Smooth Lower Bounds for Differentially Private Algorithms via Padding-and-Permuting Fingerprinting Codes

Naty Peter NATY.PETER@UTORONTO.CA
Department of Computer Schience, University of Toronto[*]

Eliad Tsfadia ELIADTSFADIA@GMAIL.COM
Department of Computer Schience, Georgetown University

Jonathan Ullman JULLMAN@CCS.NEU.EDU
Khoury College of Computer Sciences, Northeastern University

Editors: Shipra Agrawal and Aaron Roth

Abstract

Fingerprinting arguments, first introduced by Bun et al. (2014), are the most widely used method for establishing lower bounds on the sample complexity or error of approximately differentially private (DP) algorithms. Still, there are many problems in differential privacy for which we don't know suitable lower bounds, and even for problems that we do, the lower bounds are not smooth, and become vacuous when the error is larger than some threshold.

In this work, we present a new framework and tools to generate smooth lower bounds that establish strong lower bounds on the sample complexity of differentially private algorithms satisfying very weak accuracy. We illustrate the applicability of our method by providing new lower bounds in various settings: (1) A tight lower bound for DP averaging in the low-accuracy regime, which in particular implies a lower bound for the private 1-cluster problem introduced by Nissim et al. (2016). (2) A lower bound on the additive error of DP algorithms for approximate k-means clustering and general (k, z)-clustering, as a function of the multiplicative error, which is tight for a constant multiplication error. (3) A lower bound for estimating the top singular vector of a matrix under DP in low-accuracy regimes, which is a special case of the DP subspace estimation problem studied by Singhal and Steinke (2021a).

Our new tools are based on applying a padding-and-permuting transformation to a fingerprinting code. However, rather than proving our results using a black-box access to an existing fingerprinting code (e.g., Tardos' code Tardos (2008)), we develop a new fingerprinting lemma that is stronger than those of Dwork et al. (2015) and Bun et al. (2017), and prove our lower bounds directly from the lemma. Our lemma, in particular, gives a simpler fingerprinting code construction with optimal rate (up to polylogarithmic factors) that is of independent interest.

Keywords: Differential Privacy, Lower Bounds, Fingerprinting Codes.

1. Introduction

Differentially private (DP) Dwork et al. (2006b) algorithms provide a strong guarantee of privacy to the individuals who contribute their data. Informally, a DP algorithm takes data from many individuals and guarantees that no attacker, regardless of their knowledge or capabilities, can learn much more about any one individual than they would have if that individual's data had never been collected. There is a large body of literature on DP algorithms, and DP algorithms have now been deployed by large technology companies and government organizations.

[*] Work conducted while at Georgetown University.

DP Averaging. As a running example, suppose our input dataset is $x_1, \ldots, x_n \in \mathbb{R}^d$ and our goal is to estimate their average $\frac{1}{n}\sum_i x_i$. Since DP requires us to hide the influence of one data point on the average, we intuitively need to assume some kind of bounds on the data. A common way to bound the data is to assume that it lies in some ball of radius r, so there exists a center $c \in \mathbb{R}^d$ such that $\|x_i - c\|_2 \leq r$. Our goal is then to output a DP average $\hat{x}$ such that, with high probability,

$$\left\| \hat{x} - \frac{1}{n}\sum_i x_i \right\|_2 \leq \lambda r$$

If we assume that the location of the center of the ball is *known*, then the natural DP algorithm is to clip the data to lie in this known ball, and perturb the true average with noise from a Laplacian or Gaussian distribution of suitable variance. One can show that the average will satisfy the error guarantee above if the dataset has at least $n \gtrsim \sqrt{d}/\lambda$ samples.

However, in many applications of DP involving real data, we do not want to assume that the center of the ball is known. For example, algorithms like clustering, covariance estimation, and PCA are often applied to datasets as a preprocessing step to understand the general properties, and we cannot assume the user already knows the location of the data. Thus, we want to assume that the data lies in a ball whose center is *unknown*. For this problem, the FriendlyCore algorithm of Tsfadia et al. (2022) is able to achieve the same error guarantee even when the location of the data is *a priori* unknown, provided $n \gtrsim \sqrt{d}/\lambda$.

Lower Bounds for DP Averaging. The work of Bun et al. (2014) proved that $\Omega(\sqrt{d}/\lambda)$ samples are required for DP averaging, when $\lambda \lesssim 1$. Their work introduced the method of *fingerprinting codes* to differential privacy, and this technique has become the standard approach for proving lower bounds for differentially private algorithms, either by reduction to the averaging problem or by non-black-box use of the fingerprinting technique (see Related Work).

The lower bound of Bun et al. (2014) has a significant drawback that it only applies when $\lambda \lesssim 1$, and the lower bound on the sample complexity is vacuous for $\lambda \geq 1$. This limitation is inherent to the way these lower bounds work, since they construct a hard distribution over the hypercube $\{-1, 1\}^d$, which lies in a *known* ball of radius $r = \sqrt{d}$. So the DP algorithm that outputs $\hat{x} = \mathbf{0}$ satisfies

$$\left\| \hat{x} - \frac{1}{n}\sum_i x_i \right\|_2 \leq \sqrt{d} = r.$$

Thus there is no need for any samples when the error parameter is $\lambda \geq 1$, so the lower bound fully captures the hardness of DP averaging when the location of the data is known.

However, when the location of the ball is unknown, even finding a low-accuracy DP average with $\lambda \geq 1$ is non-trivial. In this work, we develop general tools for generating hard-instances for such types of problems. In particular, for DP-averaging, our results implies that $n = \tilde{\Omega}(\sqrt{d}/\lambda)$ samples are required for all λ, yielding that the above algorithms are essentially optimal.

While this low-accuracy regime may seem like an intellectual curiosity, it turns out that low-accuracy approximations of this sort are quite useful for a variety of DP approximation algorithms, and we show that our technique implies new lower bounds for other widely studied problems—computing a DP k-means clustering with a constant multiplicative approximation, and finding a DP top singular vector—that crucially rely on the fact that our lower bound applies to the low-accuracy regime.

1.1. Our Results

1.1.1. MAIN HARDNESS RESULTS

Our core new technique is a variation of Bun et al. (2014)'s method for creating strong error robust fingerprinting codes. That is, a *padding-and-permuting* transformation applied to a (weak error robust) fingerprinting

codes. In this work, we use padding size that varies as a function of the accuracy guarantee, in contrast to Bun et al. (2014) that use fixed-size padding (which suffices for robustness). Using such instances with large padding allows us to smoothly shrink the radius of the points in the hard instances while preserving the hardness. This technique allows us to give the following general construction of a hard problem in DP. Our results will follow from applying this hardness in a black-box way.

Definition 1 (**b-Marked Column**) *Given a matrix $X = (x_i^j)_{i\in[n],j\in[d]} \in \{-1,1\}^{n\times d}$ and $b \in \{-1,1\}$, we say that a column $j \in [d]$ is b-marked if $x_1^j = x_2^j = \ldots = x_n^j = b$. We denote by $\mathcal{J}_X^b \subseteq [d]$ the set of b-marked columns of X.*

Definition 2 (Strongly Agrees) *We say that a vector $q = (q^1, \ldots, q^d)$ strongly-agrees with a matrix $X \in \{-1,1\}^{n\times d}$, if*

$$\forall b \in \{-1,1\}: \quad \left|\{j \in \mathcal{J}_X^b : q^j = b\}\right| \geq 0.9\left|\mathcal{J}_X^b\right|.$$

(i.e., for both $b \in \{-1,1\}$, q agrees with at least 90% of the b-marked columns of X).

Definition 3 (**(α,β)-Weakly-Accurate Mechanism**) *Let $\alpha, \beta \in (0,1]$. We say that a mechanism $\mathsf{M}\colon \{-1,1\}^{n\times d} \to [-1,1]^d$ is (α,β)-weakly-accurate if for every input $X = (x_1,\ldots,x_n) \in \{-1,1\}^{n\times d}$ with $\left|\mathcal{J}_X^1\right|, \left|\mathcal{J}_X^{-1}\right| \geq \frac{1}{2}(1-\alpha)d$, the probability that $\mathsf{M}(X)$ strongly-agrees with X is at least β.*

Namely, for a small α, the only requirement from an (α,β)-weakly-accurate mechanism is to agree (w.p. β) with most of the 1-marked and (-1)-marked columns, but only when almost half of the input columns are 1-marked, and almost all the other half of the columns are (-1)-marked (otherwise, there is no restriction on the output).

The following theorem captures our general tool for lower bounding DP algorithms.

Theorem 4 *If $\mathsf{M}\colon (\{-1,1\}^d)^n \to [-1,1]^d$ is an (α,β)-weakly-accurate $\left(1, \frac{\beta}{4n}\right)$-DP mechanism, then $n \geq \Omega(\sqrt{\alpha d}/\log^{1.5}(\alpha d/\beta))$.*

So in order to prove a lower bound for a specific task, it suffices to prove that the assumed utility guarantee implies Definition 3.

Theorem 4 can be proven by combining our padding-and-permuting technique with an optimal *fingerprinting code*—such as Tardos' code (Tardos (2008))—in a black-box way. Moreover, in contrast with most recent constructions of DP lower bounds that use only a so-called *fingerprinting lemma*, our techniques seem to require the using of a fingerprinting code. Specifically, although fingerprinting lemmas are simpler and more flexible, they require a stronger notion of accuracy that does not fit our padding-and-permuting construction, whereas fingerprinting codes require only an extremely weak notion of accuracy to obtain hardness. In an effort to unify and simplify the techniques used to prove DP lower bounds, we give an alternative proof that makes use of a *new fingerprinting lemma* that only requires very weak accuracy to obtain hardness (see section 2 for more details).

We also consider an extension of Definition 3 to cases where the mechanism receive a dataset which consists of k clusters and is required to output a point that strongly-agrees with one of the clusters.

Definition 5 (**(k,α,β)-Weakly-Accurate Mechanism**) *Let $\alpha, \beta \in (0,1]$ and $n, k, d \in \mathbb{N}$ such that n is a multiple of k. We say that a mechanism $\mathsf{M}\colon \{-1,1\}^{n\times d} \to [-1,1]^d$ is* (k,α,β)-weakly-accurate *if the following holds: Let $X = (x_1,\ldots,x_n) \in \{-1,1\}^{n\times d}$ be an input such that for every $t \in [k]$ and every $b \in \{-1,1\}$ it holds that $\left|\mathcal{J}_{X_t}^b\right| \geq \frac{1}{2}(1-\alpha)d$ for $X_t = (x_{(t-1)n/k+1}, \ldots, x_{tn/k}) \in (\{-1,1\}^d)^{n/k}$. Then*

$$\Pr[\exists t \in [k] \text{ s.t. } \mathsf{M}(X) \text{ strongly-agrees with } X_t] \geq \beta.$$

Note that $(k = 1, \alpha, \beta)$-weakly-accurate is equivalent to (α, β)-weakly-accurate.

Using k independent padding-and-permuting fingerprinting code instances, we prove the following theorem.

Theorem 6 (Extension of theorem 4) *Let $\alpha, \beta \in (0, 1]$ and $n, k, d \in \mathbb{N}$ such that n is a multiple of k. If $\mathsf{M}\colon (\{-1, 1\}^d)^n \to [-1, 1]^d$ is an (k, α, β)-weakly-accurate $\left(1, \frac{\beta}{4n}\right)$-DP mechanism, then $n \geq \Omega(k\sqrt{\alpha d}/\log^{1.5}(\alpha d/\beta))$.*

We prove Theorems 4 and 6 using a more general framework that we developed in this work that might be useful for other types of problems with more complicated output spaces (e.g., subspace or covariance estimation). We refer to section 2.3 for more details.

1.1.2. APPLICATION: AVERAGING AND 1-CLUSTER

We first formally state our tight lower bound for DP averaging in the low-accuracy regimes. We start by defining a (λ, β)-estimator for averaging.

Definition 7 ((λ, β)-Estimator for Averaging) *A mechanism $\mathsf{M}\colon \mathbb{R}^+ \times (\mathbb{R}^d)^n \to \mathbb{R}^d$ is* (λ, β)-estimator for averaging *if given $\gamma \geq 0$ and $x_1, \ldots, x_n \in \{-1, 1\}^d$ with $\max_{i,j\in[n]}\|x_i - x_j\|_2 \leq \gamma$, it holds that*

$$\Pr\left[\left\|\mathsf{M}(\gamma, x_1, \ldots, x_n) - \frac{1}{n}\sum_{i=1}^{n} x_i\right\|_2 \leq \lambda\gamma\right] \geq \beta.$$

It is well-known how to construct DP $(\lambda, \beta = 0.99)$-estimators using $\tilde{O}(\sqrt{d}/\lambda)$ points (Karwa and Vadhan (2018); Tsfadia et al. (2022); Ashtiani and Liaw (2022); Narayanan et al. (2022)).

Fact 8 (Known upper bounds) *For $n = \tilde{O}(\sqrt{d}/\lambda)$, there exists an $(1, \frac{1}{n^2})$-DP $(\lambda, \beta = 0.99)$-estimator for averaging.*

Using Theorem 4, we prove a matching lower bound (up to low-order terms).

Theorem 9 (Our averaging lower bound) *If $\mathsf{M}\colon \mathbb{R}^+ \times (\mathbb{R}^d)^n \to \mathbb{R}^d$ is an (λ, β)-estimator for averaging for $\lambda \geq 1$ and $\mathsf{M}(\gamma, \cdot)$ is $\left(1, \frac{\beta}{4n}\right)$-DP for every $\gamma \geq 0$, then $n \geq \Omega\left(\frac{\sqrt{d}/\lambda}{\log^{1.5}\left(\frac{d}{\beta\lambda}\right)}\right)$.*

Immediate Application: The 1-Cluster Problem. An interesting application of Theorem 9 is a simple lower bound for the 1-cluster problem that is widely used in DP clustering algorithms.

In the 1-cluster problem, we are given n points from a finite domain $\mathcal{X}^d$ for $\mathcal{X} \subseteq \mathbb{R}$, and a parameter $t \leq n$. The goal is to identify a d-dimensional ball that contains almost t point, such that the size of the ball is not too far from the optimum. Formally,

Definition 10 ($(\lambda,\ \beta,\ t_{low},\ s)$-Estimator for 1-Cluster, Nissim et al. (2016)) *A mechanism $\mathsf{M}\colon (\mathcal{X}^d)^n \times [n] \to \mathbb{R}^+ \times \mathbb{R}^d$ is an $(\lambda, \beta, t_{low}, s)$-*estimator for 1-cluster *if given $\mathcal{S} \in (\mathcal{X}^d)^n$ and $t \in [t_{low}, n]$ as inputs, it outputs $r \geq 0$ and $c \in \mathbb{R}^d$ such that the following holds with probability at least β:*

1. *The ball of radius r around c contains at least $t - s$ points from $\mathcal{S}$, and*
2. *Let r_{opt} be the radius of the smallest ball in $\mathcal{X}^d$ containing at least t input points. Then $r \leq \lambda \cdot r_{opt}$.*

It is well-known how to privately solve this problem with a constant λ, whenever $t_{low} \geq \tilde{\Theta}(\sqrt{d})$.

Fact 11 (Upper bounds Nissim et al. (2016); Nissim and Stemmer (2018), simplified) *There exists an* $(1, \frac{1}{n^2})$*-DP,* $(\lambda = \Theta(1),\ \beta = 0.99,\ t_{low} = \tilde{\Theta}(\sqrt{d}),\ s = \tilde{\Theta}(1))$*-estimator for* 1*-cluster.*

Therefore, we conclude by Theorem 9 the following tight lower bound (up to low order terms) which is essentially an immediate corollary of our averaging lower bound.

Corollary 12 (Our 1-cluster lower bound) *If* M *is* $\left(1, \frac{\beta}{4n}\right)$*-DP and* $(\lambda,\ \beta,\ t_{low},\ s = n-1)$*-Estimator for* 1*-Cluster for* $\lambda \geq 1$*, then* $t_{low} \geq \Omega\left(\frac{\sqrt{d}/\lambda}{\log^{1.5}\left(\frac{d}{\beta\lambda}\right)}\right)$.

We remark that for these specific tasks of DP-averaging/1-cluster, a recent work of Narayanan et al. (2022), which provides a lower bound for user-level DP averaging, can also be used to prove a similar statement to Theorem 9 (up to poly-logarithmic terms). Their technique is very different from ours, and in particular, do not extend to proving Theorems 4 and 6 which serve as simple tools for proving lower bounds for the other problems. We refer to section 2.4 for a more detailed comparison.

1.1.3. Application: Clustering

In k-means clustering, we are given a database S of n points in $\mathbb{R}^d$, and the goal is to output k centers $C = (c_1, \ldots, c_k) \in (\mathbb{R}^d)^k$ that minimize

$$\mathrm{COST}(C; S) := \sum_{x \in \mathcal{S}} \min_{i \in [k]} \|x - c_i\|_2^2.$$

Similarly to prior works, we focus, without loss of generality, on input and output points in the d-dimensional unit ball $\mathcal{B}_d := \{x \in \mathbb{R}^d \colon \|x\|_2 \leq 1\}$. The approximation quality of a DP algorithm is measured in the literature by two parameters: a multiplicative error λ, and an additive error ξ, defined below:

Definition 13 ((λ, ξ, β)-Approximation Algorithm for k-Means) $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^k$ *is an* (λ, ξ, β)-approximation algorithm for k-means*, if for every* $S \in (\mathcal{B}_d)^n$ *it holds that*

$$\Pr_{C \sim \mathsf{M}(S)}[\mathrm{COST}(C; S) \leq \lambda \cdot \mathrm{OPT}_k(S) + \xi] \geq \beta,$$

where $\mathrm{OPT}_k(S) := \min_{C \in (\mathcal{B}_d)^k} \mathrm{COST}(C; S)$.

While non-private algorithms usually do not have an additive error, under DP an additive error is necessary. As far as we are aware, the only known lower bound on the additive error is the one of Gupta et al. (2010) (stated for k-medians), which has been extended later by Nguyen et al. (2021) (Theorem 1.2). This lower bound essentially says that the additive error ξ of any $(1, \frac{1}{n^2})$-DP algorithm for k-means (regardless of its multiplicative error) must be at least $\tilde{\Omega}(k)$. However, as far as we are aware, all known DP upper bounds have an additive error of at least $\Omega(k\sqrt{d})$. In particular, this is also the situation in the state-of-the-art upper bounds that have constant multiplicative error (Kaplan and Stemmer (2018); Ghazi et al. (2020); Nguyen et al. (2021)).

Fact 14 (General upper bounds, simplified) *There exists an* $(1, \frac{1}{n^2})$*-DP* $(\Theta(1),\ \tilde{\Theta}(k\sqrt{d}))$*-approximation algorithm for* k*-means.*

Furthermore, an additive error of $\tilde{\Theta}(k\sqrt{d})$ also appears in algorithms that provide utility only for datasets that are well-separated into k-clusters: For a small parameter $\phi \in [0, 1]$, a dataset $S \in (\mathcal{B}^d)^n$ is called ϕ-separated for k-means if $\mathrm{OPT}_k(S) \leq \phi^2 \cdot \mathrm{OPT}_{k-1}(S)$ (Ostrovsky et al. (2012)).

Fact 15 (Upper bounds for well-separated instances Shechner et al. (2020); Cohen et al. (2021), simplified) *There exists an* $(1, \frac{1}{n^2})$*-DP* $(1 + O(\phi^2),\ \tilde{\Theta}(k\sqrt{d}))$*-approximation algorithm for* k*-means of* ϕ*-separated instances.*

Using Theorem 6 we provide the first tight lower bound on the additive error for algorithms with constant multiplication error. Furthermore, since Theorem 6 is proven using k independent padding-and-permuting FPC instances which induce k obvious clusters that are far from each other, our lower bound also matches the upper bounds for well-separated instances.

Theorem 16 (Our k-means lower bound) *Let* $n, k, d \in \mathbb{N}$, $\lambda \geq 1$, $\beta \in (0, 1]$ *and* $\xi \geq 0$ *such that* $n \geq k + 80\xi$. *If* $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^k$ *is an* $\left(1, \frac{\beta}{4nk}\right)$*-DP* (λ, ξ, β)*-approximation algorithm for* k*-means, then either* $k \geq 2^{\Omega(d/\lambda)}\beta\lambda/d$ *or* $\xi \geq \Omega\left(\frac{k\sqrt{d/\lambda}}{\log^{1.5}\left(\frac{kd}{\beta\lambda}\right)}\right)$.

Note that Theorem 16 even suggests how we might expect the additive error to decrease if we increase the multiplicative error (for such cases, however, there are no matching upper bounds).

We remark that our proof is not tailored to k-means clustering. In appendix E.2 we state and prove an extension of Theorem 16 to (k, z)*-clustering* in which the cost is measured by the sum of the z^{th} powers of the distances (k-means clustering is the special case of $z = 2$).

1.1.4. APPLICATION: TOP SINGULAR VECTOR

In this problem, we are given n points $x_1, \ldots, x_n \in \mathcal{S}_d := \{v \in \mathbb{R}^d \colon \|v\|_2 = 1\}$ of unit norm as input, and the goal is to estimate the top (right) singular vector of the $n \times d$ matrix $X = (x_i^j)_{i \in [n], j \in [d]}$, which is the unit vector $v \in \mathcal{S}_d$ that maximizes $\|X \cdot v\|_2$.

The singular value decomposition of X is defined by $X = U\Sigma V^T$, where $U \in \mathbb{R}^{n \times n}$ and $V \in \mathbb{R}^{d \times d}$ are unitary matrices. The matrix $\Sigma \in \mathbb{R}^{n \times d}$ is a diagonal matrix with non-negative entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min\{n,d\}} \geq 0$ along the diagonal, called the singular values of X. The first column of V is the top right singular vector.

Dwork et al. (2014) proved a general lower bound of $n = \tilde{\Omega}(\sqrt{d})$ for any algorithm that identifies a useful approximation to the top singular vector. Yet, Singhal and Steinke (2021b) bypassed this lower bound under a distributional assumption which implies that the points are close to lying in a 1-dimensional subspace, defined by having a small ratio σ_2/σ_1. They showed that when the points are Gaussian, and the above ratio is small, the sample complexity can be made independent of d. They also consider the more general problem of estimating the span of the top k singular vectors, which we do not consider in this work.

Definition 17 ((λ, β)-Estimator of Top Singular Vector) *We say that* $\mathsf{M} \colon [0, 1] \times (\mathcal{S}_d)^n \to \mathcal{S}_d$ *is an* (λ, β)*-estimator of top singular vector, if given an* $n \times d$ *matrix* $X = (x_1, \ldots, x_n) \in (\mathcal{S}_d)^n$ *and an upper bound* $\gamma \in [0, 1]$ *on* σ_2/σ_1 *as inputs, outputs a column vector* $y \in \mathcal{S}_d$ *such that*

$$\Pr_{y \sim \mathsf{M}(\gamma, X)}\left[\|X \cdot y\|_2^2 \geq \|X \cdot v\|_2^2 - \lambda\gamma n\right] \geq \beta,$$

where v *denotes the top singular vector of* X.

Our next result is a lower bound on the sample complexity of singular vector estimation that is smooth with respect to the spectral gap parameter γ.

Theorem 18 (Our lower bound) *If* $\mathsf{M} \colon [0, 1] \times (\mathcal{S}_d)^n \to \mathcal{S}_d$ *is an* (λ, β)-estimator of top singular vector *for* $\lambda \geq 1$ *and* $\mathsf{M}(\gamma, \cdot)$ *is* $\left(1, \frac{\beta}{4n}\right)$*-DP for every* $\gamma \in [0, 1]$, *then* $n \geq \Omega\left(\frac{\sqrt{d}/\lambda}{\log^{1.5}\left(\frac{d}{\beta\lambda}\right)}\right)$.

For comparison, Dwork et al. (2014) proved a lower bound of $\tilde{\Omega}(\sqrt{d})$ on the sample complexity but it only applies when γ is larger than some specific constant, whereas our lower bound holds for the entire range of γ.

1.2. Related Work

The connection between fingerprinting codes and differential privacy was first introduced by Bun et al. (2014). Subsequent work significantly simplified and generalized the method in a number of ways Steinke and Ullman (2016); Dwork et al. (2015); Bun et al. (2017); Steinke and Ullman (2017); Kamath et al. (2019); Cai et al. (2021); Kamath et al. (2022); Cai et al. (2021); Narayanan et al. (2022), including the removal of fingerprinting codes and distilling the main technical component into a *fingerprinting lemma*. The price of this simplicity and generality is that the lower bounds rely on having a stronger accuracy requirement that constrains both marked and unmarked columns, whereas the hard instances constructed directly from fingerprinting codes require only a very weak accuracy requirement that constrains only marked columns.

Fingerprinting lower bounds have applied in many settings, including mean estimation Bun et al. (2014); Dwork et al. (2015); Kamath et al. (2019); Narayanan et al. (2022), adaptive data analysis Hardt and Ullman (2014); Steinke and Ullman (2015), empirical risk minimization Bassily et al. (2014), spectral estimation Dwork et al. (2014), combining public and private data Bassily et al. (2020), regression Cai et al. (2021, 2023), sampling Raskhodnikova et al. (2021), Gaussian covariance estimation Kamath et al. (2022); Narayanan (2024), continual observation Jain et al. (2023), and unbiased private estimation Kamath et al. (2023). Of these works, all of them except for Raskhodnikova et al. (2021) use fingerprinting lemmas, rather than the stronger construction of fingerprinting codes. Except Narayanan et al. (2022) (discussed in section 2.4), all these lower bounds are based on the same type of hard instances that lead to the data being contained in a known ball of a given radius, and thus none of them have the smoothness property we desire. Thus, we believe that our method could find other applications beyond the ones that are described in this paper.

1.3. Paper Organization

In section 2 we present a proof overview of Theorem 4 for the case $\beta \approx 1$ that uses an optimal fingerprinting code as black-box, explain how we give a direct proof using a strong fingerprinting lemma that we developed in this work, and describe additional properties of our results. We prove our strong fingerprinting lemma in section 3.

Notations, definitions and general statements used throughout the paper are given in appendix A. A robust version of our fingerprinting lemma is stated and proved in appendix B. In appendix C we present our general framework for proving DP lower bounds that is based on our fingerprinting lemma. In appendix D we present our padding-and-permuting transformation, and prove theorems 4 and 6 using our framework. In appendix E we prove theorems 9, 16 and 18, which give our main applications. In appendix F we show how to construct a simple fingerprinting code using our strong fingerprinting lemma.

2. Our Technique

In this section, we present a proof overview of Theorem 4 for $\beta \approx 1$. In section 2.1 we present a simple variant of the proof that uses an optimal fingerprinting code as black-box (e.g., Tardos (2008)). In section 2.2 we explain how we actually avoid the use of Tardos' fingerprinting code by developing a strong fingerprinting lemma. In section 2.3 we describe our more general framework, and in section 2.4 we make a detailed comparison with Narayanan et al. (2022).

2.1. Proof via an Optimal Fingerprinting Code

Fingerprinting Code (FPC). An FPC consists of two algorithms: Gen and Trace. Algorithm Gen on input n outputs a codebook (matrix) $(x_i^j)_{i\in[n],j\in[d]} \in \{-1,1\}^{n\times d}$ for $d = d(n)$, and a secret state st. An adversary who controls a coalition $\mathcal{S} \subseteq [n]$ only gets the rows $(x_i)_{i\in\mathcal{S}}$ and is required to output $q = (q^1,\ldots,q^d) \in \{-1,1\}^d$ that agrees with the "marked" columns, i.e., columns $j \in [d]$ where $x_i^j = b$ (for the same $b \in \{-1,1\}$) for every $i \in \mathcal{S}$. On "unmarked" columns $j \in [d]$, there is no restriction and the adversary is allowed to choose q^j arbitrarily. Algorithm Trace, given such legal q (and the secret state st), guarantees to output $i \in \mathcal{S}$ with high probability (i.e., to reveal at least one of the coalition members).

Fingerprinting codes were originally introduced by Boneh and Shaw (1998). Tardos (2008) constructed an optimal FPC of length $d_0 = \tilde{\Theta}(n^2)$, and Bun et al. (2014) proved that Tardos' code is actually *robust*, i.e., it enables tracing even when the adversary is allowed to be inconsistent with a small fraction of marked columns (say, 20%).[1]

Padding-and-Permuting FPC. Given a robust FPC $(\mathsf{Gen},\mathsf{Trace})$ of length $d_0 = d_0(n)$, consider the padding-and-permuting (PAP) variant of it as the following pair of algorithms $(\mathsf{Gen}',\mathsf{Trace}')$:

Algorithm Gen'**:**

- Input parameters: Number of users $n \in \mathbb{N}$ and accuracy parameter $\alpha \in [0,1]$. Let $d_0 = d_0(n)$ be the codewords' length of $\mathsf{Gen}(n)$ and let $d = d_0 + 2\ell$ for $\ell = \left\lceil \frac{d_0}{2\alpha} \right\rceil$.
- Operation:
  1. Sample a codebook $X \in \{-1,1\}^{n\times d_0}$ along with a secret state st according to $\mathsf{Gen}(n)$.
  2. Append ℓ 1-marked and ℓ (-1)-marked columns to the matrix X.
  3. Permute the columns of X according to a random permutation π over $[d]$.
  4. Output the resulting matrix $X' \in \{-1,1\}^{n\times d}$ along with the new state $\mathsf{st}' = (\pi,\mathsf{st})$.

Algorithm Trace'**:**

- Input parameters: A weakly-accurate result $q = (q^1,\ldots,q^d) \in \{-1,1\}^d$ and a secret state $\mathsf{st}' = (\pi,\mathsf{st})$.
- Operation:
  - Output $\mathsf{Trace}(\tilde{q},\mathsf{st})$ for $\tilde{q} = (q^{\pi(1)},\ldots,q^{\pi(d_0)}) \in \{-1,1\}^{d_0}$.

Note that we set the padding length as a function of the accuracy parameter α such that a weaker accuracy (i.e., smaller α) results with a larger padding. This is crucial for creating hard instances in the regime where an α-weakly-accurate mechanism must be accurate.

Proving Theorem 4 Let $(\mathsf{Gen},\mathsf{Trace})$ be a robust FPC with codewords' length $d_0 = \tilde{\Theta}(n^2)$ (e.g., Tardos (2008)). Suppose that we sample $X' = (x'_1,\ldots,x'_n) \in \{-1,1\}^{n\times d}$ according to $\mathsf{Gen}'(n,\alpha)$. By construction, X' contains at least $\frac{1}{2}(1-\alpha)d$ b-marked columns, for both $b \in \{-1,1\}$. Therefore, if $\mathsf{M}\colon \{-1,1\}^{n\times d} \to \{-1,1\}^d$ is α-weakly-accurate, then the output $q \in \{-1,1\}^d$ of $\mathsf{M}(X')$ must agree with 90% of the marked columns of X'. But because the columns are randomly permuted, M cannot distinguish between marked columns from the padding and marked columns from the original codebook X. This means that it must

1. Bun et al. (2014) did not try to optimize the constant in the fraction of errors, and only proved it for 4%. For the purpose of this proof sketch, we assume that the code is robust for 20% errors. A formal proof that relies on their result must change the constant 0.9 in Definition 2 to a constant larger than 0.96.

agree with a similar fraction of marked columns of the codebook X, which enables tracing since the code is robust. Therefore, we conclude that such a weakly-accurate mechanism cannot be DP unless $n \geq \tilde{\Omega}(\sqrt{d_0})$, i.e., $n \geq \tilde{\Omega}(\sqrt{\alpha d})$.

We remark that handling smaller values of β (the success probability of M) creates more technical challenges that we ignore for the purpose of this overview.

2.2. Proof via a Strong Fingerprinting Lemma

The disadvantage of using FPC as black-box for DP lower bounds is the fact that Tardos (2008)'s analysis is quite involved, so it is hard to gain an end-to-end understanding of the process, and in particular, to generate hard instances in more complicated settings (e.g., exponential families Kamath et al. (2022)). Therefore, later results simplified the construction and especially the analysis, at the cost of considering more restricted adversaries that must estimate the average of most coordinates, and not just the "marked" ones (which usually suffices for the aggregation tasks we are interested in). This led to the development of the Fingerprinting Lemma (described below) which serves as the most common tool for proving lower bounds for approximate DP algorithms.

Lemma 19 (Original Fingerprinting Lemma Bun et al. (2017); Dwork et al. (2015)) *Let $f\colon \{-1,1\}^n \to [-1,1]$ be a function such that for every $x = (x_1,\ldots,x_n) \in \{-1,1\}^n$, satisfies $\left|f(x) - \frac{1}{n}\sum_{i=1}^n x_i\right| \leq 1/3$. Then,*

$$\mathrm{E}_{p\leftarrow[-1,1],\, x_{1\ldots n}\sim p}\left[f(x)\cdot\sum_{i=1}^n (x_i - p)\right] \geq \Omega(1),$$

where $p \leftarrow [-1,1]$ denotes that p is sampled uniformly over $[-1,1]$, and $x_{1\ldots n} \sim p$ denotes that each $x_i \in \{-1,1\}$ is sampled independently with $\mathrm{E}[x_i] = p$.

Roughly, Lemma 19 says that if $f(x) \approx \frac{1}{n}\sum_{i=1}^n x_i$, then $f(x)$ has $\Omega(1/n)$ correlation (on average) with each x_i. In order to increase the correlation, we increase the dimension of the x_i's by using $d = \tilde{\Theta}(n^2)$ independent copies for each coordinate (column). This guarantees that the average correlation that a single word $x_i \in \{-1,1\}^d$ has with an accurate output $q \in [-1,1]^d$ is $\tilde{\Omega}(d/n)$, which is sufficiently larger than the $\tilde{O}(\sqrt{d})$ correlation that an independent row (which was not part of the input) has with q.

However, in our case, since Lemma 19 is only restricted to adversaries that must be accurate for all types of columns, and not just marked ones, we could not use it for our padding-and-permuting (PAP) technique, and therefore we developed the following stronger Fingerprinting Lemma.

Lemma 20 (Our Strong Fingerprinting Lemma) *Let $f\colon \{-1,1\}^n \to [-1,1]$ with $f(1,\ldots,1) = 1$ and $f(-1,\ldots,-1) = -1$, and let ρ be the distribution that outputs $p = \frac{e^t-1}{e^t+1}$ for $t \leftarrow [-\ln(5n), \ln(5n)]$. Then,*

$$\mathrm{E}_{p\sim\rho,\, x_{1\ldots n}\sim p}\left[f(x)\cdot\sum_{i=1}^n (x_i - p)\right] \geq \Omega(1/\log n).$$

Note that up to the $\log n$ factor, Lemma 20 is much stronger than the original fingerprinting lemma that is used in the literature (Lemma 19), since it only requires f to be fixed on the two points, $(1,\ldots,1)$ and $(-1,\ldots,-1)$, and nothing else (i.e., f can be completely arbitrary on any other input).

Now the same approach of Bun et al. (2017); Dwork et al. (2015) yields that we can increase the correlation by taking $d = \tilde{\Theta}(n^2)$ independent columns,[2] which leads to a tracing algorithm. But unlike Bun

2. An additional factor of $\log^2 n$ is hidden inside the $\tilde{\Theta}$ due to the $\log n$ factor that we lose in our new fingerprinting lemma. However, we ignore it for the sake of this presentation as it is only a low-order term.

et al. (2017); Dwork et al. (2015), we obtain a much stronger "FPC-style" tracing algorithm, which enables to apply a similar PAP approach as described in section 2.1. As a corollary that is of independent interest, the above approach results in a new fingerprinting code that has a simpler analysis than the one of Tardos (2008), described in appendix F.

An additional advantage of Lemma 20 is that it also extends to randomized functions f with $\Pr[f(1,\ldots,1)=1], \Pr[f(-1,\ldots,-1)=-1] \geq 0.9$. This yields an FPC against mechanisms M that given a codebook X, the output $\mathsf{M}(X)$ is *strongly-correlated* with X:

Definition 21 (Strongly Correlated) *We say that a random variable $Q = (Q^1,\ldots,Q^d) \in \{-1,1\}^d$ is* strongly-correlated *with a matrix $X \in \{-1,1\}^{n\times d}$, if*

$$\forall b \in \{-1,1\},\ \forall j \in \mathcal{J}_X^b : \quad \Pr\left[Q^j = b\right] \geq 0.9.$$

Note that *strongly-correlated* (Definition 21) is similar to *strongly-agrees* (Definition 2) but not exactly the same (in particular, the former is a property of a random variable while the latter is a property of a fixed vector). In our analysis, we use the *strongly-correlated* definition which makes the statements and proofs much more clean. To see this, recall that in the proof sketch in section 2.1, we claimed that an algorithm that agrees with 90% of the marked columns of X', also agrees with a "similar" fraction of marked columns of the original codebook X. But this is not exactly the same agreement fraction, and it creates some additional restrictions on the parameters, and in particular, one has to go into the specific FPC construction to argue that w.h.p., there are many marked columns (as done by Bun et al. (2014) w.r.t. Tardos' code). But in this work we observed that padding-and-permuting simply transforms an 0.9-agreement on X' (i.e., *strong-agreement*) into 0.9-correlation on X (i.e., *strong-correlation*) without any special requirements on the parameters (this is the content of Lemma 43), so we were able use our FPC directly.

2.3. More General Framework

While our tools (Theorems 4 and 6) capture various fundamental problems, they may not serve as the right abstraction for handling more complicated output spaces (e.g., matrices) that are used for other problems (e.g., subspace or covariance estimation). We therefore provide a more general framework that we hope can be applied to other types of algorithms without the need to develop similar tools from scratch.

Consider a mechanism $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{W}$ that satisfies some weak accuracy guarantee. In order to prove a lower bound on n using our approach, we need somehow to transform an FPC codebook $X \in \{-1,1\}^{n_0\times d_0}$ into hard instances $Y \in \mathcal{X}^n$ for M, and then extract from the output $w \in \mathcal{W}$ of $\mathsf{M}(Y)$ a vector $q \in \{-1,1\}^{d_0}$ that is strongly-correlated with X (n_0 and d_0 are some functions of n and d and the weak accuracy guarantee of M). Denote by $\mathsf{G}\colon \{-1,1\}^{n_0\times d_0} \times \mathcal{V} \to \mathcal{X}^n$ the algorithm that generates the hard instances using a uniformly random secret $v \leftarrow \mathcal{V}$ (i.e., v could be a random permutation, a sequence of random permutations, etc). Denote by $\mathsf{F}\colon \mathcal{V}\times\mathcal{W} \to \{-1,1\}^{d_0}$ the algorithm that extracts a good q using the secret v and the output w. Denote by $\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}(X)$ the process that samples $v \leftarrow \mathcal{V}$, and outputs $q \sim \mathsf{F}(v,\mathsf{M}(\mathsf{G}(X,v)))$.

Our framework (Lemma 40) roughly states that if M is $\left(1,\frac{\beta}{4n_0}\right)$-DP and there exists such G,F where: (1) The output of $\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}(X)$ is strongly-correlated with X w.p. at least β over the random coins of $\mathsf{M},\mathsf{F},\mathsf{G}$, and (2) G is neighboring-preserving (i.e., maps neighboring datasets to neighboring datasets), then $n_0 \geq \Omega\left(\frac{\sqrt{d_0}}{\log^{1.5}(d_0/\beta)}\right)$.

We prove Theorem 4 by applying the framework with $n_0 = n$ and $d_0 = \alpha d$, and we prove Theorem 6 by applying it with $n_0 = n/k$ and $d_0 = \alpha d$.

2.4. Comparison with Narayanan et al. (2022)

Narayanan et al. (2022) developed very different hard-instances for lower bounding *user-level* DP averaging that can be used to prove a similar statement to Theorem 9 (our DP averaging lower bound). In their setting, a

distribution vector $p = (p_1, \ldots, p_d) \in [0.5/d, 1.5/d]^d$, $\sum_{i=1}^{d} p_i = 1$, is sampled, and the goal is to estimate it. For each user (out of m), they sample $m = \lambda^2 d$ one-hot vectors according to p, and provide the average of the vectors as the user's input point. Since each $p_i \approx 1/d$, it can be shown that w.h.p., the resulting points have ℓ_2 diameter $\gamma \approx 1/\sqrt{m} = 1/\lambda\sqrt{d}$ (and also close to p up to such an additive error). So their Lemma 23 essentially states that estimating p up-to additive ℓ_2 error of $1/\sqrt{d} = \lambda \cdot \gamma$, requires $\tilde{\Omega}(\sqrt{d}/\lambda)$ users (which matches our Theorem 9 up to the low-order terms).

While both works yield similar smooth lower bound for DP averaging, the focus of the works is different: Narayanan et al. (2022) focus on *user-level* DP averaging, while our focus is on general bounds for various (*item-level*) DP problems. As part of that, we have several advantages in the *item-level* DP case:

1. Applicability: Our padding-and-permuting FPC hard-instances enable to prove more general tools (Theorems 4 and 6) which provide clean abstraction for lower bounding other fundamental problems under DP, like clustering and estimating the top-singular vector.

2. Simplicity: Our proof is conceptually cleaner and simpler, and in particular, can be explained using a simple black-box construction from an optimal fingerprinting-code (section 2.1).

3. Tighter Bounds: Narayanan et al. (2022) have an $1/\log^7(d/\lambda)$ dependency in their DP-averaging lower bound (see their Theorem 10), while we only have an $1/\log^{1.5}(d/\lambda)$ dependency.

4. Boolean instances: We create *boolean* hard-instances, while Narayanan et al. (2022) instances are not.

3. Strong Fingerprinting Lemma

In this section, we prove Lemma 20. We make use of the following lemma.

Lemma 22 (Dwork et al. (2015), Lemma 5) *Let $f\colon \{-1,1\}^n \to \mathbb{R}$. Define $g\colon [-1,1] \to \mathbb{R}$ by*

$$g(p) := \mathrm{E}_{x_{1\ldots,n}\sim p}[f(x)].$$

Then

$$\mathrm{E}_{x_{1\ldots n}\sim p}\left[f(x)\cdot\sum_{i\in[n]}(x_i - p)\right] = g'(p)\cdot(1-p^2).$$

Note that when $f(x) \approx \frac{1}{n}\sum_{i=1}^{n} x_i$, then $g(p) \approx p$. Therefore, if we choose p uniformly over $[-1,1]$, then Lemma 22 implies that we get $\Omega(1)$-advantage, which results in the original fingerprinting lemma (Lemma 19). However, in Lemma 20, we are interested in a much weaker f that only satisfies $f(1,\ldots,1) = 1$ and $f(-1,\ldots,-1) = -1$, and might be arbitrary in all other inputs in $\{-1,1\}^n$. So the only thing that we know about such f is that it induces a function $g\colon [-1,1] \to [-1,1]$ such that $g(-1) = -1$ and $g(1) = 1$, and actually we can also show that $g(-1+\varepsilon) \approx -1$ and $g(1-\varepsilon) \approx 1$ for $\varepsilon = O(1/n)$. We show that these limited properties suffice for a fingerprint lemma, by choosing p from a distribution ρ that has probability density function $\propto 1/(1-p^2)$.

Proof [Proof of Lemma 20] Recall that ρ is the distribution that outputs $p = \frac{e^t-1}{e^t+1}$ for $t \leftarrow [-\ln(5n), \ln(5n)]$. Let $\mathbf{p}$ denote a random variable that is distributed according to ρ. Let's first compute its cumulative distribution

function (CDF):

$$\begin{aligned}\Pr[\mathbf{p} \leq p] &= \Pr_{t \leftarrow [-\ln(5n), \ln(5n)]}\left[\frac{e^t - 1}{e^t + 1} \leq p\right] \\ &= \Pr_{t \leftarrow [-\ln(5n), \ln(5n)]}\left[t \leq \ln\left(\frac{1+p}{1-p}\right)\right] \\ &= \frac{\ln\left(\frac{1+p}{1-p}\right) + \ln(5n)}{2\ln(5n)} \\ &= \frac{1}{2\ln(5n)} \cdot (\ln(1+p) - \ln(1-p)) + \frac{1}{2}\end{aligned}$$

Hence, the probability density function (PDF) of $\mathbf{p}$ is

$$\begin{aligned}\rho(p) &:= \frac{d\Pr[\mathbf{p} \leq p]}{dp} \\ &= \frac{1}{2\ln(5n)} \cdot \left(\frac{1}{1+p} + \frac{1}{1-p}\right) \\ &= \frac{1}{\ln(5n)} \cdot \frac{1}{1-p^2}.\end{aligned}$$

In the following, let $g(p) := \mathrm{E}_{x_{1\ldots n} \sim p}[f(x)]$. Note that $t = \ln(5n) \implies p = 1 - \frac{2}{5n+1}$ and $t = -\ln(5n) \implies p = -1 + \frac{2}{5n+1}$. In addition, note that if $x_i \sim p$, then $\Pr[x_i = 1] = \frac{1+p}{2}$ and $\Pr[x_i = -1] = \frac{1-p}{2}$. By the assumption on f, for every p it holds that

$$g(p) \geq \Pr_{x_{1\ldots n} \sim p}[\forall i : x_i = 1] - \Pr_{x_{1\ldots n} \sim p}[\exists i : x_i = -1] = 2 \cdot \left(\frac{1+p}{2}\right)^n - 1$$

and

$$g(p) \leq -\Pr_{x_{1\ldots n} \sim p}[\forall i : x_i = -1] + \Pr_{x_{1\ldots n} \sim p}[\exists i : x_i = 1] = -2 \cdot \left(\frac{1-p}{2}\right)^n + 1$$

Hence,

$$g\left(1 - \frac{2}{5n+1}\right) \geq 2 \cdot \left(1 - \frac{1}{5n+1}\right)^n - 1 \geq 2 \cdot e^{-\frac{1.1n}{5n+1}} - 1 \geq 2 \cdot e^{-1/4} - 1 \geq 0.5$$

and

$$g\left(-1 + \frac{2}{5n+1}\right) \leq -2 \cdot \left(1 - \frac{1}{5n+1}\right)^n + 1 \leq -0.5.$$

where we used the inequality $e^{-1.1y} \leq 1 - y$ for every $y \in [0, 1/6]$.

Hence, we conclude that

$$\begin{aligned}
\mathrm{E}_{p\sim\rho,\, x_{1\ldots n}\sim p}\left[f(x)\cdot\sum_{i\in[n]}(x_i-p)\right] &= \mathrm{E}_{p\sim\rho}\left[g'(p)\cdot(1-p^2)\right] \\
&= \int_{-1+\frac{2}{5n+1}}^{1-\frac{2}{5n+1}} g'(p)\cdot(1-p^2)\cdot\rho(p)dp \\
&= \frac{1}{\ln(5n)}\cdot\int_{-1+\frac{2}{5n+1}}^{1-\frac{2}{5n+1}} g'(p)dp \\
&= \frac{1}{\ln(5n)}\cdot\left(g\left(1-\frac{2}{5n+1}\right)-g\left(-1+\frac{2}{5n+1}\right)\right) \qquad (1)\\
&\geq \frac{1}{\ln(5n)}.
\end{aligned}$$

■

Acknowledgments

Eliad Tsfadia would like to thank Edith Cohen, Haim Kaplan, Yishay Mansour and Uri Stemmer for encouraging him to tackle the problem of lower bounding DP averaging and for useful discussions.

Naty Peter was supported in part by the Massive Data Institute at Georgetown University. Eliad Tsfadia was supported in part by the Fulbright Program and a gift to Georgetown University. Jonathan Ullman is supported by NSF awards CNS-2120603, CNS-2232692, and CNS-2247484.

References

Hassan Ashtiani and Christopher Liaw. Private and polynomial time algorithms for learning gaussians and beyond. In *Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 1075–1076. PMLR, 2022.

Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *FOCS*, pages 464–473. IEEE, 2014.

Raef Bassily, Albert Cheu, Shay Moran, Aleksandar Nikolov, Jonathan Ullman, and Zhiwei Steven Wu. Private query release assisted by public data. In *International Conference on Machine Learning (ICML)*, 2020.

Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5):1897–1905, 1998.

Mark Bun, Jonathan Ullman, and Salil P. Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Symposium on Theory of Computing, STOC 2014*, pages 1–10, 2014.

Mark Bun, Thomas Steinke, and Jonathan R. Ullman. Make up your mind: The price of online queries in differential privacy. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017*, pages 1306–1325, 2017.

T. Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

T. Tony Cai, Yichen Wang, and Linjun Zhang. Score attack: A lower bound technique for optimal differentially private learning. *arXiv preprint arXiv:2303.07152*, 2023.

Edith Cohen, Haim Kaplan, Yishay Mansour, Uri Stemmer, and Eliad Tsfadia. Differentially-private clustering of easy instances. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021*, volume 139, pages 2049–2059, 2021. URL https://arxiv.org/abs/2112.14445.

Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT*, volume 4004, pages 486–503, 2006a.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, volume 3876, pages 265–284, 2006b.

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: Optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 11–20. ACM, 2014.

Cynthia Dwork, Adam D. Smith, Thomas Steinke, Jonathan R. Ullman, and Salil P. Vadhan. Robust traceability from trace amounts. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015*, pages 650–669, 2015.

Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. Differentially private clustering: Tight approximation ratios. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, 2020.

Anupam Gupta, Katrina Ligett, Frank McSherry, Aaron Roth, and Kunal Talwar. Differentially private combinatorial optimization. In *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '10, pages 1106–1125, 2010.

Moritz Hardt and Jonathan Ullman. Preventing false discovery in interactive data analysis is hard. In *FOCS*, pages 454–463, 2014.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Palak Jain, Sofya Raskhodnikova, Satchit Sivakumar, and Adam Smith. The price of differential privacy under continual observation. In *International Conference on Machine Learning*, pages 14654–14678. PMLR, 2023.

Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In *Conference on Learning Theory, COLT 2019*, volume 99, pages 1853–1902. PMLR, 2019.

Gautam Kamath, Argyris Mouzakis, and Vikrant Singhal. New lower bounds for private estimation and a generalized fingerprinting lemma. *Advances in Neural Information Processing Systems*, 35:24405–24418, 2022.

Gautam Kamath, Argyris Mouzakis, Matthew Regehr, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. A bias-variance-privacy trilemma for statistical estimation. *arXiv preprint arXiv:2301.13334*, 2023.

Haim Kaplan and Uri Stemmer. Differentially private k-means with constant multiplicative error. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*, pages 5436–5446, 2018.

Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, 2018.

Shyam Narayanan. Better and simpler lower bounds for differentially private statistical estimation, 2024.

Shyam Narayanan, Vahab S. Mirrokni, and Hossein Esfandiari. Tight and robust private mean estimation with few users. In *International Conference on Machine Learning, ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 16383–16412. PMLR, 2022.

Huy L. Nguyen, Anamay Chaturvedi, and Eric Z. Xu. Differentially private k-means via exponential mechanism and max cover. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021*, pages 9101–9108, 2021.

Kobbi Nissim and Uri Stemmer. Clustering algorithms for the centralized and local models. In *Proceedings of Algorithmic Learning Theory*, volume 83 of *Proceedings of Machine Learning Research*, pages 619–653, 2018.

Kobbi Nissim, Uri Stemmer, and Salil P. Vadhan. Locating a small cluster privately. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016*, pages 413–427, 2016.

Rafail Ostrovsky, Yuval Rabani, Leonard J. Schulman, and Chaitanya Swamy. The effectiveness of lloyd-type methods for the k-means problem. *J. ACM*, 59(6):28:1–28:22, 2012.

Sofya Raskhodnikova, Satchit Sivakumar, Adam Smith, and Marika Swanberg. Differentially private sampling from distributions. *Advances in Neural Information Processing Systems*, 34:28983–28994, 2021.

Moshe Shechner, Or Sheffet, and Uri Stemmer. Private k-means clustering with stability assumptions. In *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020*, volume 108 of *Proceedings of Machine Learning Research*, pages 2518–2528, 2020.

Vikrant Singhal and Thomas Steinke. Privately learning subspaces. *Advances in Neural Information Processing Systems*, 34, 2021a.

Vikrant Singhal and Thomas Steinke. Privately learning subspaces. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*, pages 1312–1324, 2021b.

Thomas Steinke and Jonathan Ullman. Interactive fingerprinting codes and the hardness of preventing false discovery. In *COLT*, pages 1588–1628, 2015.

Thomas Steinke and Jonathan Ullman. Tight lower bounds for differentially private selection. In *IEEE Symposium on Foundations of Computer Science*, FOCS '17, 2017.

Thomas Steinke and Jonathan R. Ullman. Between pure and approximate differential privacy. *J. Priv. Confidentiality*, 7(2), 2016.

Gábor Tardos. Optimal probabilistic fingerprint codes. *J. ACM*, 55(2), 2008.

Eliad Tsfadia, Edith Cohen, Haim Kaplan, Yishay Mansour, and Uri Stemmer. Friendlycore: Practical differentially private aggregation. In *International Conference on Machine Learning, ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 21828–21863. PMLR, 2022.

Appendix A. Preliminaries

A.1. Notations

We use calligraphic letters to denote sets and distributions, uppercase for matrices and datasets, boldface for random variables, and lowercase for vectors, values and functions. For $n \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$. Throughout this paper, we use $i \in [n]$ as a row index, and $j \in [d]$ as a column index (unless otherwise mentioned).

For a matrix $X = (x_i^j)_{i\in[n],j\in[d]}$, we denote by x_i the i^{th} row of X and by x^j the j^{th} column of X. A column vector $x \in \mathbb{R}^n$ is written as $(x_1, \ldots, x_n)$ or $x = x_{1\ldots n}$, and a row vector $y \in \mathbb{R}^d$ is written as $(y^1, \ldots, y^d)$ or $y^{1\ldots d}$. In this work we consider mechanisms who receive an $n \times d$ matrix X as input, which is treated as the dataset $X = (x_1, \ldots, x_n)$ where the rows of X are the elements (and therefore, we sometimes write $X \in (\mathbb{R}^d)^n$ instead of $X \in \mathbb{R}^{n\times d}$ to emphasize it). For $d \in \mathbb{N}$ we denote by $\mathcal{P}_d$ the set of all $d \times d$ permutation matrices.

For a vector $x \in \mathbb{R}^d$ we define $\|x\|_1 = \sum_{i=1}^n |x_i|$ (the ℓ_1 norm of x), and $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ (the ℓ_2 norm of x), and for a subset $\mathcal{S} \subseteq [d]$ we define $x_{\mathcal{S}} = (x_i)_{i\in\mathcal{S}}$, and in case x is a row vector we write $x^{\mathcal{S}}$. Given two vectors $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$, we define $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ (the inner-product of x and y). For a matrix $X = (x_i^j)_{i\in[n],j\in[d]} \in \{-1, 1\}^{n\times d}$ and $b \in \{-1, 1\}$, we define the b-marked columns of X as the subset $\mathcal{J}_X^b \subseteq [d]$ defined by $\mathcal{J}_X^b = \{j \in [d] \colon x_i^j = b \text{ for all } i \in [n]\}$.

For $z \in \mathbb{R}$, we define $\mathsf{sign}(z) := \begin{cases} 1 & z \geq 0 \\ -1 & z < 0 \end{cases}$ and for $v = (v^1, \ldots, v^d) \in \mathbb{R}^d$ we define $\mathsf{sign}(v) := (\mathsf{sign}(v^1), \ldots, \mathsf{sign}(v^d)) \in \{-1, 1\}^d$.

A.2. Distributions and Random Variables

Given a distribution $\mathcal{D}$, we write $x \sim \mathcal{D}$ to denote that x is sampled according to $\mathcal{D}$. For a set $\mathcal{S}$, we write $x \leftarrow \mathcal{S}$ to denote that x is sampled from the uniform distribution over $\mathcal{S}$.

Fact 23 (Hoeffding's inequality Hoeffding (1963)) *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random variables taking integer values in the range $[a, b]$. Also, let $\mathbf{x} = \sum_{i=1}^n \mathbf{x}_i$ denote the sum of the variables and $\mu = \mathrm{E}[\mathbf{x}]$ denote its expectation. Then, for any $t > 0$,*

$$\Pr[\mathbf{x} \leq \mu - t] \leq e^{-\frac{2t^2}{n(b-a)^2}}. \tag{2}$$

$$\Pr[\mathbf{x} \geq \mu + t] \leq e^{-\frac{2t^2}{n(b-a)^2}}. \tag{3}$$

Definition 24 (Behave the same) *We say that two random variables $\mathbf{x}$ and $\mathbf{x}'$ over $\mathcal{X}$ behave the same w.p. β, if there exists a random variable $\mathbf{y}$ over $\mathcal{Y}$ (jointly distributed with $\mathbf{x}, \mathbf{x}'$), and event $E \subseteq \mathcal{Y}$ with $\Pr[\mathbf{y} \in E] \geq \beta$ such that $\mathbf{x}|_{\mathbf{y}\in E} \equiv \mathbf{x}'|_{\mathbf{y}\in E}$.*

Fact 25 *If $\mathbf{x}$ and $\mathbf{x}'$ behave the same w.p. β, then for any event F,*

$$\Pr[\mathbf{x} \in F] \geq \Pr[\mathbf{x}' \in F] - (1 - \beta).$$

Proof Let $\mathbf{y}, E$ as in theorem 24. Compute

$$\begin{aligned}\Pr[\mathbf{x} \in F] &\geq \Pr[\mathbf{y} \in E] \cdot \Pr[\mathbf{x} \in F \mid \mathbf{y} \in E] \\ &= \Pr[\mathbf{y} \in E] \cdot \Pr\left[\mathbf{x}' \in F \mid \mathbf{y} \in E\right] \\ &\geq \Pr[\mathbf{y} \in E] \cdot \frac{\Pr[\mathbf{x}' \in F] - \Pr[\mathbf{y} \notin E]}{\Pr[\mathbf{y} \in E]} \\ &= \Pr\left[\mathbf{x}' \in F\right] - \Pr[\mathbf{y} \notin E] \geq \Pr\left[\mathbf{x}' \in F\right] - (1-\beta).\end{aligned}$$

■

A.3. Algorithms

Let M be a randomized algorithm that uses m random coins. For $r \in \{0,1\}^m$ we denote by M_r the (deterministic) algorithm M after fixing its random coins to r. We use the same notation for more specific cases, e.g., if the random choices of M consist of sampling $s \leftarrow [k]$ and $P \leftarrow \mathcal{P}_d$, then for $s \in [k]$ and $P \in \mathcal{P}_d$ we denote by $\mathsf{M}_{s,P}$ the algorithm M after fixing this random choices to s and P (respectively).

A.4. Differential Privacy

Definition 26 (Differential Privacy (Dwork et al., 2006b,a)) *A randomized mechanism $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{Y}$ is* (ε,δ)-differentially private *(in short, (ε,δ)-DP) if for every neighboring databases $X = (x_1,\ldots,x_n)$, $X' = (x'_1,\ldots,x'_n) \in \mathcal{X}^n$ (i.e., differ by exactly one entry), and every set of outputs $\mathcal{T} \subseteq \mathcal{Y}$, it holds that*

$$\Pr[\mathsf{M}(X) \in \mathcal{T}] \leq e^{\varepsilon} \cdot \Pr\left[\mathsf{M}(X') \in \mathcal{T}\right] + \delta$$

A.4.1. Known Facts

Fact 27 (Post-Processing) *If $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{Y}$ is (ε,δ)-DP then for every randomized $\mathsf{F}\colon \mathcal{Y} \to \mathcal{Z}$, the mechanism $\mathsf{F} \circ \mathsf{M}\colon \mathcal{X}^n \to \mathcal{Z}$ is (ε,δ)-DP.*

Post-processing holds when applying the function on the output of the DP mechanism. In this work we sometimes need to apply the mechanism on the output of a function. While this process does not preserve DP in general, it does so assuming the function is *neighboring-preserving*.

Definition 28 (Neighboring-Preserving Algorithm) *We say that a randomized algorithm $\mathsf{G}\colon \mathcal{X}^n \to \mathcal{Y}^m$ is* neighboring-preserving *if for every neighboring $X, X' \in \mathcal{X}^n$, the outputs $\mathsf{G}(X), \mathsf{G}(X') \in \mathcal{Y}^m$ are neighboring w.p.* 1.

Fact 29 *If $\mathsf{G}\colon \mathcal{X}^n \to \mathcal{Y}^m$ is neighboring-preserving and $\mathsf{M}\colon \mathcal{Y}^m \to \mathcal{Z}$ is (ε,δ)-DP, then $\mathsf{M} \circ \mathsf{G}\colon \mathcal{X}^n \to \mathcal{Z}$ is (ε,δ)-DP.*

We also use the following fact which states that the output of a DP mechanism cannot be too correlated with one of the input points.

Fact 30 *Let $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{Y}$ be an (ε,δ)-DP mechanism, let $i \in [n]$, let $\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{x}'_i$ be i.i.d. random variables over $\mathcal{X}$, let $\mathbf{X} = (\mathbf{x}_1,\ldots,\mathbf{x}_n)$, and let $P\colon \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a (possibly randomized) predicate. Then,*

$$\Pr[P(\mathbf{x}_i, \mathsf{M}(\mathbf{X})) = 1] \leq e^{\varepsilon} \cdot \Pr\left[P(\mathbf{x}'_i, \mathsf{M}(\mathbf{X})) = 1\right] + \delta.$$

Proof We assume without loss of generality that P is deterministic since the proof for a randomized P follows by fixing the random coins of P. In the following, for $x_i \in \mathcal{X}$ let $\mathcal{T}_{x_i} = \{y \in \mathcal{Y} \colon P(x_i, y) = 1\}$. Compute

$$\begin{aligned}
\Pr[P(\mathbf{x}_i, \mathsf{M}(\mathbf{X})) = 1] &= \mathrm{E}_{x_i \sim \mathbf{x}_i}[\Pr[\mathsf{M}(\mathbf{X}\,|_{\mathbf{x}_i = x_i}) \in \mathcal{T}_{x_i}]] \\
&= \mathrm{E}_{x_i \sim \mathbf{x}'_i}[\Pr[\mathsf{M}(\mathbf{X}\,|_{\mathbf{x}_i = x_i}) \in \mathcal{T}_{x_i}]] \\
&\leq \mathrm{E}_{x_i \sim \mathbf{x}'_i}[e^{\varepsilon} \cdot \Pr[\mathsf{M}(\mathbf{X}) \in \mathcal{T}_{x_i}] + \delta] \\
&= e^{\varepsilon} \cdot \Pr\big[P(\mathbf{x}'_i, \mathsf{M}(\mathbf{X})) = 1\big] + \delta,
\end{aligned}$$

where the second equality holds since $\mathbf{x}_i \equiv \mathbf{x}'_i$, and the inequality holds since M is (ε, δ)-DP and since $\mathbf{X}\,|_{\mathbf{x}_i = x_i}$ and $\mathbf{X}$ are neighboring. ■

Appendix B. Robust Fingerprinting Lemma

In this section we extend Lemma 20 to a robust version that allows a small error probability (Lemma 31). In appendix B.1 we increase the dimension and prove the resulting correlation in Lemma 34.

Lemma 31 (Robust version of Lemma 20) *Let $F \colon \{-1,1\}^n \to [-1,1]$ be a randomized function such that $\Pr[F(1,\ldots,1) = 1] \geq 0.9$ and $\Pr[F(-1,\ldots,-1) = -1] \geq 0.9$. Then,*

$$\mathrm{E}_{p \sim \rho,\, x_{1\ldots n} \sim p}\left[F(x) \cdot \sum_{i \in [n]} (x_i - p)\right] \geq \frac{0.4}{\ln(5n)}$$

for ρ as in Lemma 20, where the expectation is also over the randomness of F.

Proof Assume that $F(x)$ is defined by $f(x; r)$ for a random $r \leftarrow \{0,1\}^m$. Let $q = 0.9$ and $\mathcal{R} = \{r \in \{0,1\}^m \colon f(1,\ldots,1;r) = 1 \wedge f(-1,\ldots,-1;r) = -1\}$. By assumption and the union bound, $|\overline{\mathcal{R}}| \leq 2(1-q) \cdot 2^m = (2-2q) \cdot 2^m$, so $|\mathcal{R}| = 2^m - |\overline{\mathcal{R}}| \geq (2q-1) \cdot 2^m$.

By Lemma 20,

$$\forall r \in \mathcal{R}: \quad \mathrm{E}_{p \sim \rho,\, x_{1\ldots n} \sim p}\left[f(x;r) \cdot \sum_{i \in [n]} (x_i - p)\right] \geq \frac{1}{\ln(5n)}.$$

By eq. (1) and since for every r, the function $f(\cdot, r)$ induces a function g with range $[-1,1]$, it holds that

$$\begin{aligned}
\forall r \notin \mathcal{R}: \quad \mathrm{E}_{p \sim \rho,\, x_{1\ldots n} \sim p}\left[f(x;r) \cdot \sum_{i \in [n]} (x_i - p)\right] &= \frac{1}{\ln(5n)} \cdot \left(g\left(1 - \frac{2}{5n+1}\right) - g\left(-1 + \frac{2}{5n+1}\right)\right) \\
&\geq -\frac{2}{\ln(5n)}.
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}\mathrm{E}_{p\sim\rho,\, x_{1\ldots n}\sim p}\left[F(x)\cdot\sum_{i\in[n]}(x_i-p)\right] &= \mathrm{E}_{r\leftarrow\{0,1\}^m,\, p\sim\rho,\, x_{1\ldots n}\sim p}\left[f(x;r)\cdot\sum_{i\in[n]}(x_i-p)\right]\\ &\geq (2q-1)\cdot\frac{1}{\ln(5n)}-(2-2q)\cdot\frac{2}{\ln(5n)}\\ &= \frac{6q-5}{\ln(5n)}\\ &= \frac{0.4}{\ln(5n)}.\end{aligned}$$

■

B.1. Increasing the Dimension

In this section, we increase the correlation achieved by our fingerprinting lemma by increasing the number of columns (the dimension of the input points).

We start by defining a *strongly-accurate* mechanism, which is a natural extension of the one-dimension case.

Definition 32 (Strongly-Accurate Mechanism) *We say that a mechanism* $\mathsf{M}\colon \{-1,1\}^{n\times d}\to[-1,1]^d$ *is* strongly-accurate *if for every* $X\in\{-1,1\}^{n\times d}$, $\mathsf{M}(X)$ *is strongly-correlated with* X *(Definition 21).*

That is, for every b-marked column $j\in[d]$, a strongly-accurate mechanism returns b as the j^{th} element of its output with high probability.

We next define our hard distribution.

Definition 33 (Distributions $\mathcal{D}'(n,d)$ and $\mathcal{D}(n,d)$) *Let* ρ *be the distribution from Lemma 20. Define* $\mathcal{D}'(n,d)$ *to be the distribution that outputs* $(x_1,\ldots,x_n,z)\in\{-1,1\}^{(n+1)\times d}$ *where* $p^1,\ldots,p^d\sim\rho$ *are sampled (independently), each* $x_i^j\in\{-1,1\}$ *is sampled with* $\mathrm{E}\left[x_i^j\right]=p^j$*, and each* z^j *is sampled with* $\mathrm{E}\left[z^j\right]=p^j$*. Define* $\mathcal{D}(n,d)$ *as the distribution of the first* n *vectors in* $\mathcal{D}'(n,d)$ *(i.e., without* z*).*

Lemma 34 *Let* $\mathsf{M}\colon\{-1,1\}^{n\times d}\to[-1,1]^d$ *be a* strongly-accurate *mechanism, and let* $(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z})\sim\mathcal{D}'(n,d)$ *(Definition 33),* $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ *and* $\mathbf{q}=\mathsf{M}(\mathbf{X})$ *be random variables. Then,*

$$\mathrm{E}\left[\sum_{i=1}^{n}(\langle\mathbf{x}_i,\mathbf{q}\rangle-\langle\mathbf{z},\mathbf{q}\rangle)\right]\geq\frac{0.4d}{\ln(5n)}.$$

Furthermore, for every $\beta\in[0,1]$*, if* $d\geq\Theta(n^2\log^2(n)\log(1/\beta))$*, then*

$$\Pr\left[\sum_{i=1}^{n}(\langle\mathbf{x}_i,\mathbf{q}\rangle-\langle\mathbf{z},\mathbf{q}\rangle)\leq\frac{0.2d}{\ln(5n)}\right]\leq\beta.$$

Proof The first part of the lemma follows from similar arguments as in the proof of Proposition 10 in Dwork et al. (2015). In particular, for a given $j\in[d]$, we fix all columns except for the j^{th} column. Since the mechanism M only gets $\mathbf{X}$ as an input, and each columns j of $\mathbf{X}$ is sampled independently, then there is

some $F_j\colon \{-1,1\}^n \to [-1,1]^d$ for which $\mathbf{q}^j \sim F_j(\mathbf{x}^j)$, where $\mathbf{x}^j = (\mathbf{x}_1^j, \ldots, \mathbf{x}_n^j)$ (the j'th column of $\mathbf{X}$), such that Lemma 31 holds for it. Then, by this and the fact that M is a strongly-accurate mechanism,

$$\mathrm{E}\left[\mathbf{q}^j \cdot \sum_{i\in[n]}(\mathbf{x}_i^j - \mathbf{z}^j)\right] = \mathrm{E}\left[F_j(\mathbf{x}^j) \cdot \sum_{i\in[n]}(\mathbf{x}_i^j - \mathbf{p}_j)\right] \geq \frac{0.4}{\ln(5n)},$$

where $\mathbf{p}^1, \ldots, \mathbf{p}^d$ are the expectations that were sampled in the process of sampling $\mathbf{X}, \mathbf{z}$ (Definition 33). Thus, by linearity of expectation, we conclude the first part of the lemma.

The second part is proven similarly to Lemma 11 in Dwork et al. (2015). Let us define the random variable $\mathbf{a}_j = \mathbf{q}^j \cdot \sum_{i\in[n]}(\mathbf{x}_i^j - \mathbf{z}^j)$ for every $j \in [d]$. Then, we get that

$$\sum_{i=1}^{n}(\langle \mathbf{x}_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle) = \sum_{j=1}^{d}\mathbf{q}^j \sum_{i=1}^{n}(\mathbf{x}_i^j - \mathbf{z}^j) = \sum_{j=1}^{d}\mathbf{a}_j,$$

and from the first part we have $\mathrm{E}\left[\sum_{j=1}^{d}\mathbf{a}_j\right] \geq \frac{0.4d}{\ln(5n)}$. In order to use Hoeffding's inequality on the sum $\sum_{j=1}^{d}\mathbf{a}_j$, we first assume that the random variables $\mathbf{a}_1, \ldots, \mathbf{a}_d$ are independent. Later, as in Dwork et al. (2015), we show how to remove this assumption. Next, observe that $-2n \leq |\mathbf{a}_j| \leq 2n$ for every $j \in [n]$. Therefore, using an Hoeffding's inequality (2), it follows that

$$\Pr\left[\sum_{i=1}^{n}(\langle \mathbf{x}_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle) \leq \frac{0.2d}{\ln(5n)}\right] = \Pr\left[\sum_{j=1}^{d}\mathbf{a}_j \leq \frac{0.2d}{\ln(5n)}\right] \leq e^{-\frac{2(0.2d/\ln(5n))^2}{d(4n)^2}} = e^{\Theta\left(-\frac{d}{n^2\ln^2(n)}\right)} \leq \beta.$$

Now, for the case that $\mathbf{a}_1, \ldots, \mathbf{a}_d$ are not independent, we use the fact that the sum $\sum_{j=1}^{d}\mathbf{a}_j$ concentrates as if they were independent, which follows since for every $j \in [d]$, the expected value of $\mathbf{a}_j$ is the same even when conditioned on the other variables of $\{\mathbf{a}_1, \ldots, \mathbf{a}_d\} \setminus \mathbf{a}_j$. That is,

$$\mathrm{E}[\mathbf{a}_j] = \mathrm{E}[\mathbf{a}_j \mid \{\mathbf{a}_1, \ldots, \mathbf{a}_d\} \setminus \mathbf{a}_j].$$

■

Appendix C. Framework for Lower Bounds

Following section 2.3, in this section we present our general framework for lower bounding DP algorithms (stated in Lemma 40) that is based on our new hard-distribution.

Definition 35 (Algorithm $\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}$) *Let $\mathcal{V}$, $\mathcal{W}$ be domains, and let $n_0, d_0, n \in \mathbb{N}$. Let $(\mathsf{M}, \mathsf{F}, \mathsf{G})$ be a triplet of randomized algorithms of types $\mathsf{G}\colon \{-1,1\}^{n_0\times d_0} \times \mathcal{V} \to \mathcal{X}^n$, $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{W}$, and $\mathsf{F}\colon \mathcal{V}\times\mathcal{W} \to [-1,1]^{d_0}$, each uses m random coins. Let $\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}\colon \{-1,1\}^{n_0\times d_0} \to [-1,1]^{d_0}$ be the randomized algorithm that on input $X \in \{-1,1\}^{n_0\times d_0}$, samples $v \leftarrow \mathcal{V}$, $Y \sim \mathsf{G}(X, v)$ and $w \sim \mathsf{M}(Y)$, and outputs $q \sim \mathsf{F}(v, w)$.*

Definition 36 (β-Leaking)

Let $\mathsf{M}, \mathsf{F}, \mathsf{G}$ be randomized algorithms as in Definition 35, each uses m random coins, and let $\mathcal{D}(n_0, d_0)$ be the distribution from Definition 33. We say that the triplet $(\mathsf{M}, \mathsf{F}, \mathsf{G})$ is β-leaking *if*

$$\Pr_{r,r',r''\leftarrow\{0,1\}^m,\, X\leftarrow\mathcal{D}(n_0,d_0)}\left[\mathsf{A}^{\mathsf{M}_r,\mathsf{F}_{r'},\mathsf{G}_{r''}}(X) \text{ is strongly-correlated with } X\right] \geq \beta,$$

(where recall that M_r denotes the algorithm M when fixing its random coins to r, and $\mathsf{F}_{r'}, \mathsf{G}_{r''}$ are similarly defined).

In the following, fix a β-leaking triplet $(\mathsf{M},\mathsf{F},\mathsf{G})$, and let $\mathsf{A}=\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}$ and $\mathsf{A}_{r,r',r''}=\mathsf{A}^{\mathsf{M}_r,\mathsf{F}_{r'},\mathsf{G}_{r''}}$.

Claim 37 *If M is (ε,δ)-DP, and $\mathsf{G}(\cdot,v)$ is neighboring-preserving (Definition 28) for every $v\in\mathcal{V}$, then A is (ε,δ)-DP.*

Proof Fix $v\in\mathcal{V}$, and let $\mathsf{G}_v=\mathsf{G}(\cdot,v)$ and $\mathsf{F}_v=\mathsf{F}(v,\cdot)$. Since G_v is neighboring-preserving and M is (ε,δ)-DP, then $\mathsf{M}\circ\mathsf{G}_v$ is (ε,δ)-DP (Fact 29). By post-processing, $\mathsf{F}_v\circ\mathsf{M}\circ\mathsf{G}_v$ is also (ε,δ)-DP. Since the above holds for any $v\in\mathcal{V}$ and since $\mathsf{A}=(\mathsf{F}_v\circ\mathsf{M}\circ\mathsf{G}_v)_{v\leftarrow\mathcal{V}}$, we conclude that A is (ε,δ)-DP. ■

Claim 38 *Let $(\mathbf{x}_1,\ldots,\mathbf{x}_{n_0},\mathbf{z})\sim\mathcal{D}(n_0,d_0)$ (Definition 33) and let $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_{n_0})\in\{-1,1\}^{n_0\times d_0}$. Assuming $d_0\geq\Theta(n_0^2\log^2(n_0)\log(1/\beta))$, it holds that*

$$\Pr\left[\sum_{i=1}^{n_0}(\langle\mathbf{x}_i,\mathsf{A}(\mathbf{X})\rangle-\langle\mathbf{z},\mathsf{A}(\mathbf{X})\rangle)\geq\frac{0.2d_0}{\ln(5n_0)}\right]\geq\beta/2.$$

Proof Consider the algorithm $\mathsf{A}'\colon\{-1,1\}^{n_0\times d_0}\to[-1,1]^{d_0}$ that on input $X\in\{-1,1\}^{n_0\times d_0}$, samples $r,r',r''\leftarrow\{0,1\}^m$ and checks if $\mathsf{A}_{r,r',r''}(X)$ is strongly-correlated with X. If it does, it outputs $q\sim\mathsf{A}_{r,r',r''}(X)$. Otherwise, it outputs $q=(q^1,\ldots,q^{d_0})$ such that $q^j=b$ for every $b\in\{-1,1\}$ and $j\in\mathcal{J}_X^b$. By definition, A' is strongly-accurate. Therefore, by Lemma 34 we obtain that

$$\Pr\left[\sum_{i=1}^{n_0}(\langle\mathbf{x}_i,\mathsf{A}'(\mathbf{X})\rangle-\langle\mathbf{z},\mathsf{A}'(\mathbf{X})\rangle)\geq\frac{0.2d_0}{\ln(5n_0)}\right]\geq1-\beta/2.$$

But recall that $\mathsf{A}=\mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}$, where $(\mathsf{M},\mathsf{F},\mathsf{G})$ is β-leaking (Definition 36). Therefore, $\mathsf{A}'(\mathbf{X})$ behaves as $\mathsf{A}(\mathbf{X})$ with probability at least β (Definition 24). Thus by Fact 25

$$\Pr\left[\sum_{i=1}^{n_0}(\langle\mathbf{x}_i,\mathsf{A}(\mathbf{X})\rangle-\langle\mathbf{z},\mathsf{A}(\mathbf{X})\rangle)\geq\frac{0.2d_0}{\ln(5n_0)}\right]\geq(1-\beta/2)-(1-\beta)=\beta/2.$$

■

Claim 39 *If $d_0\geq\Theta(n_0^2\log^2(n_0)\log(n_0/\beta))$, then A is not $\left(1,\frac{\beta}{4n_0}\right)$-DP.*

Proof Let $(\mathbf{x}_1,\ldots,\mathbf{x}_{n_0},\mathbf{z})\sim\mathcal{D}(n_0,d_0)$, let $\mathbf{X}=(\mathbf{x}_1,\ldots,\mathbf{x}_{n_0})$, and let $\mathbf{q}=\mathsf{A}(\mathbf{X})$.

By Claim 38 and the union bound, we deduce that there exists $i\in[n_0]$ such that

$$\Pr\left[\langle\mathbf{x}_i,\mathbf{q}\rangle-\langle\mathbf{z},\mathbf{q}\rangle>\frac{0.2d_0}{n_0\ln(5n_0)}\right]\geq\frac{\beta}{2n_0}.\tag{4}$$

In the following, let $\mathbf{p}_1,\ldots,\mathbf{p}_{d_0}\in[-1,1]$ be the (random variables of the) expectations that were chosen as part of the sampling of $(\mathbf{x}_1,\ldots,\mathbf{x}_{n_0},\mathbf{z})$ from $\mathcal{D}(n_0,d_0)$ (as defined in Definition 33), and in the rest of the proof we fix $(\mathbf{p}_1,\ldots,\mathbf{p}_{d_0})=(p_1,\ldots,p_{d_0})$ such that eq. (4) holds under this fixing.

For the above $i\in[n_0]$, let $\mathbf{x}_i'\in\{-1,1\}^{d_0}$ be an independent random variable such that each $(\mathbf{x}_i')^j\sim p_j$ (independently) for every $j\in[d_0]$. Note that $\mathbf{x}_1,\ldots,\mathbf{x}_{n_0},\mathbf{x}_i',\mathbf{z}$ are i.i.d. random variables. However, while $\mathbf{q}$ depends on $\mathbf{x}_i$, it is independent of $\mathbf{x}_i'$ and $\mathbf{z}$.

Assume towards a contradiction that A is $\left(1,\frac{\beta}{4n_0}\right)$-DP. By applying Fact 30 with the random predicate $P(x_i,q)=\mathbb{1}_{\{\langle x_i,q\rangle-\langle\mathbf{z},q\rangle>\frac{0.2d_0}{n_0\ln(5n_0)}\}}$, we deduce from eq. (4) that

$$\Pr\left[\langle \mathbf{x}'_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle > \frac{0.2d_0}{n_0 \ln(5n_0)}\right] \geq e^{-1} \cdot \left(\frac{\beta}{2n_0} - \frac{\beta}{4n_0}\right) \geq \frac{\beta}{4en_0}. \tag{5}$$

In the following, we prove that since $\mathbf{x}'_i$ is independent of $\mathbf{q}$, then the above probability is much smaller, which will lead to a contradiction.

For every $j \in [d_0]$, define the random variable $\mathbf{b}_j = \mathbf{q}^j((\mathbf{x}'_i)^j - \mathbf{z}^j)$. Then, we get that

$$\langle \mathbf{x}'_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle = \sum_{j=1}^{d_0} \mathbf{q}^j((\mathbf{x}'_i)^j - \mathbf{z}^j) = \sum_{j=1}^{d_0} \mathbf{b}_j.$$

Since $\mathbf{x}'_i, \mathbf{z}$ and $\mathbf{q}$ are independent, the above equality yields that $\mathrm{E}\left[\sum_{j=1}^{d_0} \mathbf{b}_j\right] = 0$. Hence, similarly to the proof of Lemma 34 and since $-2 \leq |\mathbf{b}_j| \leq 2$ for every $j \in [n_0]$, we obtain by Hoeffding's inequality (3) that

$$\Pr\left[\langle \mathbf{x}'_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle > \frac{0.2d_0}{n_0 \ln(5n_0)}\right] = \Pr\left[\sum_{j=1}^{d_0} \mathbf{b}_j > \frac{0.2d_0}{n_0 \ln(5n_0)}\right] \leq e^{-\frac{d_0}{200n_0^2 \ln^2(5n_0)}} \leq \frac{\beta}{20n_0}, \tag{6}$$

where the last inequality holds since by assumption $d_0 \geq 200n_0^2 \ln^2(20n_0) \log\left(\frac{20n_0}{\beta}\right)$ (holds when choosing a large enough constant in the Θ expression). This is in contradiction to eq. (5), so we conclude that A is not $\left(1, \frac{\beta}{4n_0}\right)$-DP. ■

We now ready to state and prove the guarantee of our framework.

Lemma 40 (Framework for Lower Bounds) *Let $\beta \in (0,1]$, $n_0, n, d_0 \in \mathbb{N}$. Let $\mathsf{M}\colon \mathcal{X}^n \to \mathcal{W}$ be an algorithm such that there exists two algorithms $\mathsf{G}\colon \{-1,1\}^{n_0 \times d_0} \times \mathcal{V} \to \mathcal{X}^n$ and $\mathsf{F}\colon \mathcal{V} \times \mathcal{W} \to [-1,1]^{d_0}$ such that the triplet $(\mathsf{M}, \mathsf{F}, \mathsf{G})$ is β-leaking (Definition 36). If M is $\left(1, \frac{\beta}{4n_0}\right)$-DP, then $n_0 \geq \Omega\left(\frac{\sqrt{d_0}}{\log^{1.5}(d_0/\beta)}\right)$.*

Proof Let $c > 0$ be the hidden constant in the Θ expression of Claim 39. If $d_0 \geq c \cdot n_0^2 \log^2(n_0) \log(n_0/\beta)$, then $\mathsf{A} = \mathsf{A}^{\mathsf{M},\mathsf{F},\mathsf{G}}$ is not $\left(1, \frac{\beta}{4n_0}\right)$-DP (Claim 39), and therefore M is not $\left(1, \frac{\beta}{4n_0}\right)$-DP (Claim 37), contradiction. Thus $d_0 \leq c \cdot n_0^2 \log^2(n_0) \log(n_0/\beta)$ which implies that $n_0 \geq \Omega\left(\frac{\sqrt{d_0}}{\log^{1.5}(d_0/\beta)}\right)$. ■

Appendix D. Padding And Permuting (PAP) FPC

In this section, we present the PAP transformation and prove its main property (stated in Lemma 43). In appendices D.1 and D.2 we use our framework (Lemma 40) with the PAP transformation to prove Theorems 4 and 6.

Definition 41 ($\mathsf{PAP}_{n,d_0,\ell}$) *Let $\ell, n, d_0 \in \mathbb{N}$, and let $d = d_0 + 2\ell$. We define $\mathsf{PAP}_{n,d_0,\ell}\colon \{-1,1\}^{n \times d_0} \times \mathcal{P}_d \to \{-1,1\}^{n \times d}$ as the function that given $X \in \{-1,1\}^{n \times d_0}$ and a permutation matrix $P \in \mathcal{P}_d$ as inputs, outputs $X' = X'' \cdot P$ (i.e., permutes the columns of X'' according to P), where X'' is the $\{-1,1\}^{n \times d}$ matrix after appending ℓ 1-marked and ℓ (-1)-marked columns to X.*

Note that for every $n, d_0, \ell \in \mathbb{N}$ and $P \in \mathcal{P}_d$, the function $\mathsf{PAP}_{n,d_0,\ell}(\cdot, P)$ is neighboring-preserving (Definition 28).

The following lemma shows how PAP can be used to transform strong-agreement into a strong-correlation guarantee.

Lemma 42 *Let $\ell, n, d_0 \in \mathbb{N}$ such that $d = d_0 + 2\ell$. Let $X \in \{-1,1\}^{n \times d_0}$, define the random variables $\mathbf{P} \leftarrow \mathcal{P}_d$ and $\mathbf{Y} = \mathsf{PAP}_{n,d_0,\ell}(X, \mathbf{P})$, and let $\mathsf{M}\colon \{-1,1\}^{n \times d} \to [-1,1]^d$ be a mechanism that for every input $Y \in \mathrm{Supp}(\mathbf{Y})$, outputs $q \in [-1,1]^d$ that strongly-agrees with Y (Definition 2). Then $(\mathsf{M}(\mathbf{Y}) \cdot \mathbf{P}^T)^{1,\ldots,d_0}$ is strongly-correlated with X (Definition 21).*

Proof The proof follows since for every $b \in \{-1,1\}$ and $j \in \mathcal{J}_X^b$,

$$\begin{aligned} \Pr_{(Y,P)\sim(\mathbf{Y},\mathbf{P})}\left[(\mathsf{M}(Y)\cdot P^T)^j = b\right] &= \mathrm{E}_{Y\sim\mathbf{Y},\, j'\leftarrow\mathcal{J}_Y^b}\left[\Pr\left[\mathsf{M}(Y)^{j'} = b\right]\right] \\ &\geq 0.9 \cdot \mathrm{E}_{Y\sim\mathbf{Y}}[\Pr[\mathsf{M}(Y) \text{ strongly-agrees with } Y]] \\ &= 0.9, \end{aligned} \tag{7}$$

where all the probabilities are also taken over the random coins of M. The equality holds since from the point of view of M, which does not know the random permutation P, every b-marked column of Y has the same probability to be the b-marked column j of X. The first inequality holds since for every Y,

$$\mathrm{E}_{j'\leftarrow\mathcal{J}_Y^b}\left[\Pr\left[\mathsf{M}(Y)^{j'} = b \mid \mathsf{M}(Y) \text{ strongly-agrees with } Y\right]\right] \geq 0.9.$$

The last inequality in eq. (7) holds since by the assumption on M, for every $Y \in \mathrm{Supp}(\mathbf{Y})$, the output $\mathsf{M}(Y)$ strongly-agrees with Y w.p. 1. ■

We now prove the main property of our PAP technique, which transforms any probability of strong-agreement to the same probability of strong-correlation .

Lemma 43 *Let $\ell, n, d_0 \in \mathbb{N}$ such that $d = d_0 + 2\ell$. Let $\mathsf{M}\colon \{-1,1\}^{n \times d} \to [-1,1]^d$ be a mechanism that uses m random coins, define the random variable $\mathbf{P} \leftarrow \mathcal{P}_d$, and for $X \in \{-1,1\}^{n \times d_0}$ define $\mathbf{Y}_X = \mathsf{PAP}(X, \mathbf{P})$. Then for any distribution $\mathcal{D}$ over $\{-1,1\}^{n \times d_0}$:*

$$\begin{aligned} &\Pr_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}\left[(\mathsf{M}_r(\mathbf{Y}_X)\cdot\Pi^T)^{1,\ldots,d_0} \textit{ is strongly-correlated with } X\right] \\ &\geq \mathrm{E}_{X\sim\mathcal{D}}[\Pr[\mathsf{M}(\mathbf{Y}_X) \textit{ strongly-agrees with } \mathbf{Y}_X]]. \end{aligned}$$

Proof Let $\beta = \mathrm{E}_{X\sim\mathcal{D}}[\Pr[\mathsf{M}(\mathbf{Y}_X) \text{ strongly-agrees with } \mathbf{Y}_X]]$, and let $\mathsf{PAP} = \mathsf{PAP}_{n,d_0,\ell}$. Also, let $\mathsf{M}'\colon \{-1,1\}^{n \times d} \to [-1,1]^d$ be the mechanism that on input $Y \in \{-1,1\}^{n \times d}$, samples $q \sim \mathsf{M}(Y)$ and checks if it strongly-agrees with Y. If it does, it outputs q. Otherwise, it outputs an (arbitrary) vector $q' \in [-1,1]^d$ that strongly-agrees with Y. In the following, let $\mathbf{r} \leftarrow \{0,1\}^m$, $\mathbf{X} \sim \mathcal{D}(n, d_0)$, $\mathbf{P} \leftarrow \mathcal{P}_d$ and $\mathbf{Y} = \mathbf{Y}_{\mathbf{X}} (= \mathsf{PAP}(\mathbf{X}, \mathbf{P}))$ be random variables, and let $\mathbf{q} = \mathsf{M}_{\mathbf{r}}(\mathbf{Y})$ and $\mathbf{q}' = \mathsf{M}'_{\mathbf{r}}(\mathbf{Y})$. Let $E = \{(r, X, P)\colon \mathsf{M}_r(\mathsf{PAP}(X, P)) = \mathsf{M}'_r(\mathsf{PAP}(X, P))\}$, and note that $\Pr[(\mathbf{r}, \mathbf{X}, \mathbf{P}) \in E] = \beta$. In addition, for $(r, X) \in \mathrm{Supp}(\mathbf{r}) \times \mathrm{Supp}(\mathbf{X})$, let $E_{r,X} = \{P\colon (r, X, P) \in E\}$ and let $\beta_{r,X} = \Pr[\mathbf{P} \in E_{r,X}]$. By definition, the following holds:

1. $\forall (r, X) \in \mathrm{Supp}(\mathbf{r}) \times \mathrm{Supp}(\mathbf{X})$: $\mathbf{q}|_{(\mathbf{r},\mathbf{X})=(r,X)}$ behaves as $\mathbf{q}'|_{(\mathbf{r},\mathbf{X})=(r,X)}$ w.p. $\beta_{r,X}$ (Definition 24).
2. $\mathrm{E}_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}[\beta_{r,X}] = \beta$.
3. $\forall (r, X) \in \mathrm{Supp}(\mathbf{r}) \times \mathrm{Supp}(\mathbf{X})$: $(\mathbf{q}'|_{(\mathbf{r},\mathbf{X})=(r,X)} \cdot \mathbf{P}^T)^{1,\ldots,d_0}$ is strongly-correlated with X (holds by applying Lemma 42 on the mechanism M'_r).

Thus, we conclude that

$$\begin{aligned}
&\Pr_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}\Big[(\mathsf{M}_r(\mathbf{Y}_X)\cdot\mathbf{P}^T)^{1,\ldots,d_0} \text{ is strongly-correlated with } X\Big]\\
&=\Pr_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}\Big[(\mathbf{q}|_{(\mathbf{r},\mathbf{X})=(r,X)}\cdot\mathbf{P}^T)^{1,\ldots,d_0} \text{ is strongly-correlated with } X\Big]\\
&\geq \mathrm{E}_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}[1-(1-\beta_{r,X})]\\
&=\beta.
\end{aligned}$$

The inequality holds by items 1 and 3 and theorem 25, and the last equality holds by item 2. ■

D.1. Proving Theorem 4 (Basic Tool)

Theorem 44 (Restatement of Theorem 4) *If* $\mathsf{M}\colon(\{-1,1\}^d)^n\to[-1,1]^d$ *is an* (α,β)*-weakly-accurate (Definition 3)* $(1,\frac{\beta}{4n})$*-DP mechanism, then* $n\geq\Omega(\sqrt{\alpha d}/\log^{1.5}(\alpha d/\beta))$.

Proof We prove the theorem by applying Lemma 40 with $n_0=n$ and $d_0=d-2\ell$ for $\ell=\left\lceil\frac{1}{2}(1-\alpha)d\right\rceil$. Let $\mathsf{PAP}=\mathsf{PAP}_{n,d_0,\ell}$ (Definition 41), $\mathcal{D}=\mathcal{D}(n,d_0)$ (Definition 33), and define $\mathcal{V}=\mathcal{P}_d$, $\mathcal{W}=[-1,1]^d$, $\mathsf{G}=\mathsf{PAP}$, $\forall(P,w)\in\mathcal{V}\times\mathcal{W}:\mathsf{F}(P,w)=(w\cdot P^T)^{1,\ldots,d_0}$, and for $r\in\{0,1\}^m$ (random coins for M) define $\mathsf{A}^{\mathsf{M}_r,\mathsf{F},\mathsf{G}}$ as in Definition 35 (note that F and G are deterministic). By definition, $\mathsf{G}(\cdot,P)$ is neighboring-preserving for every $P\in\mathcal{P}_d$ (Definition 28). In the following, let $\mathbf{P}\leftarrow\mathcal{P}_d$ and for $X\in\{-1,1\}^{n\times d_0}$ define $\mathbf{Y}_X=\mathsf{PAP}(X,\mathbf{P})(=\mathsf{G}(X,\mathbf{P}))$. Compute

$$\begin{aligned}
&\Pr_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}\Big[\mathsf{A}^{\mathsf{M}_r,\mathsf{F},\mathsf{G}}(X) \text{ is strongly-correlated with } X\Big]\\
&=\Pr_{r\leftarrow\{0,1\}^m,\, X\sim\mathcal{D}}\Big[(\mathsf{M}_r(\mathbf{Y}_X)\cdot\mathbf{P}^T)^{1,\ldots,d_0} \text{ is strongly-correlated with } X\Big]\\
&\geq \mathrm{E}_{X\sim\mathcal{D}}[\Pr[\mathsf{M}(\mathbf{Y}_X) \text{ strongly-agrees with } \mathbf{Y}_X]]\\
&\geq\beta.
\end{aligned}$$

The first inequality hold by Lemma 43. The last inequality holds since M is (α,β)-weakly-accurate and for every $X\in\{-1,1\}^{n\times d_0}$ and $Y\in\mathrm{Supp}(Y_X)$ it holds that $\left|\mathcal{J}_Y^1\right|,\left|\mathcal{J}_Y^{-1}\right|\geq\ell\geq\frac{1}{2}(1-\alpha)d$.

Thus by Lemma 40, $n\geq\Omega(\sqrt{d_0}/\log^{1.5}(d_0/\beta))$, and the proof follows since $d_0=\Theta(\alpha d)$. ■

D.2. Proving Theorem 6 (Extended Tool)

We prove Theorem 6 using multiple PAP-FPC copies.

Theorem 45 (Restatement of Theorem 6) *Let* $\alpha,\beta\in(0,1]$, $n,k,d\in\mathbb{N}$ *such that* n *is a multiple of* k. *If* $\mathsf{M}\colon(\{-1,1\}^d)^n\to[-1,1]^d$ *is an* (k,α,β)*-weakly-accurate (Definition 5)* $(1,\frac{\beta}{4n})$*-DP mechanism, then* $n\geq\Omega(k\sqrt{\alpha d}/\log^{1.5}(k\alpha d/\beta))$.

Proof We prove the theorem by applying Lemma 40 with $n_0=n/k$ and $d_0=d-2\ell$ for $\ell=\left\lceil\frac{1}{2}(1-\alpha)d\right\rceil$. Let $\mathcal{V}=[k]\times(\mathcal{P}_d)^k$, $\mathcal{W}=[-1,1]^d$, $\mathsf{PAP}=\mathsf{PAP}_{n_0,d_0,\ell}$, and $\mathcal{D}=\mathcal{D}(n_0,d_0)$. Consider the following algorithm $\mathsf{H}\colon\{-1,1\}^{n_0\times d}\times\mathcal{V}\to\{-1,1\}^{n\times d}$ that on inputs $Y\in\{-1,1\}^{n_0\times d}$ and $v=(s,\boldsymbol{P}=(P_1,\ldots,P_k))\in\mathcal{V}$, act as follows:

1. Sample $\boldsymbol{A}=(A_1,\ldots,A_k)\sim\mathcal{D}^k$.

2. For $t \in [k]$ set $Y_t = \begin{cases} Y & t = s \\ \mathsf{PAP}(A_t, P_t) & t \neq s \end{cases}$.

   Denote $Y_t = (x'_{(t-1)n_0+1}, \ldots, x'_{tn_0}) \in \{-1,1\}^{n_0 \times d}$.

3. Output $X' = (x'_1, \ldots, x'_n) \in \{-1,1\}^{n \times d}$.

Define $\mathsf{G}\colon \{-1,1\}^{n_0 \times d_0} \times \mathcal{V} \to \{-1,1\}^{n \times d}$ as the algorithm that on inputs $X \in \{-1,1\}^{n_0 \times d_0}$ and $v = (s, \boldsymbol{P}) \in \mathcal{V}$ for $s \in [k]$ and $\boldsymbol{P} = (P_1, \ldots, P_k) \in (\mathcal{P}_d)^k$: Computes $Y = \mathsf{PAP}(X, P_s)$, and outputs $X' \sim \mathsf{H}(Y, v)$. Note that for every $v \in \mathcal{V}$, $\mathsf{G}(\cdot, v)$ is neighboring-preserving (Definition 28).

Define $\mathsf{F}\colon \mathcal{V} \times [-1,1]^d \to [-1,1]^{d_0}$ as the algorithm that on inputs $v = (s, \boldsymbol{P}) \in \mathcal{V}$ and $w \in [-1,1]^d$, outputs $(w \cdot P_s^T)^{1,\ldots,d_0}$. Note that F is deterministic, and the random choice of G is $\boldsymbol{A} \sim \mathcal{D}^k$ (chosen when executing H).

Our goal is to show that

$$\Pr_{r \leftarrow \{0,1\}^m,\, \boldsymbol{A} \sim \mathcal{D}^k,\, X \sim \mathcal{D}} \left[\mathsf{A}^{\mathsf{M}_r, \mathsf{F}, \mathsf{G}_{\boldsymbol{A}}}(X) \text{ is strongly-correlated with } X \right] \geq \beta/k. \tag{8}$$

Given that eq. (8) holds and since M is $\left(1, \frac{\beta}{4kn_0}\right)$-DP, we deduce by Lemma 40 that $n_0 \geq \Omega(\sqrt{d_0}/\log^{1.5}(kd_0/\beta))$ and the proof follows since $n_0 = n/k$ and $d_0 = \Theta(\alpha d)$.

We now focus on proving eq. (8). Consider the mechanism $\mathsf{M}'\colon \{-1,1\}^{n_0 \times d} \to [-1,1]^d$ that on input $Y \in \{-1,1\}^{n_0 \times d}$, samples $v = (s, \boldsymbol{P}) \leftarrow \mathcal{V}$, computes $X' = \mathsf{H}(Y, v)$ and outputs $w \sim \mathsf{M}(X')$.

In the following, let $\mathbf{P} \leftarrow \mathcal{P}_k$ be a random variable, and for $X \in \{-1,1\}^{n_0 \times d_0}$ define $\mathbf{Y}_X = \mathsf{PAP}(X, \mathbf{P})$. Compute

$$\begin{aligned}
&\mathrm{E}_{X \sim \mathcal{D}} \left[\Pr \left[\mathsf{M}'(\mathbf{Y}_X) \text{ strongly-agrees with } \mathbf{Y}_X \right] \right] \\
&= \mathrm{E}_{X \sim \mathcal{D},\, v=(s,\boldsymbol{P}) \leftarrow \mathcal{V},\, \boldsymbol{A} \sim \mathcal{D}^k} [\Pr[\mathsf{M}(\mathsf{G}_{\boldsymbol{A}}(X, v)) \text{ strongly-agrees with } \mathsf{PAP}(X, P_s)]] \\
&= \mathrm{E}_{v=(s,\boldsymbol{P}) \leftarrow \mathcal{V},\, \boldsymbol{A} \sim \mathcal{D}^k} [\Pr[\mathsf{M}(\mathsf{G}_{\boldsymbol{A}}(A_s, v)) \text{ strongly-agrees with } \mathsf{PAP}(A_s, P_s)]] \\
&= \mathrm{E}_{v=(s,\boldsymbol{P}) \leftarrow \mathcal{V},\, \forall t \in [k]\colon Y_t = (x'_{(t-1)n_0+1}, \ldots, x'_{tn_0}) \sim \mathsf{PAP}(\mathcal{D}, P_t)} \left[\Pr \left[\mathsf{M}(x'_1, \ldots, x'_n) \text{ strongly-agrees with } Y_s \right] \right] \\
&\geq \beta/k.
\end{aligned}$$

The first equality holds since by the definitions of M' and G, for every X, the pair $(\mathsf{M}'(\mathbf{Y}_X), \mathbf{Y}_X)$ has the same distribution as the pair $(\mathsf{M}(\mathsf{G}_{s,\boldsymbol{A}}(X, v)), \mathsf{PAP}(X, P_s))_{v=(s,\boldsymbol{P}) \leftarrow \mathcal{V},\, \boldsymbol{A} \sim \mathcal{D}^k}$. The second equality holds since $\mathsf{G}_{\boldsymbol{A}}(\cdot, (s, \boldsymbol{P}))$ does not use the value of A_s which is drawn as X. The third equality holds by the definition of G. The inequality holds since M is an (k, α, β)-weakly-accurate mechanism, and it gets $X' = (x'_1, \ldots, x'_n)$ where each $Y_t = (x'_{(t-1)n_0+1}, \ldots, x'_{tn_0})$ has $\left|\mathcal{J}^1_{Y_t}\right|, \left|\mathcal{J}^{-1}_{Y_t}\right| \geq \ell \geq \frac{1}{2}(1-\alpha)d$. Thus w.p. β, M is guaranteed to output $w \in [-1,1]^d$ that strongly-agrees with one of the $Y_1, \ldots, Y_k$. But since they are drawn independently from the same distribution, the probability to agree with Y_s for a random $s \leftarrow [k]$ decreases by a factor k, and overall is at least β/k.

We now apply Lemma 43 to get that

$$\Pr_{\substack{\boldsymbol{A} \sim \mathcal{D}^k,\, v=(s,\boldsymbol{P}) \leftarrow \mathcal{V} \\ r \leftarrow \{0,1\}^m,\, X \sim \mathcal{D}}} \left[\left(\mathsf{M}'_{v,\boldsymbol{A},r}(\mathbf{Y}_X) \cdot \mathbf{P}^T \right)^{1,\ldots,d_0} \text{ is strongly-correlated with } X \right] \geq \beta/k.$$

Since $(\mathsf{M}'_{v,\boldsymbol{A},r}(\mathbf{Y}_X) \cdot \mathbf{P}^T)_{r \leftarrow \{0,1\}^m,\, \boldsymbol{A} \sim \mathcal{D}^k,\, v \leftarrow \mathcal{V}}$ has the same distribution as $\left(\mathsf{M}_r(\mathsf{G}_{\boldsymbol{A}}(X, \mathbf{v})) \cdot \mathbf{P}_{\mathbf{s}}^T\right)_{r \leftarrow \{0,1\}^m,\, \boldsymbol{A} \sim \mathcal{D}^k}$ for $\mathbf{v} = (\mathbf{s}, \mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_k)) \leftarrow \mathcal{V}$, we deduce that

$$\Pr_{X \sim \mathcal{D},\, r \leftarrow \{0,1\}^m,\, \boldsymbol{A} \sim \mathcal{D}^k} \left[\left(\mathsf{M}_r(\mathsf{G}_{\boldsymbol{A}}(X, \mathbf{v})) \cdot \mathbf{P}_{\mathbf{s}}^T \right)^{1,\ldots,d_0} \text{ is strongly-correlated with } X \right] \geq \beta/k.$$

We thus conclude that

$$\begin{aligned}
&\Pr_{r\leftarrow\{0,1\}^m,\boldsymbol{A}\sim\mathcal{D}^k,\,X\sim\mathcal{D}}\left[\mathsf{A}^{\mathsf{M}_r,\mathsf{F},\mathsf{G}_{\boldsymbol{A}}}(X) \text{ is strongly-correlated with } X\right]\\
&= \Pr_{r\leftarrow\{0,1\}^m,\boldsymbol{A}\sim\mathcal{D}^k,\,X\sim\mathcal{D}}[\mathsf{F}(\mathbf{v},\mathsf{M}_r(\mathsf{G}_{\boldsymbol{A}}(X,\mathbf{v}))) \text{ is strongly-correlated with } X]\\
&= \Pr_{r\leftarrow\{0,1\}^m,\boldsymbol{A}\sim\mathcal{D}^k,\,X\sim\mathcal{D}}\left[(\mathsf{M}_r(\mathsf{G}_{\boldsymbol{A}}(X,\mathbf{v}))\cdot\mathbf{P}_{\mathbf{s}}^T)^{1,\ldots,d_0} \text{ is strongly-correlated with } X\right]\\
&\geq \beta/k.
\end{aligned}$$

■

Appendix E. Applications

Throughout this section, recall for $z\in\mathbb{R}$ we define $\mathsf{sign}(z) := \begin{cases}1 & z\geq 0\\ -1 & z<0\end{cases}$ and for $u=(u^1,\ldots,u^d)\in\mathbb{R}^d$ we define $\mathsf{sign}(u) := (\mathsf{sign}(u^1),\ldots,\mathsf{sign}(u^d))$.

E.1. Averaging

In this section, we prove Theorem 9.

Definition 46 ((λ,β)-Estimator for Averaging, Redefinition of Definition 7) *A mechanism* $\mathsf{M}\colon\mathbb{R}^+\times(\mathbb{R}^d)^n\to\mathbb{R}^d$ *is* (λ,β)-estimator for averaging *if given* $\gamma\geq 0$ *and* $X=(x_1,\ldots,x_n)\in\mathbb{R}^{n\times d}$ *with* $\max_{i,j\in[n]}\|x_i-x_j\|_2\leq\gamma$*, it holds that*

$$\Pr\left[\left\|\mathsf{M}(\gamma,X)-\frac{1}{n}\sum_{i=1}^n x_i\right\|_2\leq\lambda\gamma\right]\geq\beta.$$

Theorem 47 (Our averaging lower bound, Restatement of Theorem 9) *If* $\mathsf{M}\colon\mathbb{R}^+\times(\mathbb{R}^d)^n\to\mathbb{R}^d$ *is a* (λ,β)-estimator for averaging *for* $\lambda\geq 1$ *and* $\mathsf{M}(\gamma,\cdot)$ *is* $\left(1,\frac{\beta}{4n}\right)$*-DP for every* $\gamma\geq 0$*, then* $n\geq\Omega\left(\frac{\sqrt{d}/\lambda}{\log^{1.5}\left(\frac{d}{\beta\lambda}\right)}\right)$.

The proof of the theorem immediately follows by Theorem 4 and the following Claim 48 (along with post-processing of differential privacy).

Claim 48 *If* $\mathsf{M}\colon\mathbb{R}^+\times(\mathbb{R}^d)^n\to\mathbb{R}^d$ *is a* (λ,β)-estimator for averaging*, then the mechanism* $\widetilde{\mathsf{M}}\colon(\{-1,1\}^d)^n\to\{-1,1\}^d$ *defined by* $\widetilde{\mathsf{M}}(X) := \mathsf{sign}\Big(\mathsf{M}(\gamma=\sqrt{2\alpha d},\,X)\Big)$ *for* $\alpha=\frac{1}{40\lambda^2+1}$*, is* (α,β)*-weakly-accurate (Definition 3).*

Proof Fix $X=(x_1,\ldots,x_n)\in\{-1,1\}^{n\times d}$ with $\left|\mathcal{J}_X^1\right|,\left|\mathcal{J}_X^{-1}\right|\geq\frac{1}{2}(1-\alpha)d$ and let $\mu=(\mu^1,\ldots,\mu^d)=\frac{1}{n}\sum_{i=1}^n x_i$. Since $x_1,\ldots,x_n\in\{-1,1\}^d$ agree on at least $(1-\alpha)d$ coordinates, their diameter is bounded by $\gamma=\sqrt{2\alpha d}$. By the utility guarantee of M it holds that

$$\Pr_{q\sim\mathsf{M}(\gamma,X)}[\|q-\mu\|_2\leq\lambda\gamma]\geq\beta. \tag{9}$$

We prove the claim by showing that for any $q\in\mathbb{R}^d$ with $\|q-\mu\|_2\leq\lambda\gamma$ it holds that $\mathsf{sign}(q)$ strongly-agrees with X (Definition 2). Fix such q. Note that $\left|\{j\in[d]\colon\mathsf{sign}(q^j)\neq\mathsf{sign}(\mu^j)\}\right|\leq\|q-\mu\|_2^2\leq\lambda^2\gamma^2$. Since

$\mu^j = b$ for every $b \in \{-1,1\}$ and $j \in \mathcal{J}_X^b$, we deduce that $\left|\{j \in \mathcal{J}_X^b \colon \mathsf{sign}(q^j) \neq b\}\right| \leq \lambda^2\gamma^2$. Now note that for both $b \in \{-1,1\}$:

$$\lambda^2\gamma^2 = 2\alpha\lambda^2 d \leq 0.1 \cdot \frac{1}{2}(1-\alpha)d \leq 0.1 \cdot \left|\mathcal{J}_X^b\right|,$$

where the first inequality holds by our choice of α. Thus $\mathsf{sign}(q)$ strongly-agrees with X, as required. ■

E.2. Clustering

In this section, we prove an extension of Theorem 16 to (k,z)-clustering where we focus on the Euclidean metric space $(\mathbb{R}^d, d(x,y) := \|x-y\|_2)$. We start by extending the notations from section 1.1.3.

Let $\mathcal{B}_d := \{x \in \mathbb{R}^d \colon \|x\|_2 \leq 1\}$. For a database $S \in (\mathcal{B}_d)^n$, k centers $C = (c_1,\ldots,c_k) \in (\mathcal{B}_d)^k$ and a parameter $z \geq 1$, let

$$\mathrm{COST}_z(C;S) := \sum_{x \in S} \min_{i \in [k]} \|x - c_i\|_2^z \quad \text{and} \quad \mathrm{OPT}_{k,z}(S) := \min_{C \in (\mathbb{R}^d)^k} \mathrm{COST}_z(C;S).$$

Definition 49 ((λ,ξ,β)-Approximation Algorithm for (k,z)-Clustering, Redefinition of Definition 13) *We say that $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^k$ is an* (λ,ξ,β)-approximation algorithm for (k,z)-clustering, *if for every $S \in (\mathcal{B}_d)^n$ it holds that*

$$\Pr_{C \sim \mathsf{M}(S)}[\mathrm{COST}_z(C;S) \leq \lambda \cdot \mathrm{OPT}_{k,z}(S) + \xi] \geq \beta$$

Theorem 50 (Our Lower Bound, Extension of Theorem 16) *Let $n,d,k \in \mathbb{N}$, $\lambda, z \geq 1$ and $\xi \geq 0$ such that $k \geq 2$ and $n \geq k + 2 \cdot 40^{z/2}\xi$. If $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^k$ is an $(1, \frac{\beta}{4nk})$-DP (λ,ξ,β)-approximation algorithm for (k,z)-clustering, then either $k \geq 2^{\Omega(d/\lambda^{2/z})}\beta\lambda^{2/z}/d$ or $\xi \geq \Omega\left(\frac{2^{-O(z)}k\sqrt{d/\lambda^{2/z}}}{\log^{1.5}\left(\frac{kd}{\beta\lambda^{2/z}}\right)}\right)$.*

The following claim captures the main technical part in proving Theorem 50.

Claim 51 *Let $n,d,k \in \mathbb{N}$, $\lambda, z \geq 1$ and $\xi \geq 1$ such that $n \geq m$ for $m = k \cdot \left\lfloor (1 + 40^{z/2} \cdot 2\xi/k) \right\rfloor$. If $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^{k+1}$ is an $(\lambda,\ \xi,\ \beta)$-approximation algorithm for $(k+1,z)$-clustering, then the following mechanism $\widetilde{\mathsf{M}} \colon (\{-1,1\}^d)^m \to \{-1,1\}^d$ is $(k,\ \alpha = \frac{1}{160\cdot(2\lambda)^{2/z}},\ \frac{\beta}{k+1})$-weakly-accurate (Definition 5).*

An $(k, O(1/\lambda), \beta)$-weakly-accurate variant $\widetilde{\mathsf{M}}$ of the $(k+1)$-means (λ,ξ,β)-approximation algorithm M:

1. Input: $x_1,\ldots,x_m \in \{-1,1\}^d$ for $m = \left\lfloor (1 + 40^{z/2} \cdot 2\xi/k) \right\rfloor \cdot k$.
2. Operation:
   (a) Compute $(c_1,\ldots,c_{k+1}) = \mathsf{M}(\frac{1}{\sqrt{d}}x_1,\ldots,\frac{1}{\sqrt{d}}x_m, \underbrace{\mathbf{0},\ldots,\mathbf{0}}_{n-m \text{ times}})$ (where $\mathbf{0} = (\underbrace{0,\ldots,0}_{d \text{ times}})$).
   (b) Sample $j \leftarrow [k+1]$ and output $\mathsf{sign}(c_j)$.

We first prove Theorem 50 using Claim 51.

Proof [Proof of Theorem 50] Let $\mathsf{M} \colon (\mathcal{B}_d)^n \to (\mathcal{B}_d)^k$ be the mechanism from Theorem 50. By Claim 51, the mechanism $\widetilde{\mathsf{M}} \colon (\{-1,1\}^d)^m \to \{-1,1\}^d$ is $(k-1, \alpha = \frac{1}{160\cdot(2\lambda)^{2/z}}, \frac{\beta}{k})$-weakly-accurate. Furthermore,

by post-processing, $\widetilde{\mathsf{M}}$ is also $\left(1, \frac{\beta}{4nk}\right)$-DP, which in particular implies that it is $\left(1, \frac{\beta}{4n(k-1)}\right)$-DP. Hence we deduce by Theorem 6 that $m \geq \Omega\left(\frac{k\sqrt{d/\lambda^{2/z}}}{\log^{1.5}\left(\frac{dk}{\beta\lambda^{2/z}}\right)}\right)$. Since $m \geq \max\{k, 40^{z/2} \cdot 2\xi\}$, the above implies that either $k \geq \Omega\left(\frac{k\sqrt{d/\lambda^{2/z}}}{\log^{1.5}\left(\frac{dk}{\beta\lambda^{2/z}}\right)}\right)$ (which implies that $k \geq 2^{\Omega(d/\lambda^{2/z})}\beta\lambda^{2/z}/d$), or that $\xi \geq \Omega\left(\frac{2^{-O(z)}k\sqrt{d/\lambda^{2/z}}}{\log^{1.5}\left(\frac{dk}{\beta\lambda^{2/z}}\right)}\right)$. ■

We now prove Claim 51.

Proof [Proof of Claim 51] Fix $X = (x_1, \ldots, x_m) \in \{-1,1\}^{m\times d}$ such that for every $t \in [k]$ and every $b \in \{-1,1\}$ it holds that $\left|\mathcal{J}^b_{X_t}\right| \geq \frac{1}{2}(1-\alpha)d$ for $X_t = (x_{(t-1)m/k+1}, \ldots x_{tm/k})$, and let $S = (\tilde{x}_1, \ldots, \tilde{x}_m, \underbrace{\mathbf{0}, \ldots, \mathbf{0}}_{n-m \text{ times}})$ for $\tilde{x}_i = \frac{1}{\sqrt{d}}x_i$. We first argue that $\mathrm{OPT}_{k+1}(S)$ is small. For $t \in [k]$ define $c^*_t = \tilde{x}_{tm/k}$ and $c^*_{k+1} = \mathbf{0}$ (which covers all the $\mathbf{0}$'s with zero cost). By assumption, for every $t \in [k]$ and $\tilde{x}_i \in S$ with $x_i \in X_t$ it holds that

$$\|c^*_t - \tilde{x}_i\|_2^z = \left(\|c^*_t - \tilde{x}_i\|_2^2\right)^{z/2} = \left(\frac{1}{d} \cdot \left\|x_{tm/k} - x_i\right\|_2^2\right)^{z/2} \leq (4\alpha)^{z/2},$$

where the inequality holds since the number of indices that the vectors do not agree on is at most $d - \left|\mathcal{J}^1_{X_t}\right| - \left|\mathcal{J}^{-1}_{X_t}\right| \leq \alpha d$. Hence, we deduce that

$$\mathrm{OPT}_{k+1,z}(S) \leq \mathrm{COST}_z(c^*_1, \ldots, c^*_{k+1}; S) \leq (4\alpha)^{z/2}m.$$

Now, since

$$\begin{aligned}\lambda \cdot \mathrm{OPT}_{k+1,z}(S) + \xi &\leq \lambda \cdot (4\alpha)^{z/2}m + \frac{m}{2 \cdot 40^{z/2}} \\ &= \lambda \cdot \left(\frac{1}{40(2\lambda)^{2/z}}\right)^{z/2} + \frac{m}{2 \cdot 40^{z/2}} \\ &\leq m/40^{z/2},\end{aligned}$$

the utility guarantee of M implies that

$$\Pr_{C=(c_1,\ldots,c_{k+1})\sim\mathsf{M}(S)}\left[\mathrm{COST}_z(C;S) \leq m/40^{z/2}\right] \geq \beta.$$

We prove the claim by showing that for every $C = (c_1, \ldots, c_{k+1})$ with $\mathrm{COST}_z(C;S) \leq m/40^{z/2}$, there exists $s \in [k+1]$ and $t \in [k]$ such that $\mathsf{sign}(c_s)$ strongly-agrees with X_t. This will conclude the proof since the probability that the random j chosen in Step 2b of $\widetilde{\mathsf{M}}$ will hit the right s is $1/(k+1)$. Indeed if this is not the case, then for any center c_s and any non-zero point $\tilde{x}_i \in S$ where $x_i \in X_t$ we have at least one

$b \in \{-1,1\}$ such that $\left|\{j \in \mathcal{J}^b_{X_t} : \mathsf{sign}(\tilde{x}^j_i) \neq \mathsf{sign}(c^j_s)\}\right| > 0.1 \cdot \left|\mathcal{J}^b_{X_t}\right| > \frac{1}{20}(1-\alpha)d$, which yields that

$$\begin{aligned}
\|\tilde{x}_i - c_s\|_2^z &= (\|\tilde{x}_i - c_s\|_2^2)^{z/2} \\
&\geq \left(\frac{1}{d} \cdot \left|\{j \in [d] : \mathsf{sign}(\tilde{x}^j_i) \neq \mathsf{sign}(c^j_s)\}\right|\right)^{z/2} \\
&\geq \left(\frac{1}{d} \cdot \left|\{j \in \mathcal{J}^b_{X_t} : \mathsf{sign}(\tilde{x}^j_i) \neq \mathsf{sign}(c^j_s)\}\right|\right)^{z/2} \\
&> \left(\frac{1}{20}(1-\alpha)\right)^{z/2}
\end{aligned}$$

Thus $\mathrm{COST}_z(C;S) > \left(\frac{1}{20}(1-\alpha)\right)^{z/2} \cdot m > m/40^{z/2}$, a contradiction to the assumption $\mathrm{COST}_z(C;S) \leq m/40^{z/2}$. This concludes the proof of the claim.

■

E.3. Top Singular Vector

In this section, we prove Theorem 18. We start by recalling the setting.

For a matrix $X \in \mathbb{R}^{n\times d}$, the singular value decomposition of X is defined by $X = U\Sigma V^T$, where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{d\times d}$ are unitary matrices. The matrix $\Sigma \in \mathbb{R}^{n\times d}$ is a diagonal matrix with non-negative entries $\sigma_1 \geq \ldots \geq \sigma_{\min\{n,d\}} \geq 0$ along the diagonal, called the singular values of X. It is well known that $\|X\|_F^2 := \sum_{i\in[n],j\in[d]}(x^j_i)^2 = \sum_i \sigma_i^2$. The top (right) singular vector of X is defined by the first column of V (call it $v_1 \in \mathcal{S}_d$) which satisfy $\|X \cdot v_1\|_2 = \max_{v\in\mathcal{S}_d}\|X \cdot v\|_2$.

In the problem we consider, n rows $x_1, \ldots, x_n \in \mathcal{S}_d := \{v \in \mathbb{R}^d : \|v\|_2 = 1\}$ are given as input, and the goal is to estimate the top (right) singular vector of the $n \times d$ matrix $X = (x^j_i)_{i\in[n],j\in[d]}$.

Definition 52 ((λ,β)-Estimator of Top Singular Vector, Redefinition of Definition 17) *We say that $\mathsf{M} : [0,1] \times (\mathcal{S}_d)^n \to \mathcal{S}_d$ is an (λ,β)-estimator of top singular vector, if given an $n \times d$ matrix $X = (x_1, \ldots, x_n) \in (\mathcal{S}_d)^n$ and a number $\gamma \in [0,1]$ such that $\sigma_2(X) \leq \gamma \cdot \sigma_1(X)$ as inputs, outputs a column vector $y \in \mathcal{S}_d$ such that*

$$\Pr_{y\sim\mathsf{M}(\gamma,X)}\left[\|X \cdot y\|_2^2 \geq \|X \cdot v\|_2^2 - \lambda \cdot \gamma n\right] \geq \beta,$$

where v is the top (right) singular vector of X.

We now restate and prove Theorem 18.

Theorem 53 (Our lower bound, restatement of Theorem 18) *If $\mathsf{M} : [0,1] \times (\mathcal{S}_d)^n \to \mathcal{S}_d$ is a (λ,β)-*estimator of top singular vector *for $\lambda \geq 1$ and $\mathsf{M}(\gamma,\cdot)$ is $(1, \frac{\beta}{4n})$-DP for every $\gamma \in [0,1]$, then $n = \Omega\left(\frac{\sqrt{d}/\lambda}{\log\left(\frac{d}{\beta\lambda}\right)}\right)$.*

The proof of Theorem 53 immediately follows by Theorem 44 and the following claim.

Claim 54 *If $\mathsf{M} : [0,1] \times (\mathcal{S}_d)^n \to \mathcal{S}_d$ is an (λ,β)-*estimator for top singular vector *for $\lambda \geq 1$, then either the mechanism $\widetilde{\mathsf{M}} : \{-1,1\}^{n\times d} \to \{-1,1\}^d$ defined by $\widetilde{\mathsf{M}}(X) := \mathsf{sign}\left(\mathsf{M}(\gamma = \sqrt{\frac{2\alpha}{1-2\alpha}}, \frac{1}{\sqrt{d}}X)\right)$, for $\alpha = \frac{1}{4000\lambda^2}$, or the mechanism $\widetilde{\mathsf{M}}'(X) := -\widetilde{\mathsf{M}}(X)$, are $(\alpha,\beta/2)$-weakly-accurate (Definition 3).*

Proof Fix $X = (x_1, \ldots, x_n) \in \{-1,1\}^{n \times d}$ with $\left|\mathcal{J}_X^1\right|, \left|\mathcal{J}_X^{-1}\right| \geq \frac{1}{2}(1-\alpha)d$, and let $\tilde{X} = \frac{1}{\sqrt{d}}X \in (\mathcal{S}_d)^n$. Let $u = (u_1, \ldots, u_d) \in \mathcal{S}_d$ be a column vector with $u_j = \begin{cases} 1/\sqrt{d} & j \in \mathcal{J}_X^1 \\ -1/\sqrt{d} & \text{o.w.} \end{cases}$. By definition it holds that $\left\|\tilde{X} \cdot u\right\|_2^2 \geq \frac{n}{d}\left(2\left|\mathcal{J}_X^1\right| + 2\left|\mathcal{J}_X^{-1}\right| - d\right) \geq (1-2\alpha)n$. This yields that $\sigma_1(\tilde{X})^2 \geq (1-2\alpha)n$. Since $\sigma_1(\tilde{X})^2 + \sigma_2(\tilde{X})^2 \leq \left\|\tilde{X}\right\|_F^2 = n$, we deduce that $\sigma_2(\tilde{X})^2/\sigma_1(\tilde{X})^2 \leq \frac{2\alpha}{1-2\alpha} = \gamma^2$.

Therefore, by the utility guarantee of M, it holds that

$$\Pr_{y \sim \mathsf{M}(\gamma, \tilde{X})}\left[\left\|\tilde{X} \cdot y\right\|_2^2 \geq \left\|\tilde{X} \cdot u\right\|_2^2 - \lambda \cdot \gamma n\right] \geq \beta.$$

We prove the claim by showing that for any $y = (y_1, \ldots, y_d) \in \mathcal{S}_d$ with $\left\|\tilde{X} \cdot y\right\|_2^2 \geq \left\|\tilde{X} \cdot u\right\|_2^2 - \lambda \cdot \gamma n$, either $\mathsf{sign}(y)$ or $-\mathsf{sign}(y)$ strongly-agrees with X (Definition 2).

In the following, fix such y, and let $s = \sum_{j \in \mathcal{J}_X^1} y_j - \sum_{j \in \mathcal{J}_X^{-1}} y_j$. We next prove that

1. $\left\|\tilde{X} \cdot u\right\|_2^2 - \left\|\tilde{X} \cdot y\right\|_2^2 \geq \left(1 - \frac{2s^2}{d} - 4\alpha\right)n$, and
2. If $\mathsf{sign}(y)$ and $-\mathsf{sign}(y)$ do not strongly-agree with X, then $s^2 \leq \left(1 - \frac{1}{40}\right)d$.

The proof of the claim follows by items 1 and 2 since if $\mathsf{sign}(y)$ and $-\mathsf{sign}(y)$ do not strongly-agree with X, then by our choice of α it holds that

$$\left\|\tilde{X} \cdot u\right\|_2^2 - \left\|\tilde{X} \cdot y\right\|_2^2 \geq \left(\frac{1}{40} - 4\alpha\right)n > \lambda \underbrace{\sqrt{\frac{2\alpha}{1-2\alpha}}}_{\gamma} n,$$

which is a contradiction to the assumption about v.

We first prove item 1. Let $z = (z_1, \ldots, z_n) = \tilde{X} \cdot y$ and let $\mathcal{B} = [d] \setminus (\mathcal{J}_X^1 \cup \mathcal{J}_X^{-1})$. Note that for every $i \in [n]$:

$$|z_i| = \left|\frac{1}{\sqrt{d}}\sum_{j=1}^{d} x_i^j \cdot y_j\right| = \frac{1}{\sqrt{d}}\left|\sum_{j \in \mathcal{J}_X^1} y_j - \sum_{j \in \mathcal{J}_X^{-1}} y_j + \sum_{j \in \mathcal{B}} x_i^j \cdot y_j\right| \leq \frac{|s|}{\sqrt{d}} + \sqrt{\alpha},$$

where the last inequality holds since $\sum_{j \in \mathcal{B}} x_i^j \cdot y_j \leq \left\|x_i^{\mathcal{B}}\right\|_2 \cdot \|y_{\mathcal{B}}\|_2 \leq \sqrt{|\mathcal{B}|} \cdot 1 \leq \sqrt{\alpha d}$.

Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ we deduce that $\|z\|_2^2 = \sum_{i=1}^n z_i^2 \leq \left(\frac{2s^2}{d} + 2\alpha\right)n$. Hence

$$\left\|\tilde{X} \cdot u\right\|_2^2 - \|z\|_2^2 \geq (1-2\alpha)n - \left(\frac{2s^2}{d} + 2\alpha\right)n = \left(1 - \frac{2s^2}{d} - 4\alpha\right)n.$$

We next prove item 2. Assume $\mathsf{sign}(y)$ and $-\mathsf{sign}(y)$ do not strongly-agree with X. Then there exists $b, b' \in \{-1,1\}$ such that

$$\mathrm{E}_{j \leftarrow \mathcal{J}_X^b}[y_j \neq b] > 0.1 \tag{10}$$

and

$$\mathrm{E}_{j \leftarrow \mathcal{J}_X^{b'}}\left[y_j = b'\right] > 0.1. \tag{11}$$

We split into two cases:

Case 1: eqs. (10) and (11) holds for $b = b'$. In this case, we assume for simplicity that $b = b' = 1$ (the $b = b' = -1$ case holds by symmetry). Let $\mathcal{J}^+ = \{j \in \mathcal{J}_X^1 : y_j \geq 0\}$ and $\mathcal{J}^- = \{j \in \mathcal{J}_X^1 : y_j < 0\}$. Since $\mathsf{sign}(y_j) \neq 1 \implies y_j < 0$ and $\mathsf{sign}(y_j) = 1 \implies y_j \geq 0$, eqs. (10) and (11) yields that

$$\left|\mathcal{J}^+\right| \leq 0.9\left|\mathcal{J}_X^1\right| \quad \text{and} \quad \left|\mathcal{J}^-\right| \leq 0.9\left|\mathcal{J}_X^1\right|. \tag{12}$$

Now recall that $s = \sum_{j\in\mathcal{J}_X^1} y_j - \sum_{j\in\mathcal{J}_X^{-1}} y_j$. Therefore,

$$\begin{aligned} s &\leq \sum_{j\in\mathcal{J}^+} y_j - \sum_{j\in\mathcal{J}_X^{-1}} y_j \leq \sum_{j\in\mathcal{J}^+\cup\mathcal{J}_X^{-1}} |y_j| \\ &= \left\|y_{\mathcal{J}^+\cup\mathcal{J}_X^{-1}}\right\|_1 \leq \sqrt{\left|\mathcal{J}^+\cup\mathcal{J}_X^{-1}\right|}\cdot\left\|y_{\mathcal{J}^+\cup\mathcal{J}_X^{-1}}\right\|_2 \\ &\leq \sqrt{\left|\mathcal{J}^+\cup\mathcal{J}_X^{-1}\right|} \end{aligned} \tag{13}$$

where the third inequality holds since $\|w\|_1 \leq \sqrt{m}\cdot\|w\|_2$ for every $w \in \mathbb{R}^m$, and the fourth one holds since $\left\|y_{\mathcal{J}^+\cup\mathcal{J}_X^{-1}}\right\|_2 \leq \|y\|_2 = 1$.

By the right inequality in (12), a similar calculation to eq. (13) yields that

$$s \geq \sum_{j\in\mathcal{J}^-} y_j - \sum_{j\in\mathcal{J}_X^{-1}} y_j \geq -\sum_{j\in\mathcal{J}^-\cup\mathcal{J}_X^{-1}} |y_j| \geq \ldots \geq -\sqrt{\left|\mathcal{J}^-\cup\mathcal{J}_X^{-1}\right|}. \tag{14}$$

The proof in this case now follows by eqs. (13) and (14) since

$$\left|\mathcal{J}^+\cup\mathcal{J}_X^{-1}\right|,\ \left|\mathcal{J}^+\cup\mathcal{J}_X^{-1}\right| \leq d - 0.1\left|\mathcal{J}_X^1\right| \leq d - \frac{1}{20}(1-\alpha)d \leq \left(1-\frac{1}{40}\right)d,$$

where the first inequality holds by eq. (12).

Case 2: eqs. (10) and (11) holds for $b \neq b'$. In this case, we assume for simplicity that $b = 1$ and $b' = -1$ (the $b = -1$ and $b' = 1$ case holds by symmetry). Let $\mathcal{J}^1 = \{j \in \mathcal{J}_X^1 : y_j \geq 0\}$ and $\mathcal{J}^{-1} = \{j \in \mathcal{J}_X^{-1} : y_j \geq 0\}$. Since $\mathsf{sign}(y_j) \neq 1 \implies y_j < 0$, eqs. (10) and (11) yield that

$$\left|\mathcal{J}^1\right| \leq 0.9\left|\mathcal{J}_X^1\right| \quad \text{and} \quad \left|\mathcal{J}^{-1}\right| \leq 0.9\left|\mathcal{J}_X^{-1}\right| \tag{15}$$

The left inequality in eq. (15) implies that

$$s \leq \sum_{j\in\mathcal{J}^1} y_j - \sum_{j\in\mathcal{J}_X^{-1}} y_j \leq \sum_{j\in\mathcal{J}^1\cup\mathcal{J}_X^{-1}} |y_j| \leq \sqrt{\left|\mathcal{J}^1\cup\mathcal{J}_X^{-1}\right|} \tag{16}$$

where the last inequality holds similarly to eq. (13). Similarly, the right inequality in eq. (15) implies that

$$s \geq \sum_{j\in\mathcal{J}_X^1} y_j - \sum_{j\in\mathcal{J}^{-1}} y_j \geq -\sum_{\mathcal{J}_X^1\cup\mathcal{J}^{-1}} |y_j| \geq -\sqrt{\left|\mathcal{J}_X^1\cup\mathcal{J}^{-1}\right|}. \tag{17}$$

The proof now follows by eqs. (16) and (17) since $\left|\mathcal{J}^1\cup\mathcal{J}_X^{-1}\right| \leq d - 0.1\left|\mathcal{J}_X^1\right| \leq \left(1-\frac{1}{40}\right)d$ and $\left|\mathcal{J}_X^1\cup\mathcal{J}^{-1}\right| \leq d - 0.1\left|\mathcal{J}_X^{-1}\right| \leq \left(1-\frac{1}{40}\right)d$.

■

Appendix F. Fingerprinting Codes

An (n, d)-fingerprinting code, where $[n]$ is the set of users and d is the code length, is a pair $(\mathsf{Gen}, \mathsf{Trace})$. The first element Gen is a randomized algorithm that generates and outputs a codebook $X \in \{-1, 1\}^{n \times d}$, where for every $i \in [n]$, the i^{th} row $x_i \in \{-1, 1\}^d$ of X is the code of user i. The second element Trace is a randomized algorithm that receives as an input a codebook $X \in \{-1, 1\}^{n \times d}$ and a word $q \in \{-1, 1\}^d$ that was (possibly) generated by a malicious subset of users $\mathcal{S} \subseteq [n]$ (from their codes), and outputs a user $i \in [n]$ or $\perp$.

The security requirements of fingerprinting codes says that (1) given every code x_i, one can be verify that user i holds it (that is, it is known that user i is the one to receive the code x_i), and that (2) given a codebook X and a word q generated by a malicious subset of users $\mathcal{S}$ using their codes, the algorithm Trace returns a user $i \in \mathcal{S}$ with high probability (so, it returns a user $i \notin \mathcal{S}$ or $\perp$ with low probability).

The basic assumption of fingerprinting code is that for every $j \in [d]$, the j^{th} element of q must be equal to the j^{th} element of x_i for some $i \in \mathcal{S}$. In particular, if the j^{th} element of all the codes of $\mathcal{S}$ is equal to some $b \in \{-1, 1\}$, then the j^{th} element of the word q must also be b. We say that such j's are *b-marked in* $(X, \mathcal{S})$.

First, the set of all feasible words $q \in \{-1, 1\}^d$ for a codebook $X \in \{-1, 1\}^{n \times d}$ and a malicious subset $\mathcal{S} \subseteq [n]$ is denoted as

$$F(X, \mathcal{S}) := \{q \in \{-1, 1\}^d \mid \forall j \in [d], \exists i \in \mathcal{S} : \ q^j = x_i^j\}.$$

Now, we are ready to formally define fingerprinting codes, similarly to Bun et al. (2014).

Definition 55 (Fingerprinting Codes) *For every $n, d \in \mathbb{N}$ and $\beta \in [0, 1]$, a pair of algorithms $(\mathsf{Gen}, \mathsf{Trace})$ is an (n, d)-*fingerprinting code with security β *if Gen outputs a codebook $\mathbf{X} \in \{-1, 1\}^{n \times d}$, and for every subset $\mathcal{S} \subseteq [n]$ and every (possibly randomized) adversary $\mathsf{A} \colon \{-1, 1\}^{|\mathcal{S}| \times d} \to \{-1, 1\}^d$ such that $\mathbf{q} \sim \mathsf{A}(\mathbf{X}_{\mathcal{S}})$, it holds that*

- $\Pr[\mathbf{q} \in F(\mathbf{X}, \mathcal{S}) \wedge \mathsf{Trace}(\mathbf{X}, \mathbf{q}) = \perp] \leq \beta$, *and*
- $\Pr[\mathsf{Trace}(\mathbf{X}, \mathbf{q}) \in [n] \setminus \mathcal{S}] \leq \beta$,

where the probability is over the randomness of Gen, Trace and A. Moreover, the algorithm Gen can share an additional output $z \in \{-1, 1\}^*$ with the algorithm Trace.

Application: A Simple Fingerprinting Code. The following algorithm Gen generates a fingerprinting code of length d as in Lemma 34 for n users.

Algorithm Gen

1. Input: Number of users $n \in \mathbb{N}$ and a confidence parameter $\beta \in [0, 1]$.
2. Operation:
   (a) Let $d \geq \Theta(n^2 \log^2(n) \log(1/\beta))$ be the length of the code.
   (b) Sample $(x_1, \ldots, x_n, z) \sim \mathcal{D}'(n, d)$ (Definition 33).
   (c) Output $X = (x_i^j)_{i \in [n], j \in [d]}$ and share $z = (z^1, \ldots, z^d)$ with Trace.

The next tracing algorithm Trace receives a codebook of n codes of length d (namely, a fingerprinting code as generated by the algorithm Gen), a word of length d, possibly generated by a subset of malicious users using their codes, and a shared state of length d shared by the algorithm Gen. It outputs a user from the malicious subset (if there is such user) or report that there is no such user, with high probability.

Algorithm Trace

1. Input: A codebook $X \in \{-1,1\}^{n\times d}$, a word $q \in \{-1,1\}^d$, and a shared state $z \in \{-1,1\}^d$.
2. Operation:
   (a) For every $i \in [n]$: If $\langle x_i, q\rangle - \langle z, q\rangle > \frac{0.2d}{n\cdot\ln(5n)}$, output i.
   (b) Else (no such $i \in [n]$), output $\perp$.

Next, we prove that the pair of algorithms $(\mathsf{Gen}, \mathsf{Trace})$ is a fingerprinting code.

Claim 56 *For $n \in \mathbb{N}$, $\beta \in [0,1]$, and $d \geq \Theta(n^2 \log^2(n)\log(1/\beta))$, the pair $(\mathsf{Gen}, \mathsf{Trace})$ is an (n,d)-fingerprinting code with security β according to Definition 55.*

Proof We start by denoting $\mathbf{X} \in \{-1,1\}^{n\times d}$ as the output of Gen and $\mathbf{z} \in \{-1,1\}^d$ as the state shared by Gen with Trace. We want to show that for every subset $\mathcal{S} \subseteq [n]$ and every adversary $\mathsf{A}\colon \{-1,1\}^{|\mathcal{S}|\times d} \to \{-1,1\}^d$ such that $\mathbf{q} \sim \mathsf{A}(\mathbf{X}_\mathcal{S})$ is generated by A,

$$\Pr[\mathbf{q} \in F(\mathbf{X},\mathcal{S}) \wedge \mathsf{Trace}(\mathbf{X},\mathbf{q}) = \perp] \leq \beta \quad \text{and} \quad \Pr[\mathsf{Trace}(\mathbf{X},\mathbf{q}) \in [n]\setminus\mathcal{S}] \leq \beta.$$

To prove the first item, define the random variable $\mathbf{q}'$ which is equal to $\mathbf{q}$ if $\mathbf{q} \in F(\mathbf{X},\mathcal{S})$ and otherwise takes a value in $F(\mathbf{X},\mathcal{S})$ (e.g., the first lexicographic vector there). Furthermore, consider an algorithm $\mathsf{A}'\colon \{-1,1\}^{|\mathcal{S}|\times d} \to \{-1,1\}^d$ that given $X_\mathcal{S}$, samples $q \sim \mathsf{A}(X_\mathcal{S})$ and checks if $q \in F(X,\mathcal{S})$ (i.e., perfectly agrees with all the marked columns of $X_\mathcal{S}$). If it does, it outputs q. Otherwise, it outputs the first vector in $F(X,\mathcal{S})$ (which again can be done only using $\mathbf{X}_\mathcal{S}$). Observe that by definition, A' is strongly-accurate (Definition 32) and also $\mathbf{q}'$ has the same distribution as $\mathsf{A}'(\mathbf{X}_\mathcal{S})$. Hence

$$\begin{aligned}
&\Pr[\mathbf{q} \in F(\mathbf{X},\mathcal{S}) \wedge \mathsf{Trace}(\mathbf{X},\mathbf{q}) = \perp] \\
&= \Pr\left[\mathbf{q} \in F(\mathbf{X},\mathcal{S}) \wedge \mathsf{Trace}(\mathbf{X},\mathbf{q}) = \perp \middle|\, \mathbf{q} = \mathbf{q}'\right] \cdot \Pr\left[\mathbf{q} = \mathbf{q}'\right] + 0 \cdot \Pr\left[\mathbf{q} \neq \mathbf{q}'\right] \\
&= \Pr\left[\mathsf{Trace}(\mathbf{X},\mathbf{q}') = \perp \middle|\, \mathbf{q} = \mathbf{q}'\right] \cdot \Pr\left[\mathbf{q} = \mathbf{q}'\right] \\
&\leq \Pr\left[\mathsf{Trace}(\mathbf{X},\mathbf{q}') = \perp\right] = \Pr\left[\mathsf{Trace}(\mathbf{X},\mathsf{A}'(\mathbf{X}_\mathcal{S})) = \perp\right] \leq \beta,
\end{aligned}$$

where the last inequality holds by Lemma 34.

To prove the second item, fix $i \in [n]\setminus\mathcal{S}$. Let $\mathbf{p}^1,\ldots,\mathbf{p}^d$ be the expectations that were sampled in the process of sampling $\mathbf{X},\mathbf{z}$ (Definition 33), and in the following assume they were fixed to some values $(p^1,\ldots,p^d)$. Recall that $\mathbf{q}$ is a function of $\mathbf{X}_\mathcal{S}$. Therefore, $\mathbf{q}$ is independent of $\mathbf{x}_i$. The same calculation done in eq. (6) yields that

$$\Pr\left[\langle \mathbf{x}_i, \mathbf{q}\rangle - \langle \mathbf{z}, \mathbf{q}\rangle > \frac{0.2d}{n\ln(5n)}\right] \leq \frac{\beta}{20n}.$$

The proof of the second item is now concluded by the union bound over $j \in [n]\setminus\mathcal{S}$.

■

Remark 57 *In a* robust fingerprinting code*, we use a relaxed assumption for the set $F(X,\mathcal{S})$, as in Bun et al. (2014), and only require that the j^{th} element of q is equal to the j^{th} element of x_i for some $i \in \mathcal{S}$* with high probability. *It is possible to prove that $(\mathsf{Gen}, \mathsf{Trace})$ is also a robust fingerprinting code. We can easily achieve robustness if we require the same fraction of mistakes in both b-marked sets $\mathcal{J}^b_{X_\mathcal{S}}$, but robustness can be achieved anyway since the number of 1-marked columns and the number of (-1)-marked columns are almost the same, and there are many such columns (a simple calculation yields that $\Omega(1/\log n)$ fraction of the columns are marked).*