ, specifically by their sum at each level, which we denote

$$\Delta_j \coloneqq \sum_{\theta \in \{0,1\}^j} \operatorname{diam}(\Omega_\theta) \tag{4.1}$$

and adopt the notation $\Delta_{-1} \coloneqq \Delta_0 = \operatorname{diam}(\Omega)$. In addition to Δ_j, the accuracy is affected by the *resolution* of the partition, which is the maximum diameter of the cells, denoted by

$$\delta \coloneqq \max_{\theta \in \{0,1\}^r} \operatorname{diam}(\Omega_\theta).$$

Theorem 10 (Accuracy of Algorithm 4) *Algorithm 4 that transforms true data $\mathcal{X}$ into synthetic data $\mathcal{Y}$ has the following expected accuracy in the Wasserstein metric:*

$$\mathbb{E}\, W_1(\mu_\mathcal{X}, \mu_\mathcal{Y}) \le \frac{2\sqrt{2}}{n} \sum_{j=0}^{r} \sigma_j \Delta_{j-1} + \delta.$$

Here $\mu_\mathcal{X}$ and $\mu_\mathcal{Y}$ are the empirical probability distributions on the true and synthetic data, respectively.

The privacy and accuracy guarantees of Algorithm 4 (Theorems 9 and 10) hold for any choice of noise levels σ_j. By optimizing σ_j, we can achieve the best accuracy for a given level of privacy.

Theorem 11 (Optimized Accuracy) *With the optimal choice of magnitude levels (A.4), Algorithm 4 that transforms true data $\mathcal{X}$ into synthetic data $\mathcal{Y}$ is ε-differential private, and has the following expected accuracy in the 1-Wasserstein distance:*

$$\mathbb{E}\, W_1(\mu_\mathcal{X}, \mu_\mathcal{Y}) \le \frac{\sqrt{2}}{\varepsilon n} \Big(\sum_{j=0}^{r} \sqrt{\Delta_{j-1}} \Big)^2 + \delta.$$

Here $\mu_\mathcal{X}$ and $\mu_\mathcal{Y}$ are the empirical measures of the true and synthetic data, respectively.

Corollary 12 (Optimized accuracy for hypercubes) *When $\Omega = [0,1]^d$ equipped with the ℓ^∞ metric, with the optimal choice of magnitude levels (A.4) and the optimal choice of*

$$r = \begin{cases} \log_2(\varepsilon n) - 1 & \text{if } d = 1, \\ \log_2(\varepsilon n) & \text{if } d \geq 2, \end{cases}$$

we have

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \lesssim \begin{cases} \dfrac{\log^2(\varepsilon n)}{\varepsilon n}, & \text{if } d = 1, \\ (\varepsilon n)^{-1/d}, & \text{if } d \geq 2. \end{cases}$$

Remark 13 (Computational efficiency of Algorithm 4) *Since a binary hierarchical partition has 2^r cells in total, the running time of Algorithm 4 is $O(2^r)$. When $\Omega = [0,1]^d$, with the same optimal choice of r in Corollary 12, the running time of PMM becomes $O(\varepsilon d n)$.*

4.5. Proof of Theorem 10

For the proof of Theorem 10, we introduce a quantitative notion for the incomparability of two vectors on the plane. For vectors $a, b \in \mathbb{Z}_+^2$, we define

$$\operatorname{flux}(a,b) := \begin{cases} 0 & \text{if } a \text{ and } b \text{ are comparable,} \\ \min\left(|a_1 - b_1|, |a_2 - b_2|\right) & \text{otherwise.} \end{cases}$$

Lemma 14 (Flux as incomparability) *$\operatorname{flux}(a,b)$ is the ℓ_∞-distance from a to the set of points that are comparable to b.*

For example, if $a = (1,9)$ and $b = (6,7)$, then $\operatorname{flux}(a,b) = 2$. Note that a has a distance 2 to the vector $(1,7)$ which is comparable with b.

Lemma 15 (Flux as transfer) *Suppose we have two bins with a_1 and a_2 balls in them. Then one can achieve b_1 and b_2 balls in these bins by:*

(a) first making the total number of balls correct by adding a total of $(b_1 + b_2) - (a_1 - a_2)$ balls to the two bins (or removing, if that number is negative);

(b) then transferring $\operatorname{flux}\big((a_1,a_2),\,(b_1,b_2)\big)$ balls from one bin to the other.

For example, suppose that one bin has 1 ball and the other has 9. Then we can achieve 6 and 7 balls in these bins by first adding 3 balls to the first bin and transferring 2 balls from the second to the first bin. As we noted above, 2 is the flux between the vectors $(1,9)$ and $b = (6,7)$.

Lemma 15 can be generalized to the hierarchical binary partition of Ω as follows.

Lemma 16 *Consider any data set $\mathcal{X} \in \Omega^n$, and let $(n_\theta)_{\theta \in \{0,1\}^r}$ be its counts. Consider any consistent vector of non-negative integers $(m_\theta)_{\theta \in \{0,1\}^r}$. Then one can transform $\mathcal{X}$ into a set $\mathcal{Z} \in \Omega^m$ that has counts $(m_\theta)_{\theta \in \{0,1\}^r}$ by:*

(a) first making the total number of points correct by adding a total of $m - n$ points to Ω (or remove, if that number is negative);

(b) *then transferring* flux $\big((n_{\theta 0}, n_{\theta 1}),\ (m_{\theta 0}, m_{\theta 1})\big)$ *points from* $\Omega_{\theta 0}$ *to* $\Omega_{\theta 1}$ *or vice versa, for all* $j = 0, \ldots, r-1$ *and* $\theta \in \{0,1\}^j$.

Combining the concept of the flux and our algorithm, the following two lemmas are useful in the proof of Theorem 10.

Lemma 17 *In Algorithm 4, we have*

$$\text{flux}\big((n_{\theta 0}, n_{\theta 1}),\ (m_{\theta 0}, m_{\theta 1})\big) \le \max\big(|\lambda_{\theta 0}|, |\lambda_{\theta 1}|\big)$$

for all $j = 0, \ldots, r-1$ *and* $\theta \in \{0,1\}^j$.

Lemma 18 *For any finite multisets* $U \subset V$ *such that all elements in* U *are from* Ω*, one has*

$$W_1(\mu_U, \mu_V) \le \frac{|V \setminus U|}{|V|} \cdot \operatorname{diam}(\Omega).$$

Proof *(Proof of Theorem 10)* Owing to Lemma 16 and Lemma 17, the creation of synthetic data from the true data $\mathcal{X} \mapsto \mathcal{Y}$, described by Algorithm 4, can be achieved by the following three steps.

1. Transform the n-point input set $\mathcal{X}$ to an m-point set $\mathcal{X}_1$ by adding or removing $|m-n|$ points.

2. Transform $\mathcal{X}_1$ to $\mathcal{X}_2$ by moving at most $\max\big(|\lambda_{\theta 0}|, |\lambda_{\theta 1}|\big)$ many data points for each $j = 0, 1, \ldots, r-1$ and $\theta \in \{0,1\}^j$ between the two parts of the region Ω_θ.

3. Transforms $\mathcal{X}_2$ to the output data $\mathcal{Y}$ by relocating points within their cells.

We will analyze the accuracy of these steps one at a time.
Analyzing Step 2. The total distance the points are moved at this step is bounded by

$$\sum_{j=0}^{r-1} \sum_{\theta \in \{0,1\}^j} \max\big(|\lambda_{\theta 0}|, |\lambda_{\theta 1}|\big) \operatorname{diam}(\Omega_\theta) =: D. \tag{4.2}$$

Since $|\mathcal{X}_1| = m$, it follows that

$$W_1(\mu_{\mathcal{X}_1}, \mu_{\mathcal{X}_2}) \le \frac{D}{m}. \tag{4.3}$$

Combining Steps 1 and 2. Recall that step 1 transforms the input data $\mathcal{X}$ with $|\mathcal{X}| = n$ into $\mathcal{X}_1$ with $|\mathcal{X}_1| = m = n + \operatorname{sign}(\lambda) \cdot \lfloor |\lambda| \rfloor$ by adding or removing points, depending on the sign of λ.
Case 1: $\lambda \ge 0$. Here $\mathcal{X}_1$ is obtained from $\mathcal{X}$ by adding $\lfloor \lambda \rfloor$ points, so Lemma 18 gives

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{X}_1}) \le \frac{\lambda}{m} \cdot \Delta_0.$$

Combining this with (4.3) by triangle inequality, we conclude that

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{X}_2}) \le \frac{\lambda \Delta_0 + D}{m} \le \frac{\lambda \Delta_0 + D}{n}.$$

Case 2: $\lambda < 0$. Here $\mathcal{X}_1$ is obtained from $\mathcal{X}$ by removing a set $\mathcal{X}_0$ of $n - m = \lfloor |\lambda| \rfloor$ points. Furthermore, by our analysis of step 2, $\mathcal{X}_2$ is obtained from $\mathcal{X}_1$ by moving points the total distance at most D. Therefore, $\mathcal{X}_2 \cup \mathcal{X}_0$ (as a multiset) is obtained from $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_0$ by moving points the total distance at most D, too. (The points in $\mathcal{X}_0$ remain unmoved.) Since $|\mathcal{X}| = n$, it follows that

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{X}_2 \cup \mathcal{X}_0}) \leq \frac{D}{n}.$$

Moreover, Lemma 18 gives

$$W_1(\mu_{\mathcal{X}_2}, \mu_{\mathcal{X}_2 \cup \mathcal{X}_0}) \leq \frac{|\mathcal{X}_0|}{|\mathcal{X}_2 \cup \mathcal{X}_0|} \cdot \operatorname{diam}(\Omega) \leq \frac{|\lambda|\, \Delta_0}{n}.$$

(Here we used that the multiset $\mathcal{X}_2 \cup \mathcal{X}_0$ has the same number of points as $\mathcal{X}$, which is n.) Combining the two bounds by triangle inequality, we obtain

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{X}_2}) \leq \frac{|\lambda|\, \Delta_0 + D}{n}. \tag{4.4}$$

In other words, this bound holds in both cases.

Analyzing Step 3. This step is the easiest to analyze: since $\mathcal{Y}$ is obtained from $\mathcal{X}_2$ by relocating the points are relocated within their cells, and the maximal diameter of the cells is δ, we have $W_1(\mu_{\mathcal{X}_2}, \mu_{\mathcal{Y}}) \leq \delta$. Combining this with (4.4) by triangle inequality, we conclude that

$$W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \leq \frac{|\lambda|\, \Delta_0 + D}{n} + \delta.$$

Taking expectation. Recall the definition of D from (4.2). We get

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \leq \frac{1}{n} \left[\mathbb{E}\left[|\lambda|\right] \Delta_0 + \sum_{j=0}^{r-1} \sum_{\theta \in \{0,1\}^j} \mathbb{E}\left[\max\left(|\lambda_{\theta 0}|, |\lambda_{\theta 1}|\right)\right] \operatorname{diam}(\Omega_\theta) \right] + \delta.$$

Since $\lambda \sim \mathrm{Lap}_{\mathbb{Z}}(\sigma_0)$, by (B.1) we have $\mathbb{E}\left[|\lambda|\right] \leq (\mathbb{E}(\lambda)^2)^{1/2} \leq \sqrt{2}\sigma_0$. Similarly, since $\lambda_{\theta 0}$ and $\lambda_{\theta 1}$ are independent $\mathrm{Lap}_{\mathbb{Z}}(\sigma_{j+1})$ random variables, $\mathbb{E}\left[\max\left(|\lambda_{\theta 0}|, |\lambda_{\theta 1}|\right)\right] \leq 2\sqrt{2}\sigma_{j+1}$. Substituting these estimates and rearranging the terms of the sum will complete the proof. ■

Acknowledgments

R.V. acknowledges support from NSF DMS–1954233, NSF DMS–2027299, U.S. Army 76649–CS, and NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning. Y.Z. is partially supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

The authors thank March Boedihardjo, Girish Kumar, and Thomas Strohmer for helpful discussions.

References

Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

John Abowd, Robert Ashmead, Garfinkel Simson, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, and William Sexton. Census topdown: Differentially private data, incremental schemas, and consistency with public knowledge. *US Census Bureau*, 2019.

John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Simson Garfinkel, Micah Heineck, Christine Heiss, Robert Johns, Daniel Kifer, Philip Leclerc, Ashwin Machanavajjhala, et al. The 2020 census disclosure avoidance system topdown algorithm. *Harvard Data Science Review*, (Special Issue 2), 2022.

Khanh Do Ba, Huy L Nguyen, Huy N Nguyen, and Ronitt Rubinfeld. Sublinear time algorithms for earth mover's distance. *Theory of Computing Systems*, 48:428–442, 2011.

Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 273–282, 2007.

Peter L Bartlett and Shahar Mendelson. Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482, 2002.

Steven M Bellovin, Preetam K Dutta, and Nathan Reitinger. Privacy and synthetic datasets. *Stan. Tech. L. Rev.*, 22:1, 2019.

Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *Journal of the ACM (JACM)*, 60(2):1–25, 2013.

Sergey G Bobkov and Michel Ledoux. A simple fourier analytic proof of the AKT optimal matching theorem. *The Annals of Applied Probability*, 31(6):2567–2584, 2021.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Covariance's loss is privacy's gain: Computationally efficient, private and accurate synthetic data. *Foundations of Computational Mathematics*, pages 1–48, 2022a.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Privacy of synthetic data: A statistical framework. *IEEE Transactions on Information Theory*, 69(1):520–527, 2022b.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Private measures, random walks, and synthetic data. *arXiv preprint arXiv:2204.09167*, 2022c.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Private sampling: a noiseless approach for generating differentially private synthetic data. *SIAM Journal on Mathematics of Data Science*, 4(3):1082–1115, 2022d.

March Boedihardjo, Thomas Strohmer, and Roman Vershynin. Covariance loss, Szemeredi regularity, and differential privacy. *arXiv preprint arXiv:2301.02705*, 2023.

Sébastien Bubeck and Mark Sellke. A universal law of robustness via isoperimetry. *Advances in Neural Information Processing Systems*, 34:28811–28822, 2021.

T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. *Advances in neural information processing systems*, 21, 2008.

Steffen Dereich, Michael Scheutzow, and Reik Schottstedt. Constructive quantization: Approximation by empirical measures. *Annales de l'IHP Probabilités et statistiques*, 49(4):1183–1203, 2013.

Sjoerd Dirksen. Tail bounds via generic chaining. *Electron. J. Probab*, 20(53):1–29, 2015.

Josep Domingo-Ferrer, David Sánchez, and Alberto Blanco-Justicia. The limits of differential privacy (and its misuse in data release and machine learning). *Communications of the ACM*, 64(7): 33–35, 2021.

John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 715–724, 2010.

Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53:650–673, 2015.

Dylan J Foster and Alexander Rakhlin. ℓ^∞ vector contraction for Rademacher complexity. *arXiv preprint arXiv:1911.06468*, 6, 2019.

Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. Adaptive metric dimensionality reduction. *Theoretical Computer Science*, 620:105–118, 2016.

Moritz Hardt, Katrina Ligett, and Frank McSherry. A simple and practical algorithm for differentially private data release. *Advances in neural information processing systems*, 25, 2012.

Mathew E Hauer and Alexis R Santos-Lozada. Differential privacy in the 2020 census will distort covid-19 rates. *Socius*, 7:2378023121994014, 2021.

Michael B Hawes. Implementing differential privacy: Seven lessons from the 2020 United States Census. *Harvard Data Science Review*, 2(2), 2020.

Seidu Inusah and Tomasz Kozubowski. A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136:1090–1102, 03 2006. doi: 10.1016/j.jspi.2004.08.014.

James Jordon, Jinsung Yoon, and Mihaela Van Der Schaar. PATE-GAN: Generating synthetic data with differential privacy guarantees. In *International conference on learning representations*, 2019.

Leonid V Kovalev. Lipschitz clustering in metric spaces. *The Journal of Geometric Analysis*, 32 (7):188, 2022.

Terrance Liu, Giuseppe Vietri, and Steven Z Wu. Iterative methods for private synthetic data: Unifying framework and new methods. *Advances in Neural Information Processing Systems*, 34: 690–702, 2021.

Shahar Mendelson. Empirical processes with a bounded ψ 1 diameter. *Geometric and Functional Analysis*, 20(4):988–1027, 2010.

Laurent Meunier, Blaise J Delattre, Alexandre Araujo, and Alexandre Allauzen. A dynamical system perspective for Lipschitz neural networks. In *International Conference on Machine Learning*, pages 15484–15500. PMLR, 2022.

Shuang Song, Kamalika Chaudhuri, and Anand D Sarwate. Stochastic gradient descent with differentially private updates. In *2013 IEEE global conference on signal and information processing*, pages 245–248. IEEE, 2013.

Dong Su, Jianneng Cao, Ninghui Li, Elisa Bertino, and Hongxia Jin. Differentially private k-means clustering. In *Proceedings of the sixth ACM conference on data and application security and privacy*, pages 26–37, 2016.

Michel Talagrand. Matching random samples in many dimensions. *The Annals of Applied Probability*, pages 846–856, 1992.

Michel Talagrand. *The generic chaining: upper and lower bounds of stochastic processes*. Springer Science & Business Media, 2005.

Justin Thaler, Jonathan Ullman, and Salil Vadhan. Faster algorithms for privately releasing marginals. In *Automata, Languages, and Programming: 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part I 39*, pages 810–821. Springer, 2012.

VM Tikhomirov. ε-entropy and ε-capacity of sets in functional spaces. In *Selected works of AN Kolmogorov*, pages 86–170. Springer, 1993.

Jonathan Ullman and Salil Vadhan. PCPs and the hardness of generating private synthetic data. In *Theory of Cryptography: 8th Theory of Cryptography Conference, TCC 2011, Providence, RI, USA, March 28-30, 2011. Proceedings 8*, pages 400–416. Springer, 2011.

Salil Vadhan. The complexity of differential privacy. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, pages 347–450, 2017.

Vijay V Vazirani. *Approximation algorithms*, volume 1. Springer, 2001.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Giuseppe Vietri, Cedric Archambeau, Sergul Aydore, William Brown, Michael Kearns, Aaron Roth, Ankit Siva, Shuai Tang, and Steven Wu. Private synthetic data for multitask learning and marginal queries. In *Advances in Neural Information Processing Systems*, 2022.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with Lipschitz functions. *J. Mach. Learn. Res.*, 5(Jun):669–695, 2004.

Ziteng Wang, Chi Jin, Kai Fan, Jiaqi Zhang, Junliang Huang, Yiqiao Zhong, and Liwei Wang. Differentially private data releasing for smooth queries. *The Journal of Machine Learning Research*, 17(1):1779–1820, 2016.

Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in wasserstein distance. *Bernoulli*, 25(4 A):2620–2648, 2019.

Appendix A. Additional proofs

A.1. Proof of Proposition 3

Proof We will apply the chaining argument (see, e.g., (Vershynin, 2018, Chapter 8)) to deduce a bound similar to Dudley's inequality.

Step 1: (Finding nets)

Define $\varepsilon_j = 2^{-j}$ for $j \in \mathbb{Z}$ and consider an ε_j-net T_j of $\mathcal{F}$ of size $\mathcal{N}(\mathcal{F}, \varepsilon_j, \|\cdot\|_\infty)$. Then for any $f \in \mathcal{F}$ and any level j, we can find the closest element in the net, denoted $\pi_j(f)$. In other words, there exists $\pi_j(f)$ s.t.

$$\pi_j(f) \in T_j, \quad \|f - \pi_j(f)\|_\infty \leq \varepsilon_j.$$

Let m be a positive integer to be determined later, we have the telescope sum together with triangle inequality

$$\begin{aligned} \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^{n} f(X_i)\lambda_i \right| &\leq \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^{n} \big(f - \pi_m(f)\big)(X_i) \cdot \lambda_i \right| \\ &\quad + \sum_{j=j_0+1}^{m} \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^{n} \big(\pi_j(f) - \pi_{j-1}(f)\big)(X_i) \cdot \lambda_i \right|. \end{aligned}$$

Note that when $j = j_0$ is small enough, Ω can be covered by $\pi_{j_0}(f) \equiv 0$.

Step 2: (Bounding the telescoping sum)

For a fixed $j_0 < j \leq m$, we consider the quantity

$$\mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^{n} \big(\pi_j(f) - \pi_{j-1}(f)\big)(X_i) \cdot \lambda_i \right|.$$

For simplicity we will denote $a_i = a_i(f)$ as the coefficient $\frac{1}{n}\left(\pi_j(f) - \pi_{j-1}(f)\right)(X_i)$. Then we have

$$|a_i| \leq \frac{1}{n}\left\|f - \pi_{j-1}(f)\right\|_\infty + \frac{1}{n}\left\|\pi_j(f) - f\right\|_\infty \leq \frac{1}{n}(\varepsilon_j + \varepsilon_{j-1}) \leq \frac{3\varepsilon_j}{n}.$$

Since $\{\lambda_i\}_{i\in[n]}$ are independent subexponential random variables, we can apply Bernstein's inequality to the sum $\sum_i a_i\lambda_i$. Let $K = 3\varepsilon_j$, we have

$$\begin{aligned}
\mathbb{P}\left\{\left|\sum_{i=1}^n a_i\lambda_i\right| > t\right\} &\leq 2\exp\left[-c\min\left(\frac{t^2}{\|a\|_2^2}, \frac{t}{\|a\|_\infty}\right)\right] \\
&\leq 2\exp\left[-c\min\left(\frac{t^2}{K^2/n}, \frac{t}{K/n}\right)\right] \\
&= 2\exp\left[-cn\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right],
\end{aligned}$$

Then we can use the union bound to control the supreme. Define $N = |T_j|\cdot|T_{j-1}| \leq |T_j|^2$,

$$\begin{aligned}
\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n a_i\lambda_i\right| > t\right\} &\leq 2N\exp\left[-cn\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right]\wedge 1 \\
&= 2\exp\left[\log N - cn\min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right]\wedge 1 \\
&\leq 2\exp\left(\log N - cn\frac{t^2}{K^2}\right)\wedge 1 \\
&\quad + 2\exp\left(\log N - cn\frac{t}{K}\right)\wedge 1
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n a_i\lambda_i\right| &= \int_0^\infty 2\exp\left(\log N - cn\frac{t^2}{K^2}\right)\wedge 1\mathrm{d}t \\
&\quad + \int_0^\infty 2\exp\left(\log N - cn\frac{t}{K}\right)\wedge 1\mathrm{d}t \\
&:= I_2 + I_1.
\end{aligned}$$

We will compute them separately.

$$\begin{aligned}
I_1 &= \int_0^\infty 2\exp\left(\log N - cn\frac{t}{K}\right) \wedge 1\mathrm{d}t \\
&= \frac{K\log N}{cn} + \int_{K\log N/cn}^\infty 2\exp\left(\log N - cn\frac{t}{K}\right) \\
&= \frac{K\log N}{cn} + \int_0^\infty 2\exp\left(-cn\frac{t}{K}\right) \\
&\le CK\frac{\log N}{n}
\end{aligned}$$

$$\begin{aligned}
I_2 &= \int_0^\infty 2\exp\left(\log N - cn\frac{t^2}{K^2}\right) \wedge 1\mathrm{d}t \\
&= \sqrt{\frac{K^2\log N}{cn}} + \int_{\sqrt{K^2\log N/cn}}^\infty 2\exp\left(\log N - cn\frac{t^2}{K^2}\right) \\
&= \sqrt{\frac{K^2\log N}{cn}} + \int_0^\infty 2\exp\left(-cn\frac{t^2}{K^2} - 2\sqrt{cn\log N}\frac{t}{K}\right) \\
&\le \sqrt{\frac{K^2\log N}{cn}} + \frac{K}{\sqrt{cn\log N}} \\
&\le CK\sqrt{\frac{\log N}{n}}.
\end{aligned}$$

Therefore we concluded that for a fixed level j,

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n a_i\lambda_i\right| \le CK\left(\frac{\log N}{n} + \sqrt{\frac{\log N}{n}}\right) \lesssim \varepsilon_j\left(\frac{\log N}{n} + \sqrt{\frac{\log N}{n}}\right)$$

Step 3: (Bounding the last entry)

For the last entry in the telescoping sum, similarly, we denote $a_i := \frac{1}{n}\left(f - \pi_m(f)\right)(X_i)$ and we have $|a_i| \le \varepsilon_m/n$. Then

$$\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n a_i\lambda_i\right| \le \frac{\varepsilon_m}{n}\sum_{i=1}^n |\lambda_i|,$$

and the expectation satisfies

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n a_i\lambda_i\right| \le \frac{\varepsilon_m}{n}\sum_{i=1}^n \mathbb{E}\,|\lambda_i| \lesssim \varepsilon_m.$$

Step 4: (Combining the bound and choosing m) Combining the two integrals together, we deduce that for any $X_1, \dots, X_n \in \Omega$,

$$\mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n f(X_i)\lambda_i \right| \le C \left(\varepsilon_m + \sum_{j=j_0+1}^{m} \varepsilon_j \left(\frac{\log \mathcal{N}(\mathcal{F}, \varepsilon_j, \|\cdot\|_\infty)}{n} + \sqrt{\frac{\log \mathcal{N}(\mathcal{F}, \varepsilon_j, \|\cdot\|_\infty)}{n}} \right) \right).$$

Then for any $\alpha > 0$, we can always choose m such that $2\alpha \le \varepsilon_m < 4\alpha$ and bound the sum above with integral

$$\mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \left| \sum_{i=1}^n f(X_i)\lambda_i \right| \le C \left(2\alpha + \frac{1}{\sqrt{n}} \int_\alpha^\infty \sqrt{\log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty)} \mathrm{d}u + \frac{1}{n} \int_\alpha^\infty \log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \mathrm{d}u \right). \tag{A.1}$$

Taking infimum over α completes the proof of the first inequality.

Now assume $\mathcal{F}$ is the set of all functions f with $\|f\|_{\mathrm{Lip}} \le 1$. From (Gottlieb et al., 2016, Lemma 4.2), we can bound the covering number of $\mathcal{F}$ by the covering number of Ω as follows:

$$\log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \le \log(8/u)\, \mathcal{N}(\Omega, u/2, \rho).$$

As a result, for any $\alpha > 0$,

$$L_n(\mathcal{F}) \le C \left(2\alpha + \frac{1}{\sqrt{n}} \int_\alpha^\infty \sqrt{\log(8/u)\, \mathcal{N}(\Omega, u/2, \rho)} \mathrm{d}u + \frac{1}{n} \int_\alpha^\infty \log(8/u)\, \mathcal{N}(\Omega, u/2, \rho) \mathrm{d}u \right).$$

This completes the proof. ■

A.2. Proof of Corollary 4

Proof For $\Omega = [0,1]^d$ with l_∞-norm, we have $\mathrm{diam}(\Omega) = 1$ and the covering number

$$\mathcal{N}([0,1]^d, u, \|\cdot\|_\infty) \le u^{-d}.$$

Then, as the domain $\Omega = [0,1]^d$ is connected and centered, we can apply the bound for the covering number of $\mathcal{F}$ from (von Luxburg and Bousquet, 2004, Theorem 17):

$$\mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \le \left(2 \lceil 2/u \rceil + 1 \right) 2^{\mathcal{N}\left([0,1]^d, u/2, \|\cdot\|_\infty\right)},$$

$$\Longrightarrow \log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) \lesssim \mathcal{N}(\Omega, u/2, \|\cdot\|_\infty) \lesssim (u/2)^{-d}.$$

Applying the inequality above to (A.1), we get

$$L_n \le C \left(2\alpha + \frac{1}{\sqrt{n}} \int_\alpha^\infty (u/2)^{-d/2} \mathrm{d}u + \frac{1}{n} \int_\alpha^\infty (u/2)^{-d} \mathrm{d}u \right). \tag{A.2}$$

Compute the integral for the case $d = 2$ and $d \geq 3$,

$$L_n(f) \leq \begin{cases} C\left(2\alpha + \frac{2}{\sqrt{n}} \log\frac{2}{\alpha} + \frac{2}{n}\left(\frac{\alpha}{2}\right)^{-1}\right) & \text{if } d = 2. \\ C\left(2\alpha + \frac{2}{\sqrt{n}} \cdot \frac{1}{\frac{d}{2}-1}\left(\frac{\alpha}{2}\right)^{1-\frac{d}{2}} + \frac{2}{n} \cdot \frac{1}{d-1}\left(\frac{\alpha}{2}\right)^{1-d}\right) & \text{if } d \geq 3. \end{cases}$$

Choosing $\alpha = 2n^{-1/d}$ finishes the cases for $d \geq 2$.

When $d = 1$, the Dudley integral in (A.2) is divergent. However, note that $\mathrm{diam}(\mathcal{F}) \leq 2$ and hence $\log \mathcal{N}(\mathcal{F}, u, \|\cdot\|_\infty) = 0$ for $u > 1$. From (A.1), we have

$$\begin{aligned} L_n(\mathcal{F}) &\leq C\left(2\alpha + \frac{1}{\sqrt{n}}\int_\alpha^1 (u/2)^{-1/2}\mathrm{d}u + \frac{1}{n}\int_\alpha^1 (u/2)^{-1}\mathrm{d}u\right) \\ &\leq C\left(2\alpha + \frac{2(\sqrt{2}-\sqrt{\alpha})}{\sqrt{n}} + \frac{2}{n}\log\frac{1}{\alpha}\right). \end{aligned}$$

The optimal choice of α is $\alpha \sim n^{-1/2}$, which gives us the result for $d = 1$. ■

A.3. Proof of Proposition 5

Proof It suffices to prove that the steps from $\mathcal{X}$ to the sign measure ν in Algorithm 1 is ε-differentially private since the remaining steps are only based on ν. Notice that both $\mu_\mathcal{Y}, \nu$ are supported on $Y_1, \dots, Y_m$, we can identify the two discrete measures as m dimensional vectors in the standard simplex, denoted $\overline{\mu_\mathcal{Y}}, \overline{\nu}$, respectively. Consider two data sets $\mathcal{X}_1$ and $\mathcal{X}_2$ differ in one point. Suppose we deduced $\mu_{\mathcal{Y}_1}, \mu_{\mathcal{Y}_2}$ and ν_1, ν_2 through the first four steps of Algorithm 1 from $\mathcal{X}_1, \mathcal{X}_2$, respectively. We know two vectors $\overline{\mu_{\mathcal{Y}_1}}, \overline{\mu_{\mathcal{Y}_2}}$ are different at one coordinate, where the difference is bounded by $1/n$.

Then

$$\begin{aligned} \frac{\mathbb{P}\left\{\nu_1 = \eta\right\}}{\mathbb{P}\left\{\nu_2 = \eta\right\}} &= \prod_{i=1}^m \frac{\mathbb{P}\left\{\lambda_i = n(\eta - \overline{\mu_{\mathcal{Y}_1}})_i\right\}}{\mathbb{P}\left\{\lambda_i = n(\eta - \overline{\mu_{\mathcal{Y}_2}})_i\right\}} = \prod_{i=1}^m \frac{\exp(-\varepsilon n|(\eta - \overline{\mu_{\mathcal{Y}_1}})_i|)}{\exp(-\varepsilon n|(\eta - \overline{\mu_{\mathcal{Y}_2}})_i|)} \\ &\leq \exp\left(\varepsilon n \|\mu_{\mathcal{Y}_2} - \mu_{\mathcal{Y}_1}\|_1\right) \leq e^\varepsilon. \end{aligned}$$

By writing $\mathbb{P}\left\{\nu_i \in S\right\} = \sum_{\eta \in S} \mathbb{P}\left\{\nu_i = \eta\right\}$ for $i = 1, 2$, the inequality above implies Algorithm 1 is ε-differentially private. ■

A.4. Proof of Proposition 6

Proof For two signed measures τ, ν supported on $\mathcal{Y}$, the d_{BL}-distance between τ and ν is

$$d_{\mathrm{BL}}(\tau, \nu) = \sup_{\|f\|_{\mathrm{Lip}} \leq 1}\left|\sum_{i=1}^m f(y_i)\left(\tau(\{y_i\}) - \nu(\{y_i\})\right)\right|.$$

For simplicity, we denote $f_i = f(y_i)$, $\nu_i = \nu(\{y_i\})$ and $\tau_i = \tau(\{y_i\})$. Then we note that for any f with $\|f\|_{\text{Lip}} \le 1$, only $(f_i)_{i\in[m]}$ matters in the definition above. Therefore, suppose ν and τ are fixed, computing the d_{BL}-distance is equivalent to the following linear programming problem:

$$\begin{aligned}
\max \quad & \sum_{i=1}^{m}(\nu_i - \tau_i) f_i & \\
\text{s.t.} \quad & f_i - f_j \le \|y_i - y_j\|_\infty, & \forall i,j \le m, i \ne j,\\
& -f_i + f_j \le \|y_i - y_j\|_\infty, & \forall i,j \le m, i \ne j,\\
& -1 \le f_i \le 1, & \forall i \le m.
\end{aligned}$$

After a change of variable $f'_i = f_i + 1$, we can rewrite it as

$$\begin{aligned}
\max \quad & \sum_{i=1}^{m}(\nu_i - \tau_i) f'_i - \big(\nu(\Omega) - 1\big) & \\
\text{s.t.} \quad & f'_i - f'_j \le \|y_i - y_j\|_\infty, & \forall i,j \le m, i \ne j,\\
& -f'_i + f'_j \le \|y_i - y_j\|_\infty, & \forall i,j \le m, i \ne j,\\
& 0 \le f'_i \le 2, & \forall i \le m.
\end{aligned}$$

Next, we can consider the dual problem of the linear programming problem above. The duality theory in linear programming (Vazirani, 2001, Chapter 12) showed that the original problem and the dual problem have the same optimal solution. Let $u_{ij}, u'_{ij} \ge 0$ be the dual variable for the linear constraints about $f'_i - f'_j$ and $-f'_i + f'_j$, and let $v_i \ge 0$ be the dual variable for the equation $f'_i \le 2$. As the linear programming above is in the standard form, by the duality theory, it is equivalent to

$$\begin{aligned}
\min \quad & \sum_{i\ne j} \|y_i - y_j\|_\infty (u_{ij} + u'_{ij}) + 2\, v_i - \big(\nu(\Omega) - 1\big) & \\
\text{s.t.} \quad & \sum_{j\ne i}(u_{ij} - u'_{ij}) + v_i \ge \nu_i - \tau_i, & \forall i \le m,\\
& u_{ij}, u'_{ij}, v_i \ge 0 & \forall i,j \le m, i \ne j.
\end{aligned}$$

To find the minimizer τ for a given ν, we regard τ_i as variables and add the constraints of τ being a probability measure. Also, we can eliminate the constant $\nu(\Omega) - 1$ in the target function. So we get the linear programming problem:

$$\begin{aligned}
\min \quad & \sum_{i\ne j} \|y_i - y_j\|_\infty (u_{ij} + u'_{ij}) + 2\, v_i & \\
\text{s.t.} \quad & \sum_{j\ne i}(u_{ij} - u'_{ij}) + v_i + \tau_i \ge \nu_i, & \forall i \le m,\\
& \sum_{i=1}^{m} \tau_i = 1, & \\
& u_{ij}, u'_{ij}, v_i, \tau_i \ge 0 & \forall i,j \le m, i \ne j.
\end{aligned}$$

There are $2m^2$ variables in total and $m+1$ linear constraints, and the minimizer $(\tau_i)_{i=1}^m$ is what we want. ■

A.5. Proof of Theorem 7

Proof We transformed the original data measure $\mu_{\mathcal{X}}$ with three steps: $\mu_{\mathcal{X}} \longrightarrow \mu_{\mathcal{Y}} \longrightarrow \nu \longrightarrow \hat{\nu}$.
Step 1: For the first step in the algorithm, we have $W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \leq \max_i \operatorname{diam}(\Omega_i)$. This follows from the definition of 1-Wasserstein distance.
Step 2: In this step, ν is no longer a probability measure, and we consider $d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu)$ instead:

$$\begin{aligned}
\mathbb{E} d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu) &= \mathbb{E} \sup_{\|f\|_{\mathrm{Lip}} \leq 1} \left| \int f \mathrm{d}\mu_{\mathcal{Y}} - \int f \mathrm{d}\nu \right| \\
&= \mathbb{E} \sup_{\|f\|_{\mathrm{Lip}} \leq 1} \left| \sum_{i=1}^{m} f(y_i) \left(\frac{n_i}{n} + \frac{\lambda_i}{n} - \frac{n_i}{n} \right) \right| = \frac{m}{\varepsilon n} \widetilde{L}_m(\mathcal{F}). \qquad \text{(A.3)}
\end{aligned}$$

Step 3: For the last step, we have $d_{\mathrm{BL}}(\nu, \hat{\nu}) \leq d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu)$ because $\hat{\nu}$ is the closest probability measure to ν from Proposition 6. As a result, we have

$$\begin{aligned}
W_1(\mu, \hat{\nu}) = d_{\mathrm{BL}}(\mu, \hat{\nu}) &\leq d_{\mathrm{BL}}(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) + d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu) + d_{\mathrm{BL}}(\nu, \hat{\nu}) \\
&\leq W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) + 2 d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu) \leq \max_i \operatorname{diam}(\Omega_i) + 2 d_{\mathrm{BL}}(\mu_{\mathcal{Y}}, \nu).
\end{aligned}$$

After taking the expectation, we can apply (A.3) to get the desired inequality. ■

A.6. Proof of Corollary 8

Proof Using Theorem 7, we have

$$\mathbb{E}\, W_1(\mu_{\mathcal{X}}, \hat{\nu}) \leq \max_i \operatorname{diam}(\Omega_i) + \frac{2m}{\varepsilon n} \widetilde{L}_m(\mathcal{F}).$$

By assumption we have $\max_i \operatorname{diam}(\Omega_i) \asymp m^{-1/d} \asymp (\varepsilon n)^{-1/d}$. And by 4 we have the bound for the Laplacian complexity

$$\widetilde{L}_m(\mathcal{F}) \leq \begin{cases} C(\varepsilon n)^{-1/2} & \text{if } d = 1, \\ C \log n \cdot (\varepsilon n)^{-1/2} & \text{if } d = 2, \\ C(\varepsilon n)^{-1/d} & \text{if } d \geq 3. \end{cases}$$

When $d \geq 3$, the two terms are comparable. And when $d = 1, 2$, the Laplacian complexity dominates the error. Combining the two inequalities gives the result. ■

A.7. Proof of Theorem 9

Theorem 9 can be obtained by applying the parallel composition lemma (Dwork and Roth, 2014). Here we present a self-contained proof by considering an inhomogeneous version of the classical Laplacian mechanism (Dwork and Roth, 2014).

Lemma 19 (Inhomogeneous Laplace mechanism) *Let $F : \Omega^n \to \mathbb{R}^k$ be any map, $s = (s_i)_{i=1}^k \in \mathbb{R}_+^k$ be a fixed vector, and $\lambda = (\lambda_i)_{i=1}^k$ be a random vector with independent coordinates $\lambda_i \sim \mathrm{Lap}_{\mathbb{Z}}(s_i)$. Then the map $x \mapsto F(x) + \lambda$ is ε-differentially private, where*

$$\varepsilon = \sup_{x,\tilde{x}} \left\| F(x) - F(\tilde{x}) \right\|_{\ell^1(s)}.$$

Here the supremum is over all pairs of input vectors in Ω^n that differ in one coordinate, and $\|z\|_{\ell^1(s)} = \sum_{i=1}^k |z_i| / s_i$.

Proof [Proof of Lemma 19] Suppose $x, \tilde{x} \in \Omega^n$ differs in exactly one coordinate. Consider the density functions of the inputs having the same output $y = F(x) + \lambda = F(\tilde{x}) + \tilde{\lambda} \in \mathbb{Z}^k$. We have

$$\begin{aligned}
\frac{\mathbb{P}\left\{F(x) + \lambda = y\right\}}{\mathbb{P}\left\{F(\tilde{x}) + \tilde{\lambda} = y\right\}} &= \frac{\mathbb{P}\left\{\lambda = y - F(x)\right\}}{\mathbb{P}\left\{\tilde{\lambda} = y - F(\tilde{x})\right\}} \\
&= \frac{\prod_{i=1}^k \exp\left(-\frac{|(y - F(x))_i|}{s_i}\right)}{\prod_{i=1}^k \exp\left(-\frac{|(y - F(\tilde{x}))_i|}{s_i}\right)} \\
&= \exp\left(-\sum_{i=1}^k \frac{1}{s_i}\left(|(y - F(x))_i| - |(y - F(\tilde{x}))_i|\right)\right) \\
&\le \exp\left(\|F(x) - F(\tilde{x})\|_{\ell^1(s)}\right) \\
&\le e^{\varepsilon}
\end{aligned}$$

Therefore, we know $x \mapsto F(x) + \lambda$ is ε-differentially private. ■

Proof [Proof of Theorem 9] Consider the map $F(\mathcal{X}) = (n_\theta)$ that transforms the input data into the vector of counts. Suppose a pair of input data $\mathcal{X}$ and $\tilde{\mathcal{X}}$ differ in one point x_i. Consider the corresponding vectors of counts (n_θ) and $(\tilde{n}_\theta)$. For each level $j = 0, \dots, r$, the vectors of counts differ for a single $\theta \in \{0,1\}^j$, namely for the θ that corresponds to the region Ω_θ containing x_i. Moreover, whenever such a difference occurs, we have $|n_\theta - \tilde{n}_\theta| = 1$. Thus, extending the vector $(\sigma_j)_{j=0}^r$ to $(\sigma_\theta)_{\theta \in \{0,1\}^{\le r}}$ trivially (by converting σ_j to σ_θ for all $\theta \in \{0,1\}^j$), we have

$$\left\| F(\mathcal{X}) - F(\tilde{\mathcal{X}}) \right\|_{\ell^1(\sigma)} = \sum_{j=0}^r \frac{1}{\sigma_j} \sum_{\theta \in \{0,1\}^j} |n_\theta - \tilde{n}_\theta| = \sum_{j=0}^r \frac{1}{\sigma_j} = \varepsilon.$$

Applying Lemma 19, we conclude that the map $\mathcal{X} \mapsto (n_\theta + \lambda_\theta)$ is ε-differentially private. ■

A.8. Proof of Theorem 11

Proof We will use the Lagrange multipliers procedure to find the optimal choices of σ_j. Given the maximal layer r, recall Theorem 9, we should use our privacy budget as

$$\varepsilon = \sum_{j=0}^r \frac{1}{\sigma_j}.$$

Therefore, we aim to minimize the accuracy bound with the specified privacy budget, namely

$$\min \mathbb{E}\, W_1(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \quad \text{s.t. } \varepsilon = \sum_{j=0}^{r} \frac{1}{\sigma_j}.$$

Recall the result in Theorem 10. Here ε, n are given and δ is fixed as long as we determine the maximal level r. So the minimization problem is

$$\min \sum_{j=0}^{r} \sigma_j \Delta_{j-1} \quad \text{s.t. } \varepsilon = \sum_{j=0}^{r} \frac{1}{\sigma_j}.$$

Consider the Lagrangian function

$$f(\sigma_0, \ldots, \sigma_r; t) := \sum_{j=0}^{r} \sigma_j \Delta_{j-1} - t \left(\sum_{j=0}^{r} \frac{1}{\sigma_j} - \varepsilon \right)$$

and the corresponding equation

$$\frac{\partial f}{\partial \sigma_0} = \cdots = \frac{\partial f}{\partial \sigma_r} = \frac{\partial f}{\partial t} = 0.$$

One can easily check that the equations above have a unique solution

$$\sigma_j = \frac{S}{\varepsilon \sqrt{\Delta_{j-1}}} \quad \text{where} \quad S = \sum_{j=0}^{r} \sqrt{\Delta_{i-1}}. \tag{A.4}$$

and it is indeed a minimal point for $f(\sigma_0, \ldots, \sigma_r; t)$.

As a result, if we fix ε and want Algorithm 4 to be ε-differentially private, we should choose the noise magnitudes as (A.4). Substituting these noise magnitudes into the accuracy Theorem 10, we see that the accuracy gets bounded by $\frac{\sqrt{2}}{\varepsilon n} S^2 + \delta$. ■

A.9. Proof of Corollary 12

Proof Let $\Omega = [0, 1]$ with the ℓ^∞ metric. The natural hierarchical binary decomposition of $[0, 1]$ (cut through the middle) makes subintervals of length $\operatorname{diam}(\Omega_\theta) = 2^{-j}$ for $\theta \in \{0, 1\}^j$, so $\Delta_j = 1$ for all j, and the resolution is $\delta = 2^{-r}$. Theorem 11 makes ε-differential private synthetic data with accuracy

$$\mathbb{E}\, W_1 (\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \le \frac{\sqrt{2}(r+1)^2}{\varepsilon n} + 2^{-r}.$$

A nearly optimal choice for r is $r = \log_2(\varepsilon n) - 1$, which yields

$$\mathbb{E}\, W_1 (\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}) \le \frac{(2 + \sqrt{2}) \log_2^2(\varepsilon n)}{\varepsilon n}.$$

The optimal noise magnitudes, per (A.4), are $\sigma_j = \log_2^2(\varepsilon n)/\varepsilon$. In other words, the noise *does not decay* with the level.

Let $\Omega = [0,1]^d$ for $d > 1$. The natural hierarchical binary decomposition of $[0,1]^d$ (cut through the middle along a coordinate hyperplane) makes subintervals of length $\mathrm{diam}(\Omega_\theta) \asymp 2^{-j/d}$ for $\theta \in \{0,1\}^j$, so $\Delta_j = 2^j \cdot 2^{-j/d} = 2^{(1-1/d)j}$ for all j, and the resolution is $\delta = 2^{-r/d}$. Thus,

$$S = \sum_{j=0}^{r} \sqrt{\Delta_{j-1}} \sim 2^{\frac{1}{2}(1-\frac{1}{d})r}.$$

Theorem 11 makes a ε-differential private synthetic data with accuracy

$$\mathbb{E}\, W_1\left(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}\right) \lesssim \frac{2^{(1-\frac{1}{d})r}}{\varepsilon n} + 2^{-r/d}.$$

A nearly optimal choice for the depth of the partition is $r = \log_2(\varepsilon n)$, which yields

$$\mathbb{E}\, W_1\left(\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}\right) \lesssim (\varepsilon n)^{-1/d}.$$

The optimal noise magnitudes, per (A.4), are

$$\sigma_j \sim \varepsilon^{-1} 2^{\frac{1}{2}(1-\frac{1}{d})(r-j)}.$$

Thus, the *noise decays* with the level j, becoming $O(1)$ per region for the smallest regions. ∎

A.10. Proof of Lemma 14

Proof If a, b are comparable, both values are zero. If a, b is not comparable, we can assume $a_1 > b_1$, $a_2 < b_2$ without loss of generality. The set of points that are comparable to b is

$$\{(x_1, x_2) \in \mathbb{Z}_+^2 \mid x_1 \le b_1, x_2 \le b_2\} \cup \{(x_1, x_2) \in \mathbb{Z}_+^2 \mid x_1 \ge b_1, x_2 \ge b_2\}.$$

Note that the distance from a to the first set is $|a_1 - b_1|$ and the distance from a to the second set is $|a_2 - b_2|$. Then $\mathrm{flux}(a, b)$ is the smaller one of the two distances, which is also the distance from a to the union set. ∎

A.11. Proof of Lemma 15

Proof *Case 1: $a = (a_1, a_2)$ and $b = (b_1, b_2)$ are comparable.* If $a \preceq b$, remove $b_1 - a_1$ balls from bin 1 and $b_2 - a_2$ balls from bin 2 to achieve the result. If $b \preceq a$, adding $a_1 - b_1$ balls to bin 1 and $a_2 - b_2$ balls to bin 2 to achieve the result.

Case 2: $a = (a_1, a_2)$ and $b = (b_1, b_2)$ are incomparable. Without loss of generality, we can assume that $a_1 - b_1 \ge 0$, $a_2 - b_2 \le 0$.

Assume first that $a_1 - b_1 \ge b_2 - a_2$. Then $\mathrm{flux}(a, b) = b_2 - a_2 := M$. Then $\Delta =: (a_1 + a_2) - (b_1 + b_2) > 0$. Removing Δ balls from bin 1 and transferring M balls from bin 1 to bin 2 achieves the result. Note that there are enough balls in bin 1 to transfer, since $M + \Delta = a_1 - b_1 \in [0, a_1]$.

Now assume that $a_1 - b_1 \le b_2 - a_2$. Then $\mathrm{flux}(a, b) = a_1 - b_1 := M$. Then $\Delta =: (b_1 + b_2) - (a_1 + a_2) > 0$. Adding Δ balls to bin 2 and transferring M balls from bin 1 to bin 2 achieves the result. ∎

A.12. Proof of Lemma 16

Proof First, we make the total number of points in Ω correct by adding $m-n$ points to Ω (or removing, if that number is negative).

Apply Lemma 15 for the two parts of Ω: bin Ω_0 that contains n_0 points and bin Ω_1 that contains n_1 points. Since Ω already contains the correct total number of points m, we can make the two bins contain the correct number of points, i.e. m_0 and m_1 respectively, by transferring flux $\big((n_0,n_1),\,(m_0,m_1)\big)$ points from one bin to the other.

Apply Lemma 15 for the two parts of Ω_0: bin Ω_{00} that contains n_{00} points and bin Ω_{01} that contains n_{01} points. Since Ω_0 already contains the correct number of points m_0, we can make the two bins contain the correct number of points, i.e. m_{00} and m_{01} respectively, by transferring flux $\big((n_{00},n_{01}),\,(m_{00},m_{01})\big)$ points from one bin to the other.

Similarly, since Ω_1 already has the correct number of points m_1, we can make Ω_{10} and Ω_{11} contain the correct number of points m_{10} and m_{11} by transferring flux $\big((n_{00},n_{01}),\,(m_{00},m_{01})\big)$ points from one bin to the other.

Continuing this way, we can complete the proof. Note that the steps of the iteration procedure we described are interlocked. Each next step determines which subregion the transferred points are selected from, and which subregion they are moved to in the previous step. For example, the original step calls to add (or remove) $m-n$ points to or from Ω, but does not specify how these points are distributed between the two parts Ω_0 and Ω_1. The application of Lemma 15 at the next step determines this. ∎

A.13. Proof of Lemma 17

Proof We will derive this result from Lemma 14. First, let us compute the distance from $a = (n_{\theta 0}, n_{\theta 1})$ to $b' = (n'_{\theta 0}, n'_{\theta 1}) = \big((n_{\theta 0}+\lambda_{\theta 0})_+,\,(n_{\theta 1}+\lambda_{\theta 1})_+\big)$. Since the map $x \mapsto x_+$ is 1-Lipschitz, we have

$$\left\|a-b'\right\|_\infty \le \max\left(\left|\lambda_{\theta 0}\right|, \left|\lambda_{\theta 1}\right|\right).$$

Furthermore, recall that by Algorithm 3, b' is comparable to $b = (m_{\theta 0}, m_{\theta 1})$. An application of Lemma 14 completes the proof. ∎

A.14. Proof of Lemma 18

Proof Finding the 1-Wasserstein distance in the discrete case is equivalent to solving the optimal transformation problem. In fact, we can obtain μ_U from μ_V by moving $\left|V \setminus U\right|$ atoms of μ_V, each having mass $1/|V|$, and distributing their mass uniformly over U. The distance for each movement is bounded by $\mathrm{diam}(\Omega)$. Therefore the 1-Wasserstein distance between μ_U and ν_V is bounded by $\frac{|V\setminus U|}{|V|}\,\mathrm{diam}(\Omega)$. ∎

Appendix B. Discrete Laplacian distribution

Recall that the classical Laplacian distribution $\mathrm{Lap}_{\mathbb{R}}(\sigma)$ is a continuous distribution with density

$$f(x) = \frac{1}{2\sigma}\exp\left(-\left|x\right|/\sigma\right), \quad x \in \mathbb{R}.$$

A random variable $X \sim \mathrm{Lap}_{\mathbb{R}}(\sigma)$ has zero mean and

$$\mathrm{Var}(Z) = 2\sigma^2.$$

To deal with counts, it is more convenient to use the *discrete* Laplacian distribution $\mathrm{Lap}_{\mathbb{Z}}(\sigma)$, see (Inusah and Kozubowski, 2006), which has probability mass function

$$f(z) = \frac{1-p_\sigma}{1+p_\sigma} \exp\left(-|z|/\sigma\right), \quad z \in \mathbb{Z}$$

where $p_\sigma = \exp(-1/\sigma)$. A random variable $Z \sim \mathrm{Lap}_{\mathbb{Z}}(\sigma)$ has zero mean and

$$\mathrm{Var}(Z) = \frac{2p_\sigma}{(1-p_\sigma)^2}.$$

Thus, one can verify that discrete Laplacian has a smaller variance than its continuous counterpart:

$$\mathrm{Var}(Z) < 2\sigma^2, \tag{B.1}$$

but the gap vanishes for large σ:

$$\mathrm{Var}(Z) \to 2\sigma^2 \quad \text{as } \sigma \to \infty.$$