The Power of Factorization Mechanisms in Local and Central Differential Privacy

Alexander Edmonds
University of Toronto
Toronto, Canada
edmonds@cs.toronto.edu

Aleksandar Nikolov
University of Toronto
Toronto, Canada
anikolov@cs.toronto.edu

Jonathan Ullman
Northeastern University
Boston, USA
jullman@ccs.neu.edu

ABSTRACT

We give new characterizations of the sample complexity of answering linear queries (statistical queries) in the local and central models of differential privacy: (1) In the non-interactive local model, we give the first approximate characterization of the sample complexity. Informally our bounds are tight to within polylogarithmic factors in the number of queries and desired accuracy. Our characterization extends to agnostic learning in the local model. (2) In the central model, we give a characterization of the sample complexity in the high-accuracy regime that is analogous to that of Nikolov, Talwar, and Zhang (STOC 2013), but is both quantitatively tighter and has a dramatically simpler proof.

Our lower bounds apply equally to the empirical and population estimation problems. In both cases, our characterizations show that a particular factorization mechanism is approximately optimal, and the optimal sample complexity is bounded from above and below by well studied factorization norms of a matrix associated with the queries.

CCS CONCEPTS

• **Theory of computation → Design and analysis of algorithms**.

KEYWORDS

Differential privacy, local differential privacy, matrix factorization, matrix mechanism, factorization mechanism, statistical queries, PAC learning.

ACM Reference Format:
Alexander Edmonds, Aleksandar Nikolov, and Jonathan Ullman. 2020. The Power of Factorization Mechanisms in Local and Central Differential Privacy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC '20), June 22–26, 2020, Chicago, IL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3357713.3384297

1 INTRODUCTION

Differential privacy [DMNS06] is a rigorous mathematical framework for protecting individual privacy that is well suited to statistical data analysis. In addition to a rich academic literature, differential privacy is now being deployed on a large scale by Apple [App17], Google [EPK14, BEM⁺17, WZL+19], Uber [JNS18], and the US Census Bureau [DLS+17].

To compute statistics of the data with differential privacy—or any notion of privacy—we have to inject noise into the computation of these statistics [DN03]. The amount of noise is highly dependent on the particular statistic, and thus a central problem in differential privacy is to determine how much error is necessary to compute a given statistic.

In this work we consider the class of *linear queries* (also called *statistical queries* [Kea93]). The simplest example of a linear query is "What fraction of individuals in the data have property P?" Workloads of linear queries capture a variety of statistical tasks: computing histograms and PDFs, answering range queries and computing CDFs, estimating the mean, computing correlations and higher-order marginals, and estimating the risk of a classifier.

The power of differentially private algorithms for answering a *worst-case* workload of linear queries is well understood [BUV14], and known bounds are essentially tight as a function of the dataset size, the data domain, and the size of the workload. However, many workloads, such as those corresponding to computing PDFs or CDFs, have additional structure that makes it possible to answer them with less error than these worst-case workloads. Thus, a central question is

> *Can we characterize the amount of error required to estimate a given workload of linear queries subject to differential privacy in terms of natural properties of the workload, and can we achieve this error via computationally efficient algorithms?*

In the central model, there has been dramatic progress on this question [HT10, BDKT12, NTZ16, Nik15, BBNS19], giving approximate characterizations for every workload of linear queries. We extend this line of work in two ways:

(1) We give the first approximate characterization for *non-interactive local differential privacy* [DMNS06, KLN+08]. This result is also much sharper than analogous results for the central model of differential privacy.
(2) We give a new approximate characterization for the *central model of differential privacy* in the high-accuracy regime (equivalently, in the large-dataset regime). This characterization is analogous to a result of [NTZ16], but it is quantitatively tighter and its proof is dramatically simpler. For ℓ_2^2 error, our characterization is tight up to a constant factor.

In particular, our results show that a natural and well studied type of *factorization mechanism* is approximately optimal in these settings. Factorization mechanisms capture a number of special-purpose mechanisms from the theory literature [BCD+07, DNPR10,

STOC '20, June 22–26, 2020, Chicago, IL, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6979-4/20/06…$15.00
https://doi.org/10.1145/3357713.3384297

CSS11, TUV12, CTUW14], were involved in previous characterizations, and also roughly capture the *matrix mechanisms* [LHR+10, MMHM18] from the databases literature, which have been developed into practical algorithms for US Census Data.[1]

Our characterization in the local model extends to agnostic PAC learning, and shows that the optimal learner for any family of queries is to use the optimal factorization mechanism to estimate the error of every concept. Our characterization is sharper than the previous characterization of [KLN+08], which loses polynomial factors in the SQ dimension [BFJ+94].

1.1 Background: Linear Queries and Factorization Mechanisms

We start by briefly introducing the relevant concepts and definitions necessary to state our results. See Section 2 for a more thorough treatment of the necessary background.

Linear Queries. Suppose we are given a *dataset* $X = (x_1, \ldots, x_n) \in \mathcal{X}^n$, where each entry x_i is the data of one individual and $\mathcal{X}$ is some *data universe*. We will treat the size of the dataset n as public information. A linear query is specified by a bounded function $q : \mathcal{X} \to \mathbb{R}$ and (abusing notation) its answer is $q(X) = \frac{1}{n}\sum_{i=1}^{n} q(x_i)$. A *workload* is a set of linear queries $Q = \{q_1, \ldots, q_k\}$, and we use $Q(X) = (q_1(X), \ldots, q_k(X))$ to denote the answers.

Given a workload of queries, we can associate a *workload matrix* $W \in \mathbb{R}^{Q \times \mathcal{X}}$, defined by $W_{q,x} = q(x)$. The convention of calling the above queries "linear" stems from the fact that they can be written as the product of the workload matrix with the histogram vector of the dataset. As such, we will sometimes use Q and W interchangeably.

Error and Sample Complexity. Our goal is to design an (ε, δ)-differentially private mechanism $\mathcal{M}$ that takes a dataset X and accurately estimates $Q(X)$ for an appropriate measure of accuracy. In this work we primarily consider accuracy in the ℓ_∞ norm, and define

$$\mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n) = \max_{X \in \mathcal{X}^n} \mathop{\mathbb{E}}_{\mathcal{M}}[\|\mathcal{M}(X) - Q(X)\|_\infty],$$
$$\mathrm{err}^{\ell_\infty}_{\varepsilon,\delta}(Q, n) = \min_{(\varepsilon,\delta)\text{-DP } \mathcal{M}} \mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n).$$

Privacy becomes easier to achieve as the dataset size n grows. We are interested in the *sample complexity*, which is the smallest size of dataset on which it is possible to achieve a specified error α for given privacy parameters ε and δ:

$$\mathrm{sc}^{\ell_\infty}_{\varepsilon,\delta}(Q, \alpha) = \min\left\{n : \mathrm{err}^{\ell_\infty}_{\varepsilon,\delta}(Q, n) \le \alpha\right\}.$$

The Approximate Factorization Mechanisms. One of the most basic tools in the central-model of differential privacy is the *Gaussian mechanism* (see e.g. [DR14]). This mechanism computes the vector of answers to the queries $Q(X)$ and perturbs it with spherical Gaussian noise scaled to the ℓ_2*-sensitivity* of the workload. In particular, the sample complexity of this mechanism is

$$O\left(\frac{\|W\|_{1\to 2}\sqrt{\log(1/\delta)\log k}}{\varepsilon\alpha}\right).$$

where $\|W\|_{1\to 2}$ denotes the largest ℓ_2 norm of any column of W, which is the ℓ_2-sensitivity.

One can try to improve this mechanism by replacing W with a simpler workload of queries A, and then attempting to reconstruct the answer to W by applying a linear transform R such that $W = RA$. One can show that the overall mechanism has error $\|R\|_{2\to\infty}\|A\|_{1\to 2}$, where $\|R\|_{2\to\infty}$ denotes the maximum ℓ_2 norm of any row of R. This quantity can be dramatically smaller than $\|W\|_{1\to\infty}$, for example if W contains many copies of the same query.

The *factorization mechanism* chooses the optimal factorization $W = RA$, giving error proportional to the *factorization norm*

$$\gamma_2(W) = \min\{\|R\|_{2\to\infty}\|A\|_{1\to 2} : W = RA\}.$$

The sample complexity of this mechanism is thus

$$\mathrm{sc}^{\ell_\infty}(\mathcal{M}_{\gamma_2}, Q, \alpha) = O\left(\frac{\gamma_2(W)\sqrt{\log(1/\delta)\log|Q|}}{\varepsilon\alpha}\right).$$

We note that that the factorization norm $\gamma_2(W)$ and an optimal factorization $W = RA$ can be computed in time polynomial in the size of W via semidefinite programming [LS09].

Finally, we can try to further improve the mechanism using an *approximate factorization mechanism* that approximates the workload W with a simpler workload $\widetilde{W}$ that is entrywise close to W, and applying the factorization mechanism to $\widetilde{W}$. The error of this mechanism is proportional to the *approximate factorization norm*

$$\gamma_2(W, \alpha) = \min\{\gamma_2(\widetilde{W}) : \|W - \widetilde{W}\|_{1\to\infty} \le \alpha/2\},$$

where $\|W - \widetilde{W}\|_{1\to\infty}$ is the maximum absolute difference between entries of W and $\widetilde{W}$. The sample complexity of this mechanism is thus

$$\mathrm{sc}^{\ell_\infty}(\mathcal{M}_{\gamma_2,\alpha}, Q, \alpha) = O\left(\frac{\gamma_2(W, \alpha/2)\sqrt{\log(1/\delta)\log|Q|}}{\varepsilon\alpha}\right).$$

The Local Model. Although we have discussed the factorization mechanism in the context of central differential privacy, these ideas can all be adapted to *(non-interactive) local differential privacy*. In this model, each user will apply a separate (ε, δ)-differentially private mechanism $\mathcal{M}_1, \ldots, \mathcal{M}_n$ to their own data, and the output can then be postprocessed using an arbitrary algorithm $\mathcal{A}$, so the mechanism can be expressed as

$$\mathcal{M}(X) = \mathcal{A}(\mathcal{M}_1(X_1), \ldots, \mathcal{M}_n(X_n))$$

We define $\mathrm{err}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,\delta}$, and $\mathrm{sc}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,\delta}$ analogously to the central model, but with the minimum taken over mechanisms that are (ε, δ)-DP in the local model.

Since the queries are linear, we can simply have each user apply the approximate factorization mechanism to their own data and average the results. One can show that randomizing each individual's data independently increases the variance of the noise by a factor of $\sqrt{n}$ compared to the central model version of the mechanism. One can also achieve $(\varepsilon, 0)$-differential privacy by replacing Gaussian

[1] In a nutshell, the matrix mechanism is a particular factorization mechanism designed for the special case of ℓ_2^2 error, and combined with various optimizations and post-processing techniques to improve computational efficiency and utility. Usually the matrix mechanism is presented in the special case of pure differential privacy.

noise with a different subgaussian noise distribution. Putting it together, the resulting sample complexity becomes

$$\mathrm{sc}^{\ell_\infty}(\mathcal{M}^{\mathrm{loc}}_{\gamma_2,\alpha}, Q, \alpha) = O\left(\frac{\gamma_2(W, \alpha/2)^2 \log |Q|}{\varepsilon^2\alpha^2}\right). \tag{1}$$

1.2 Our Results

1.2.1 Linear Queries in the Local Model. Our main result in the local model shows that the approximate factorization mechanism described above is approximately optimal among all non-interactive locally differentially private mechanisms.

Theorem 1 (Informal). *Let $\alpha, \varepsilon, \delta > 0$ be smaller than some absolute constants and let Q be a workload of linear queries with workload matrix W. Then, for some $\alpha' = \Omega(\alpha/\log(1/\alpha))$,*

$$\mathrm{sc}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,0}(Q, \alpha') = \Omega\left(\frac{\gamma_2(W, \alpha/2)^2}{\varepsilon^2\alpha^2}\right).$$

To interpret the theorem, it helps to start by imagining that $\gamma_2(W, \alpha'/2) = \gamma_2(W, \alpha/2)$, in which case the theorem would show that the sample complexity of answering queries up to error α' is

$$\Omega\left(\frac{\gamma_2(W, \alpha'/2)^2}{\varepsilon^2\alpha^2}\right),$$

which differs from the sample complexity of the local approximate factorization mechanism, given in (1), by a factor of just $O(\log(1/\alpha')^2 \log |Q|)$. The fact that we take $\alpha' < \alpha$ means that $\gamma_2(W, \alpha/2)$ can be much smaller than $\gamma_2(W, \alpha'/2)$.[2] Nevertheless, for many natural families of queries and choices of α, $\gamma_2(W, \alpha/2)$ will be relatively stable to small changes in α, in which case our lower bound will be tight up to this $O(\log(1/\alpha)^2 \log |Q|)$ factor. In contrast, existing characterizations for the central model [HT10, BDKT12, NTZ16, Nik15, BBNS19] lose a poly$(1/\alpha)$ factor, or else they lose a polylog$|\mathcal{X}|$ factor that is typically large.

Remark 2. *Our proof of Theorem 1, in fact, shows that the lower bound holds in the distributional setting where X is sampled i.i.d. from an unknown distribution μ, and the goal is to estimate the quantity $q(\mu) = \mathbb{E}_{x\sim\mu}[q(x)]$ for every query $q \in Q$ up to error at most α.*

Remark 3. *Theorem 1 crucially assumes that the error is bounded in the ℓ_∞ metric. If we consider the less stringent ℓ_2^2 error metric (appropriately scaled to reflect the error per query), then one can achieve sample complexity $O(\log |\mathcal{X}|/\varepsilon^2\alpha^4)$ for any workload of queries [BBNS19], which can be exponentially smaller than the lower bound we prove for ℓ_∞ error. In many applications, such as releasing the PDF, CDF, or marginals of the data, the ℓ_∞ error metric is standard in the literature on these problems, and is more practical, since, for natural datasets, the weaker ℓ_2^2 guarantee can be achieved by mechanisms that ignore the data.*

Using Theorem 1, we obtain new lower bounds for three well studied families of queries:

(1) *Threshold queries,* which are also known as range queries, and equivalent to computing the CDF of the data.
(2) *Parity queries,* which capture the covariance and higher-order moments of the data.
(3) *Marginal queries,* also known as conjunctions, which capture the marginal distribution on subsets of the attributes.

Corollary 4 (Thresholds / CDFs). *Let Q^{cdf}_T be the family of statistical queries over the domain $\mathcal{X} = [T]$ that, for every $1 \le t \le T$, contains the statistical query $q_t(x) = \mathbb{I}\{x \le t\}$. Then for every $T \in \mathbb{N}$ and ε, α smaller than an absolute constant,*

$$\mathrm{sc}^{\ell_\infty,loc}_{\varepsilon,0}(Q^{cdf}_T, \alpha) = \Omega\left(\log^2 T\right).$$

We obtain this corollary by combining Theorem 1 with results from [FSSS03]. Corollary 4 should be compared to the upper bound of $O(\log^3 T)$ that can be obtained from the local analogue of the *binary tree mechanism* [DNPR10, CSS11]. Ours is the first lower bound to go beyond the easy $\Omega(\log T)$ lower bound for this problem, which follows easily via a so-called packing argument.

Corollary 5 (Parities). *Let $Q^{parity}_{d,w}$ be the family of statistical queries over the domain $\mathcal{X} = \{\pm 1\}^d$ that, for every $S \subseteq [d], |S| \le w$, contains the statistical query $q_S(x) = \prod_{j\in S} x_j$. Then for every $k \le d \in \mathbb{N}$ and ε, α smaller than an absolute constant,*

$$\mathrm{sc}^{\ell_\infty,loc}_{\varepsilon,0}(Q^{parity}_{d,w}, \alpha) = \Omega((d/w)^w).$$

Corollary 5 says that adding independent Gaussian noise to each query is optimal up to a $O(w \log(d/w))$ factor. Using similar techniques, one can also obtain a direct proof that gives a tight lower bound up to constant factors, even for the simpler problem of finding the subset S of size at most w that maximizes $q_S(X)$.

Corollary 6 (Marginals). *Let $Q^{marginal}_{d,w}$ be the family of statistical queries over the domain $\mathcal{X} = \{0, 1\}^d$ that, for every $S \subseteq [d], |S| \le w$, contains the statistical query $q_S(x) = \prod_{j\in S} x_j$. Then for every $k \le d \in \mathbb{N}$ and ε, α smaller than an absolute constant,*

$$\mathrm{sc}^{\ell_\infty,loc}_{\varepsilon,0}(Q^{marginal}_{d,w}, \alpha) = (d/w)^{\Omega(\sqrt{w})}.$$

Marginal queries have been extremely well studied in differential privacy [BCD+07, KRSU10, GHRU11, HRS12, TUV12, CTUW14, DNT15]. Corollary 6 shows that a natural local analogue of the algorithm of [TUV12] is optimal for answering marginal queries up to the hidden constant factor in the exponent.

1.2.2 Agnostic Learning in the Local Model. Theorem 1 extends to characterizing *agnostic PAC learning* [KSS94] in the local model. In agnostic PAC learning, the dataset consists of labeled examples $X = ((x_1, y_1), \ldots, (x_n, y_n))$, where $x_i \in \mathcal{X}$, and $y_i \in \{\pm 1\}$, and each pair (x_i, y_i) is sampled independently from an unknown distribution μ. The goal is to find a concept $c : \mathcal{X} \to \{\pm 1\}$ in a concept class C that approximately maximizes $\mathbb{E}_{(x,y)\sim\mu}[c(x)y]$.

The correlation of each concept c with the labels in the data is a linear query, and one natural approach to agnostic PAC learning is to estimate all these linear queries, and output the concept that corresponds to the largest query value. Thus, we can apply the local approximate factorization mechanism to the family of queries C to obtain the same sample complexity upper bound in (1). Interestingly, the proof of our lower bound in Theorem 1 shows that the same lower bound also applies to this a priori easier problem of

[2] For example, if every entry of W is at most α in absolute value, then $\gamma_2(W, \alpha) = 0$ whereas $\gamma_2(W, \alpha')$ can be arbitrarily large for $\alpha' < \alpha$, but this behavior typically does not happen for "non-trivial" values of α.

agnostic PAC learning, showing that the local approximate factorization mechanism gives an approximately optimal way to learn any concept class C.

Prior results of Kasisiviswanathan et al. [KLN+08] connecting learning algorithms in the local model with the SQ model, together with characterizations of sample complexity in the SQ model [BFJ+94, Szö09], give upper and lower bounds on sample complexity of learning in the local model in terms of SQ dimension. These results, however, are only tight up to polynomial factors in the SQ dimension—which can be polynomial in $|C|$—whereas our results are sharper. We remark that, technically, the results are not comparable, since the the characterization via the SQ model holds for sequentially interactive, rather than non-interactive, mechanisms.

1.2.3 *Linear Queries in the Central Model.*

Our second set of results quantitatively strengthens—and simplifies the proof of—the central model characterization of [NTZ16]. In contrast to the local model, the sample complexity of answering many natural workloads of linear queries exhibits two distinct regimes, depending on the desired accuracy. For example, for a worst-case workload of linear queries, the sample complexity is at most

$$\min\left\{\frac{\log^{1/2}|\mathcal{X}|\log|Q|}{\varepsilon\alpha^2}, \frac{|Q|^{1/2}}{\varepsilon\alpha}\right\}.$$

Thus, the sample complexity behaves very differently when α goes below some critical value. Our results concern this *high-accuracy regime* where α is quite small. In these results, we consider the ℓ_2^2 error (scaled to be directly comparable to the ℓ_∞ error), which is

$$\mathrm{err}^{\ell_2^2}(\mathcal{M}, Q, n) = \max_{X\in\mathcal{X}^n} \mathop{\mathbb{E}}_{\mathcal{M}}\left[\tfrac{1}{|Q|}\|\mathcal{M}(X) - Q(X)\|_2^2\right]^{1/2}$$

with the related quantities defined analogously. Notice that we have scaled the ℓ_2^2 error so that $\mathrm{err}^{\ell_2^2}(\mathcal{M}, Q, n) \le \mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n)$. For ℓ_2^2 error, the natural factorization norm that describes the error of the factorization mechanisms is

$$\gamma_F(W) = \left\{\tfrac{1}{|Q|^{1/2}}\|R\|_F\|A\|_{1\to 2} : W = RA\right\},$$

where $\|R\|_F = \sqrt{\sum_{i,j} R_{i,j}^2}$ is the Frobenius norm of R.

In this high-accuracy regime, a combination of [NTZ16] and [NT15] (see also the thesis [Nik14]) shows that, for every workload of linear queries, there is some α^* such that

$$\forall\alpha \le \alpha^* \;\; \Omega(\log^{-1}|Q|)\cdot\frac{\gamma_F(W)}{\varepsilon\alpha} \le \mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q,\alpha) \le O(1)\cdot\frac{\gamma_F(W)}{\varepsilon\alpha}\cdot\log(1/\delta).$$

Note that the upper and lower bound differ by a factor of $O(\log|Q|\cdot\log(1/\delta))$. The upper bound above is precisely what is given by the factorization mechanism. Our next theorem closes the gap between the upper and lower bounds in terms of $|Q|$, and thus gives a characterization up to $O(\log(1/\delta))$ for ℓ_2^2.

Theorem 7. *Let $\varepsilon, \delta > 0$ be smaller than some absolute constants and let Q be a workload of linear queries with workload matrix W. There exists some $\alpha^* > 0$ such that for every $\alpha \le \alpha^*$,*

$$\mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q,\alpha) = \Omega\left(\frac{\gamma_F(W)}{\varepsilon\alpha}\right).$$

In addition to being sharper, our proof of Theorem 7 is dramatically simpler than the lower bounds in [NTZ16, NT15].

Remark 8. *By a trivial reduction, Theorem 1, in fact, gives lower bounds for the distributional setting where X is sampled i.i.d. from an unknown distribution μ, and the goal is to estimate the quantity $q(\mu) = \mathop{\mathbb{E}}_{x\sim\mu}[q(x)]$ for every query $q \in Q$ up to error at most α.*

Data-Independent Mechanisms. Along the way, we prove a simple result that this sample complexity bound holds for *every* choice of α, provided we restrict attention to *data-independent mechanisms.* These mechanisms can be written in the form $\mathcal{M}(X) = Q(X) + Z/n$ for some fixed random variable Z that depends only on Q and not on the data.

For such mechanisms we show that the sample complexity is always $\Omega(\gamma_F(W)/\varepsilon\alpha)$, regardless of α.[3]

Data-independent mechanisms are interesting on their own, since the fact that we add noise from a known distribution makes them simpler to implement, and also means that we can give precise confidence intervals on the error of the mechanism. One application of our lower bound for data-independent mechanisms is an $\Omega(\log T)$ lower bound on the sample complexity of any mechanism for answering threshold queries over $[T]$ in ℓ_2^2 error, which matches the data-independent binary tree mechanism.

1.3 Techniques

Below we give a brief overview of the techniques used to prove Theorems 1 and 7.

Lower bound in the local model. As mentioned above, Theorem 1 is proved in the distributional setting, where the dataset X consists of n i.i.d. samples from some distribution μ, and the goal is to estimate the expectation of each query $q \in Q$ on μ. Our approach is to design two families of hard distributions $\{\lambda_1, \dots, \lambda_k\}$ and $\{\mu_1, \dots, \mu_k\}$ with the following properties: first, any locally differentially private mechanism requires many samples to distinguish these two families; second, the two families give very different answers to the queries.

To show that the distributions are hard to distinguish, we prove an upper bound on the KL-divergence between: (1) the transcript of a private mechanism in the local model when run on n samples from a random distribution in $\{\lambda_1, \dots, \lambda_k\}$, and (2) the same, but for a random distribution in $\{\mu_1, \dots, \mu_k\}$. Intuitively, the bound shows that the KL-divergence between transcripts is small when no bounded test function can simultaneously distinguish between λ_v and μ_v on average over a random choice of $v \in [k]$. This bound is a slight extension of a similar bound from [DJW18]. In particular, the upper bound on the KL-divergence is in terms of the $\infty \to 2$ operator norm of a matrix M derived from the two families of distributions.

Thus, what remains is to find families distributions $\{\lambda_1, \dots, \lambda_k\}$ and $\{\mu_1, \dots, \mu_k\}$, for which the $\infty \to 2$ operator norm of M is small, but the expectations of the queries in Q are sufficiently different on the two families. Recall that our goal is to prove a lower bound

[3] Technically, we require $\alpha \le \|W\|_{1\to\infty}$, but in nearly all applications of interest $\|W\|_{1\to\infty} = 1$.

in terms of the approximate norm $\gamma_2(W, \alpha)$, where W is the workload matrix. Since $\gamma_2(W, \alpha)$ is the value of a convex minimization problem, it admits a dual characterization, showing that $\gamma_2(W, \alpha)$ is equal to the value of a maximization problem over matrices U. We take an optimal dual solution U, and use it to derive distributions $\{\lambda_1, \ldots, \lambda_k\}$ and $\{\mu_1, \ldots, \mu_k\}$. The objective function of the dual problem guarantees that these distributions are such that the expectation of any query $q \in Q$ on any λ_v is small, yet the expectation of the query q_v on μ_v is large. Moreover, the dual objective, together with classical arguments in functional analysis, also guarantees an upper bound on the $\infty \to 2$ norm of the appropriate matrix M, giving us both ingredients for our lower bound.

Lower bound in the central model. The main ingredient of the proof of Theorem 7 is a lower bound of $\Omega(\gamma_F(W)/\varepsilon\alpha)$ on the sample complexity of data-independent mechanisms. Recall that a mechanism $\mathcal{M}$ is data-independent if $\mathcal{M}(X) = Q(X) + \frac{1}{n}Z$ for a random variable $Z \in \mathbb{R}^Q$. Our key observation is that, if Σ is the covariance matrix of Z, then the mechanism

$$Q(X) + \frac{O(\log(1/\delta))}{n} \cdot \mathcal{N}(0, \Sigma)$$

that uses Gaussian noise in place of Z is also (ε, δ)-differentially private. Moreover, the ℓ_2^2 error of $\mathcal{M}$ is equal to $\mathrm{Tr}(\Sigma)/|Q|^{1/2}$, so, up to a factor of $O(\log(1/\delta))$, the optimal data-independent mechanism with respect to ℓ_2^2-error can be assumed to use correlated Gaussian noise. It is easy to see that the class of all such mechanism is equivalent to the class of all factorization mechanisms, and, hence, the optimal achievable ℓ_2^2-error is $O(\gamma_F(W)/\varepsilon n)$.

To give a lower bound for arbitrary mechanisms in the high-accuracy regime, we use a clever transformation from [BDKT12] that turns a data-dependent mechanisms that is accurate for large datasets into a data-independent mechanism.

2 PRELIMINARIES

In this section we recount basic notation and definitions used throughout the paper.

2.1 Norms

For a set $\mathcal{S}$, the ℓ_1, ℓ_2 and ℓ_∞ norms on $\mathbb{R}^{\mathcal{S}}$ are given respectively by

$$\|a\|_1 = \sum_{v \in \mathcal{S}} |a_v|, \quad \|a\|_2 = \sqrt{\sum_{v \in \mathcal{S}} (a_v)^2}, \quad \|a\|_\infty = \max_{v \in \mathcal{S}} |a_v|.$$

Given a probability distribution π on $\mathcal{S}$, we consider the norms $\|\cdot\|_{L_1(\pi)}$ and $\|\cdot\|_{L_2(\pi)}$ on $\mathbb{R}^{\mathcal{S}}$, given by

$$\|a\|_{L_1(\pi)} = \sum_{v \in \mathcal{S}} \pi(v)|a_v|, \quad \|a\|_{L_2(\pi)} = \sqrt{\sum_{v \in \mathcal{S}} \pi(v)(a_v)^2}.$$

We also take advantage of a number of matrix norms. For norms $\|\cdot\|_\zeta$ and $\|\cdot\|_\xi$ on $\mathbb{R}^{\mathcal{S}}$ and $\mathbb{R}^{\mathcal{S}'}$ respectively, we consider the *matrix operator norm* of $M \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}'}$ given by

$$\|M\|_{\zeta \to \xi} = \max_{x \in \mathbb{R}^{\mathcal{S}} \setminus \{0\}} \frac{\|Mx\|_\xi}{\|x\|_\zeta}.$$

For the special case of $\|M\|_{\ell_s \to \ell_t}$, we will simply write $\|M\|_{s \to t}$. Of particular importance are $\|M\|_{1 \to \infty}$ which corresponds to the largest entries of M, $\|M\|_{1 \to 2}$, which corresponds to the maximum ℓ_2-norm of a column of M, and $\|M\|_{2 \to \infty}$, which corresponds to the maximum ℓ_2-norm of a row of M.

The *inner product* of two matrices M and N in $\mathbb{R}^{\mathcal{S} \times \mathcal{S}'}$ is defined by $M \bullet N = \mathrm{Tr}(M^\top N) = \sum_{u \in \mathcal{S}, v \in \mathcal{S}'} m_{u,v} n_{u,v}$. The *Frobenius norm* of $M \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}'}$ is given by $\|M\|_F = \sqrt{M \bullet M}$.

Lastly, the *factorization norms* γ_F and γ_2 central to this work are given for $M \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}'}$ by

$$\gamma_F(M) = \min\left\{ \frac{1}{|\mathcal{S}|^{1/2}} \|R\|_F \|A\|_{1 \to 2} : RA = M \right\},$$
$$\gamma_2(M) = \min\{ \|R\|_{2 \to \infty} \|A\|_{1 \to 2} : RA = M \}.$$

2.2 Differential Privacy

Let $\mathcal{X}$ denote the *data universe*. A generic element from $\mathcal{X}$ will be denoted by x. We consider *datasets* of the form $X = (x_1, \ldots, x_n) \in \mathcal{X}^n$, each of which is identified with its *histogram* $h \in \mathbb{Z}_{\geq 0}^{\mathcal{X}}$ where, for every $x \in \mathcal{X}$, $h_x = |\{i : x_i = x\}|$, so that $\|h\|_1 = n$. To refer to a dataset, we use X and h interchangeably. A pair of datasets $X = (x_1, \ldots, x_i, \ldots, x_n)$ and $X' = (x_1, \ldots, x_i', \ldots, x_n)$ are called *adjacent* if X' is obtained from X by replacing an element x_i of X with a new universe element x_i'.

For parameters $\varepsilon, \delta > 0$, an *(ε, δ)-differentially private mechanism* [DMNS06] (or $(\varepsilon, \delta) - DP$ for short) is a randomized function $\mathcal{M} : \mathcal{X}^n \to \Omega$ which, for all adjacent datasets X and X', for all outcomes $S \subseteq \Omega$, satisfies

$$\Pr_{\mathcal{M}}[\mathcal{M}(X) \in S] \leq e^\varepsilon \Pr_{\mathcal{M}}[\mathcal{M}(X') \in S] + \delta.$$

A mechanism which is $(\varepsilon, 0)$-differentially private will be referred to as being simply *ε-differentially private* (or ε-DP for short).

Of special interest are (ε, δ)-differentially private mechanisms $\mathcal{M}_i : \mathcal{X} \to \bar{\Omega}$ which take a singleton dataset $X = \{x\}$ as input. These are referred to as *local randomizers*. A sequence of (ε, δ)-differentially private local randomizers $\mathcal{M}_1, \ldots, \mathcal{M}_n$ together with a *post-processing function* $\mathcal{A} : \bar{\Omega}^n \to \Omega$ specify a *(non-interactive) locally (ε, δ)-differentially private mechanism* $\mathcal{M} : \mathcal{X}^n \to \Omega$ [EGS03, DMNS06, KLN+08]. In short, we say that such mechanisms are (ε, δ)-LDP, or ε-LDP when $\delta = 0$. When the local mechanism $\mathcal{M}$ is applied to a dataset X, we refer to

$$\mathcal{T}_{\mathcal{M}}(X) = (\mathcal{M}_1(x_1), \ldots, \mathcal{M}_n(x_n))$$

as the *transcript* of the mechanism. Then the output of the mechanism is given by $\mathcal{M}(X) = \mathcal{A}(\mathcal{T}_{\mathcal{M}}(X))$.

2.3 Linear Queries

A *linear query* is specified by a bounded function $q : \mathcal{X} \to \mathbb{R}$. Abusing notation slightly, its answer on a dataset X is given by $q(X) = \frac{1}{n}\sum_{i=1}^n q(x_i)$. We also extend this notation to distributions: if μ is a distribution on $\mathcal{X}$, then we write $q(\mu)$ for $\mathop{\mathbb{E}}_{x \sim \mu}[q(x)]$. A *workload* is a set of linear queries $Q = \{q_1, \ldots, q_k\}$, and $Q(X) = (q_1(X), \ldots, q_k(X))$ is used to denote their answers. The answers on a distribution μ on $\mathcal{X}$ are denoted by $Q(\mu) = (q_1(\mu), \ldots, q_k(\mu))$. We will often represent Q by its *workload matrix* $W \in \mathbb{R}^{Q \times \mathcal{X}}$ with entries $w_{q,x} = q(x)$. In this notation, the answers to the queries are given by $\frac{1}{n}Wh$. We will often use Q and W interchangeably.

2.4 Error and Sample Complexity

The ℓ_∞ *and* ℓ_2^2*-error* of a mechanism $\mathcal{M}$, which takes a dataset of size n, on the query workload Q are given by

$$\mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n) = \max_{X \in \mathcal{X}^n} \mathop{\mathbb{E}}_{\mathcal{M}}[\|\mathcal{M}(X) - Q(X)\|_\infty],$$

$$\mathrm{err}^{\ell_2^2}(\mathcal{M}, Q, n) = \max_{X \in \mathcal{X}^n} \mathop{\mathbb{E}}_{\mathcal{M}}\left[\tfrac{1}{|Q|}\|\mathcal{M}(X) - Q(X)\|_2^2\right]^{1/2}.$$

We can then define the *sample complexity* of a mechanism $\mathcal{M}$ for a given ℓ_∞ error α by

$$\mathrm{sc}^{\ell_\infty}_{\varepsilon,\delta}(\mathcal{M}, Q, \alpha) = \min\{n : \mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n) \le \alpha\}.$$

The sample complexity with respect to ℓ_2^2 error $\mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q, \alpha)$ is defined analogously.

Having defined error and sample complexity for a fixed mechanism, we can define the optimal error and sample complexity by

$$\mathrm{err}^{\ell_\infty}_{\varepsilon,\delta}(Q, n) = \min_{\mathcal{M} \text{ is } (\varepsilon,\delta)\text{-DP}} \mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n),$$

$$\mathrm{sc}^{\ell_\infty}_{\varepsilon,\delta}(Q, \alpha) = \min_{\mathcal{M} \text{ is } (\varepsilon,\delta)\text{-DP}} \mathrm{sc}^{\ell_\infty}(\mathcal{M}, Q, n).$$

The analogous quantities $\mathrm{err}^{\ell_2^2}_{\varepsilon,\delta}(Q, n)$ and $\mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q, \alpha)$ for ℓ_2^2-error are defined similarly. The optimal error and sample complexity for the local model are denoted $\mathrm{err}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,\delta}(Q, n)$ and $\mathrm{sc}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,\delta}(Q, \alpha)$, and are defined in the same way but with the minimum taken over (ε, δ)-LDP mechanisms.

2.5 Factorization Mechanisms

The Gaussian mechanism [DN03, DN04, DMNS06] is defined as

$$\mathcal{M}_{\mathrm{Gauss}}(W, h) = \frac{1}{n}Wh + Z, \quad Z \sim \mathcal{N}\left(0, \left(\frac{\sigma_{\varepsilon,\delta}\|W\|_{1\to 2}}{n}\right)^2 \cdot I\right),$$

where $\sigma_{\varepsilon,\delta} = O(\sqrt{\log(1/\delta)}/\varepsilon)$ depends only on the privacy parameters. Given a factorization $W = RA$, we consider the mechanism

$$\begin{aligned}\mathcal{M}_{R,A}(h) &= R\ \mathcal{M}_{\mathrm{Gauss}}(W, h)\\ &= \frac{1}{n}Wh + Z, \qquad Z \sim \mathcal{N}\left(0, \left(\frac{\sigma_{\varepsilon,\delta}\|A\|_{1\to 2}}{n}\right)^2 \cdot RR^\top\right),\end{aligned}$$

and, utilizing Gaussian tail bounds, one can show that the error is

$$\mathrm{err}^{\ell_\infty}(\mathcal{M}_{R,A}, Q, n) = O\left(\frac{\|R\|_{2\to\infty}\|A\|_{1\to 2}\sqrt{\log(1/\delta)}\log|Q|}{\varepsilon n}\right).$$

We define the *factorization mechanism* $\mathcal{M}_{\gamma_2}$ to be the mechanism that chooses R, A to minimize this expression, and its error is proportional to the *factorization norm*

$$\gamma_2(W) = \min\{\|R\|_{2\to\infty}\|A\|_{1\to 2} : W = RA\}.$$

The sample complexity of this mechanism is thus

$$\mathrm{sc}^{\ell_\infty}(\mathcal{M}_{\gamma_2}, Q, \alpha) = O\left(\frac{\gamma_2(W)\sqrt{\log(1/\delta)}\log|Q|}{\alpha}\right).$$

This mechanism is implicit in [NTZ16], and is stated in this form in [Nik14].

Analogously, we can show that

$$\mathrm{err}^{\ell_2^2}(\mathcal{M}_{R,A}, Q, n) = O\left(\frac{|Q|^{-1/2}\|R\|_F\|A\|_{1\to 2}\sqrt{\log(1/\delta)}}{\varepsilon n}\right).$$

Optimizing this error bound over the choice of R and A gives error proportional to the factorization norm

$$\gamma_F(W) = \min\{|Q|^{-1/2}\|R\|_F\|A\|_{1\to 2} : W = RA\},$$

and the mechanism $\mathcal{M}_{\gamma_F}$ that runs $\mathcal{M}_{R,A}$ with the R and A achieving $\gamma_F(W)$ has sample complexity

$$\mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q, \alpha) = O\left(\frac{\gamma_F(W)\sqrt{\log(1/\delta)}}{\alpha}\right).$$

This factorization mechanism is equivalent to the Gaussian noise matrix mechanism in [LHR+10].

3 NON-INTERACTIVE LOCAL DP: LINEAR QUERIES

In this section we give details about our results for answering linear queries in the local model. We first present the local approximate factorization mechanism. Then we give an information theoretic lemma that bounds the KL-divergence between the transcripts of mechanisms in the local model on inputs drawn from mixtures of product distributions. We then use a dual formulation of the approximate γ_2 norm to construct distributions to use with the information theoretic lemma in order to prove the lower bound in Theorem 1.

3.1 Approximate Factorization

Here we give details of the approximate factorization mechanism, which was sketched in the introduction. Recall that the approximate γ_2 norm is defined by

$$\gamma_2(W, \alpha) = \min\{\gamma_2(\widetilde{W}) : \|W - \widetilde{W}\|_{1\to\infty} \le \alpha/2\},$$

where $\gamma_2(\widetilde{W}) = \min\{\|R\|_{2\to\infty}\|A\|_{1\to 2} : W = RA\}$. Matrices $\widetilde{W}$, R, and A achieving the minimum to any degree of accuracy can be computed in polynomial time via semidefinite programming, as shown in [LS09]. Our main positive result shows that the sample complexity of the corresponding approximate factorization mechanism is bounded above by the approximate γ_2 norm. As sketched in the introduction, this can be achieved via a local version of the Gaussian noise mechanism, which can then be transformed into a purely private mechanism using the results of [BNS18]. This gives, however, a slightly suboptimal bound, and, instead, we use the local randomizer from [BBNS19], which is a variant of a local randomizer from [DJW18]. The relevant properties of this local randomizer are captured by the next lemma. We recall that a random variable Z over $\mathbb{R}$ is σ-subgaussian if $\mathbb{E}\exp(Z^2/\sigma^2) \le 2$, and a random variable Z over $\mathbb{R}^d$ is σ-subgaussian if $\theta^\top Z$ is σ-subgaussian for every vector θ such that $\|\theta\|_2 = 1$.

Lemma 9 ([BBNS19]). *There exists an ε-DP mechanism $\mathcal{M}$ which takes as input a single datapoint $x \in \mathbb{R}^d$ such that $\|x\|_2 \le 1$, and outputs a random $Y_x := \mathcal{M}(x) \in \mathbb{R}^d$ such that*

(1) *Y_x can be sampled in time polynomial in d on input x,*
(2) $\mathbb{E}[Y_x] = x$,

(3) *$Y_x - x$ is σ-subgaussian with $\sigma = O(\varepsilon^{-1})$.*

Based on this local randomizer, and the approximate factorizations, we prove the following upper bound in Appendix A.

Theorem 10 (Approximate Factorization Mechanism). *There exists an ε-LDP mechanism $\mathcal{M}^{loc}_{\gamma_2,\alpha}$ such that, for any k statistical queries Q with workload matrix W, we have*

$$\mathrm{sc}^{\ell_\infty}(\mathcal{M}^{loc}_{\gamma_2,\alpha}, Q, \alpha) = O\left(\frac{\gamma_2(W,\alpha/2)^2 \log k}{\varepsilon^2\alpha^2}\right),$$

and the mechanism runs in time polynomial in n, k, and $|\mathcal{X}|$.

3.2 Bounding KL-Divergence

Our lower bound will rely on the construction, based on a workload Q, of families $\{\lambda_1, \ldots, \lambda_k\}$ and $\{\mu_1, \ldots, \mu_k\}$ of distributions on $\mathcal{X}$. Together with these, we consider a distribution π over $[k]$. For any $v \in [k]$, let λ_v^n be the product distribution induced by sampling n times independently from λ_v, and let λ_π^n be the mixture $\sum_{v=1}^k \pi(v)\lambda_v^n$. Define μ_v^n and μ_π^n analogously. Note that λ_π^n and μ_π^n are *not* product distributions, but mixtures of such distributions. For a mechanism $\mathcal{M}$ in the local model, and a probability distribution ν on $\mathcal{X}^n$, we use $\mathcal{T}_\mathcal{M}(\nu)$ to denote the distribution on random transcripts $\mathcal{T}_\mathcal{M}(X)$ when X is sampled from ν. Similarly, if ν is a distribution on $\mathcal{X}$, we use the notation $\mathcal{M}_i(\nu)$ for the distribution of $\mathcal{M}_i(x)$, when x is sampled from ν.

We approach the task of showing that $\lambda_1, \ldots, \lambda_k$ and $\mu_1, \ldots, \mu_k$ are "hard" distributions on which to evaluate Q in two steps. On the one hand, we wish to argue that being able to estimate Q on the distributions $\lambda_1, \ldots, \lambda_k$ and $\mu_1, \ldots, \mu_k$ enables us to distinguish between λ_π^n and μ_π^n. On the other hand, we show a lower bound on the number of samples required for a locally private mechanism to distinguish between λ_π^n and μ_π^n. The second of these objectives will be met by way of the following bound on KL-divergence. Similar bounds were proved in [DJW18, DR18] when only one of the two distributions is a mixture of products, and our proof is similar to the proof of Theorem 2 in [DR18]. Our proof is in Appendix B.

Lemma 11. *Let $\varepsilon \in (0,1]$, and let $\mathcal{M}$ be an ε-DP mechanism in the local model. Then, for families $\{\lambda_1, \ldots, \lambda_k\}$ and $\{\mu_1, \ldots, \mu_k\}$ of distributions on $\mathcal{X}$, together with a distribution π over $[k]$,*

$$\begin{aligned}&\mathrm{D}_{KL}(\mathcal{T}_\mathcal{M}(\lambda_\pi^n)\|\mathcal{T}_\mathcal{M}(\mu_\pi^n))\\ &\le O(n\varepsilon^2)\cdot \max_{f\in\mathbb{R}^\mathcal{X}:\|f\|_\infty\le 1} \mathop{\mathbb{E}}_{V\sim\pi}\left[\left(\mathop{\mathbb{E}}_{x\sim\lambda_V}[f_x] - \mathop{\mathbb{E}}_{x\sim\mu_V}[f_x]\right)^2\right].\end{aligned}$$

In matrix notation, define the matrix $M \in \mathbb{R}^{[K]\times\mathcal{X}}$ by $m_{v,x} = (\lambda_v(x) - \mu_v(x))$. Then

$$\mathrm{D}_{KL}(\mathcal{T}_\mathcal{M}(\lambda_\pi^n)\|\mathcal{T}_\mathcal{M}(\mu_\pi^n)) \le O(n\varepsilon^2)\cdot\|M\|^2_{\ell_\infty\to L_2(\pi)}.$$

Being able to distinguish between $\mathcal{T}_\mathcal{M}(\lambda_\pi^n)$ and $\mathcal{T}_\mathcal{M}(\mu_\pi^n)$ with constant probability implies, by Pinsker's inequality, that

$$\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_\mathcal{M}(\lambda_\pi^n)\|\mathcal{T}_\mathcal{M}(\mu_\pi^n)) \ge \Omega(1).$$

Together with Lemma 11, this would imply

$$n = \Omega\left(\frac{1}{\varepsilon^2\cdot\|M\|^2_{\ell_\infty\to L_2(\pi)}}\right).$$

Hence, our goal will be to define our distributions so that that $\|M\|^2_{\ell_\infty\to L_2(\pi)}$ is small while still meeting the requirement that estimating the queries Q allows us to distinguish between λ_π^n and μ_π^n.

It is worth noting that Lemma 11 is not known to hold when the protocol is allowed to be sequentially interactive. Indeed, this is the bottleneck to generalizing our lower bound to the case of sequentially interactive local privacy. See the proof of Lemma 11 for further discussion.

3.3 Duality for $\gamma_2(W,\alpha)$ and the Dual Norm

Recall that our goal is to prove a lower bound on the sample complexity of mechanisms in the local model in terms of the approximate γ_2 norm. We will do so via Lemma 11, and the distributions $\{\lambda_1, \ldots, \lambda_k\}$ and $\{\mu_1, \ldots, \mu_k\}$ will serve as a certificate of a lower bound on the sample complexity. On the other hand, convex duality can certify a lower bound on the approximate γ_2 norm. In the proof of our lower bounds, we will show that these dual certificates for which the approximate γ_2 norm is large can be turned into hard families of distributions to use in Lemma 11.

The key duality statement follows. This dual formulation for the $\gamma_2(W,\alpha)$ was also given in [LS09] for the special case when W has entries in $\{-1,+1\}$.[4] For completeness, here we rederive it in Appendix C by directly applying the hyperplane separator theorem.

Lemma 12. *For any $k \times T$ matrix W and α,*

$$\gamma_2(W,\alpha) = \max\left\{\frac{W\bullet U - \alpha\|U\|_1}{\gamma_2^*(U)} \;:\; U\in\mathbb{R}^{k\times T},\ U\ne 0\right\},$$

where γ_2^* is the dual norm to γ_2 given by

$$\begin{aligned}\gamma_2^*(U) &= \max\{U\bullet V : V\in\mathbb{R}^{k\times T},\ \gamma_2(V)\le 1\}\\ &= \max_{\substack{a_1,\ldots,a_k\\ b_1,\ldots,b_T}} \sum_{i=1}^k\sum_{j=1}^T u_{i,j}a_i^\top b_j,\end{aligned}$$

where $a_1, \ldots, a_k$ and $b_1, \ldots, b_T$ range over vectors with unit ℓ_2 norm in $\mathbb{R}^{k+T}$.

The expression

$$\gamma_2^*(U) = \max\sum_{i=1}^k\sum_{j=1}^T u_{i,j}a_i^\top b_j,$$

with the max over unit vectors $a_1, \ldots a_k$ and $b_1, \ldots, b_T$ can be easily formulated as a semidefinite program, and, in fact, is exactly the semidefinite program that appears in Grothendieck's inequality (see, e.g., [KN12, Pis12]). It is straightforward to check (just take all the a_i and b_j co-linear) that

$$\gamma_2^*(U) \ge \max\{y^\top Uz : y\in\{-1,1\}^m, z\in\{-1,1\}^N\} = \|U\|_{\infty\to 1}. \tag{2}$$

Moreover, Grothendieck showed that this inequality is always tight up to a universal constant [Gro53], although this fact will not be used here. Instead, we will need the following lemma, which can be derived from SDP duality, and is also due to Grothendieck. For a proof using the Hahn-Banach theorem, see [Pis12].

[4] Note that in [LS09], Linial and Shraibman use the notation $\gamma_2^\alpha(W) = \inf\{\gamma_2(\widetilde{W}) : 1 \le \widetilde{w}_{ij}w_{ij} \le \alpha\ \forall i,j\}$. For sign matrices W this is equal to $\frac{\alpha+1}{2}\gamma_2(W,(\alpha-1)/(\alpha+1))$ in our notation.

Lemma 13 ([Gro53]). *For any $k \times T$ matrix U, $\gamma_2^*(U) \le t$ if and only if there exist diagonal matrices $P \in \mathbb{R}^{k \times k}$ and $Q \in \mathbb{R}^{T \times T}$, and a matrix $\widetilde{U} \in \mathbb{R}^{k \times T}$ such that $\mathrm{Tr}(P^2) = \mathrm{Tr}(Q^2) = 1$, $U = P\widetilde{U}Q$, and $\|\widetilde{U}\|_{2\to 2} \le t$.*

By (2), the $\gamma_2^*(\cdot)$ norm is an upper bound on the $\|\cdot\|_{\infty\to 1}$ norm. We use Lemma 13 to show a similar upper bound on the $\|\cdot\|_{\infty\to 2}$, which allows projecting out some of the rows of the matrix, but is quantitatively stronger. The reason we are interested in the $\|\cdot\|_{\infty\to 2}$ norm is that this is the norm that appears in the statement of Lemma 11.

Lemma 14. *For any matrix $U \in \mathbb{R}^{k \times T}$, there exists a set $S \subseteq [k]$ of size $|S| \ge \frac{k}{2}$ such that $\sqrt{\frac{k}{2}}\|\Pi_S U\|_{\infty\to 2} \le \gamma_2^*(U)$, where Π_S is the projection onto the subspace $\mathbb{R}^S$.*

The next lemma slightly strengthens Lemma 14 to allow for weights on the rows of the matrix. This is the key fact about the γ_2^* norm that we need for our lower bounds.

Lemma 15. *Let U and M be $k \times T$ matrices, and let π be a probability distribution on $[k]$ where, for any $i \in [k]$, $j \in [T]$, we have $u_{i,j} = \pi(i) m_{i,j}$. Then there exists a probability distribution $\widehat{\pi}$ on $[k]$, with support contained in the support of π, such that $\|M\|_{\ell_\infty \to L_2(\widehat{\pi})} \le 4\gamma_2^*(U)$.*

Lemmas 14 and 15 are proved in Appendix C.

3.4 Symmetrization

For our lower bound, it will be convenient to narrow our attention to the following restricted class of 'symmetric' query workloads.

Definition 16. *Let Q be a workload of statistical queries with workload matrix $W \in \mathbb{R}^{Q \times \mathcal{X}}$. Suppose there exists a partition of $\mathcal{X}$ into sets $\mathcal{X}^+$ and $\mathcal{X}^-$, $|\mathcal{X}^+| = |\mathcal{X}^-|$, where each element x of $\mathcal{X}^+$ is identified with a distinct element of $\mathcal{X}^-$, denoted $-x$, such that, for all $q \in Q$, for all $x \in \mathcal{X}$, $q(-x) = -q(x)$. In other words, W can be expressed as (W^+, W^-), where $W^+ \in \mathbb{R}^{Q \times \mathcal{X}^+}$ and $W^- \in \mathbb{R}^{Q \times \mathcal{X}^-}$ are the restrictions of W to $Q \times \mathcal{X}^+$ and $Q \times \mathcal{X}^-$ respectively, with each entry $w^+_{q,x}$ of W^+ and the corresponding entry $w^-_{q,-x}$ of W^- satisfying $w^-_{q,x} = -w^+_{q,-x}$. Also write Q^+ to denote the collection of queries with workload matrix W^+ so that the queries $q^+ : \mathcal{X}^+ \to \mathbb{R}$ of Q^+ are obtained by restricting queries $q : \mathcal{X} \to \mathbb{R}$ of Q to the input space $\mathcal{X}^+$; define Q^- analogously. Then Q, and also W, are called symmetric.*

The following result will allow us to translate our lower bound for the symmetric query workloads into a lower bound for general query workloads. Its proof is given in Appendix D.

Lemma 17. *Let $\alpha, \epsilon > 0$. Let Q be a symmetric workload of statistical queries and take Q^+ as given by Definition 16. Suppose there exists a non-interactive locally ε-LDP mechanism $\mathcal{M}^+$ which takes n samples as input and achieves $\mathrm{err}^{\ell_\infty}(\mathcal{M}^+, Q^+, n) \le \alpha$. Then there exists a local 3ε-LDP mechanism $\mathcal{M}$ which takes $n' = \max\{n, \frac{1}{\varepsilon^2\alpha^2}\}$ samples as input and achieves $\mathrm{err}^{\ell_\infty}(\mathcal{M}, Q, n') \le 4\alpha$.*

Lemma 18 allows us to relate $\gamma_2(W)$ and $\gamma_2(W^+)$ and their witnesses. Its proof is also given in Appendix D.

Lemma 18. *Let $\alpha > 0$ and let $W \in \mathbb{R}^{Q \times \mathcal{X}}$ be a symmetric workload matrix with $\mathcal{X}^+$ and W^+ as given by Definition 16. Then it holds that $\gamma_2(W) = \gamma_2(W^+)$ and $\gamma_2(W, \alpha) = \gamma_2(W^+, \alpha)$. Moreover, if, for some $U^+ \in \mathbb{R}^{Q \times \mathcal{X}^+}$,*

$$\gamma_2(W^+, \alpha) = \frac{W^+ \bullet U^+ - \alpha\|U^+\|_1}{\gamma_2^*(U^+)},$$

then

$$\gamma_2(W, \alpha) = \frac{W \bullet U - \alpha\|U\|_1}{\gamma_2^*(U)},$$

where $U = \frac{1}{2}(U^+, U^-)$ is a matrix in $\mathbb{R}^{Q \times \mathcal{X}}$ such that the submatrix U^- is indexed by $\mathcal{X}^-$ and has entries $u^-_{q,-x} = -u^+_{q,x}$ for all $x \in \mathcal{X}^+$ and $q \in Q$.

3.5 Lower Bound Based on Dual Solutions

In this section we put together the different tools we have already set up – the KL-divergence lower bound, and the duality of the approximate γ_2 norm – in order to prove our main lower bound result Theorem 1.

For this section, it is convenient to consider the enumeration $q_1, \ldots, q_k$ of the queries of a symmetric workload Q with workload matrix $W \in \mathbb{R}^{[k] \times \mathcal{X}}$. Let U be the dual witness to the lower bound on $\gamma_2(W, \alpha)$, as given by Lemma 12, so that

$$\gamma_2(W, \alpha) = \frac{W \bullet U - \alpha\|U\|_1}{\gamma_2^*(U)}. \tag{3}$$

By Lemma 18, we may assume without loss of generality that U is of the form (U^+, U^-) where each entry of U^- is the additive inverse of the corresponding entry of U^+. Furthermore, by dividing each entry of U by $\|U\|_1$ if necessary, then we may assume without loss of generality that $\|U\|_1 = 1$. In this case,

$$\gamma_2(W, \alpha) = \frac{W \bullet U - \alpha}{\gamma_2^*(U)}.$$

Let us make a first attempt at constructing our collection of "hard" distributions $\lambda_1, \ldots, \lambda_k$ and $\mu_1, \ldots, \mu_k$ for Q. Since $\|U\|_1 = 1$, then

$$\pi(v) = \sum_{x \in \mathcal{X}} |u_{v,x}| \tag{4}$$

defines a valid probability distribution over $[k]$. For each $v \in [k]$, we then define a pair of distributions λ_v and μ_v given by

$$\forall x \in \mathcal{X}^+ : \ \lambda_v(x) = \lambda_v(-x) = |u_{v,x}|/\pi(v) \tag{5}$$

$$\forall x \in \mathcal{X}^+ : \ \mu_v(x) = \begin{cases} 2|u_{v,x}|/\pi(v) & \text{if } u_{v,x} \ge 0 \\ 0 & \text{if } u_{v,x} < 0 \end{cases} \tag{6}$$

$$\mu_v(-x) = \begin{cases} 0 & \text{if } u_{v,x} \ge 0 \\ 2|u_{v,x}|/\pi(v) & \text{if } u_{v,x} < 0 \end{cases} \tag{7}$$

Then, for all $i, v \in [k]$, the symmetry of λ_v implies $q_i(\lambda_v) = 0$. By contrast, it holds for all $v \in [k]$ that

$$\begin{aligned} q_v(\mu_v) &= \sum_{x \in \mathcal{X}} q_v(x)\mu_v(x) \\ &= \sum_{x \in \mathcal{X}^+} q_v(x)(\mu_v(x) - \mu_v(-x)) \\ &= 2W^+ \bullet U^+ = W \bullet U. \end{aligned}$$

Hence,

$$\mathop{\mathbb{E}}_{V\sim\pi}\left[\max_{i\in[k]} q_i(\mu_V)\right] \geq \mathop{\mathbb{E}}_{V\sim\pi}[q_V(\mu_V)] = W \bullet U.$$

Since $W \bullet U = \gamma_2^*(U)\gamma_2(W,\alpha) + \alpha \geq \alpha$ by Lemma 12, then

$$\mathbb{E}_{V\sim\pi}[\max_{i\in[k]} q_i(\mu_V)] \geq \alpha.$$

If we could guarantee that $q_V(\mu_V)$ was close to its expectation when $V \sim \pi$, then estimating each of the queries q_i of Q with error less than α would allow us to distinguish the distributions $\lambda_1,\dots,\lambda_k$ from the distributions $\mu_1,\dots,\mu_k$. The following result modifies our distributions in a way that resolves this issue.

Lemma 19. *Let Q be a collection of symmetric queries with workload matrix $W \in \mathbb{R}^{[k]\times\mathcal{X}}$. Let $U \in \mathbb{R}^{[k]\times\mathcal{X}}$ be the dual witness so that (3) is satisfied. Then there exist probability distributions $\widetilde{\lambda}_1,\dots,\widetilde{\lambda}_k$ and $\widetilde{\mu}_1,\dots,\widetilde{\mu}_k$ over $\mathcal{X}$, and a distribution $\widetilde{\pi}$ over $[k]$ such that:*

1. *$q_i(\widetilde{\lambda}_v) = 0$ for all $i, v \in [k]$;*
2. *for all v in the support of $\widetilde{\pi}$, $q_v(\widetilde{\mu}_v) \geq \frac{W\bullet U - \alpha/4}{O(\log(1/\alpha))}$;*
3. *the matrix $\widetilde{U} \in \mathbb{R}^{[Q]\times\mathcal{X}}$ with entries $\widetilde{u}_{v,x} = \widetilde{\pi}(v)(\widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x))$ satisfies $\gamma_2^*(\widetilde{U}) \leq \gamma_2^*(U)$.*

The proof of Lemma 19 will take advantage of the following exponential binning lemma. A proof is given in Appendix E.

Lemma 20. *Suppose that $a_1,\dots,a_k \in [0,1]$ and that π is a probability distribution over $[k]$. Then for any $\beta \in (0,1]$, there exists a set $S \subseteq [k]$ such that $\pi(S)\cdot\min_{v\in S} a_v \geq \frac{\sum_{v=1}^k \pi(v)a_v - \beta}{O(\log(1/\beta))}$.*

Proof of Lemma 19. Let $\lambda_1,\dots,\lambda_k,\mu_1,\dots,\mu_k$, and π be as given by equations (4) - (7). Since $q_v(\mu_v) > 0$ for all v, we may apply Lemma 20 with $a_v = q_v(\mu_v)$ and $\beta = \alpha/4$ to obtain a subset $S \subseteq [k]$ for which

$$\pi(S)\cdot\min_{v\in S} q_v(\mu_v) \geq \frac{\mathbb{E}_{V\sim\pi} q_v(\mu_v) - \alpha/4}{O(\log(1/\alpha))} = \frac{W\bullet U - \alpha/4}{O(\log(1/\alpha))}.$$

Now define $\widetilde{\pi}$ as π conditional on S. In particular,

$$\widetilde{\pi}(v) = \begin{cases} \pi(v)/\pi(S), & \text{if } v \in S \\ 0, & \text{otherwise.} \end{cases}$$

Then, for all $v \in [k]$, define $\widetilde{\lambda}_v = \lambda_v$ and $\widetilde{\mu}_v = \pi(S)\mu_v + (1-\pi(S))\lambda_v$. This implies

$$\forall i, v \in [k]: \quad q(\widetilde{\lambda}_v) = q(\lambda_v) = 0,$$

$$\forall v \in [k]: \quad q_v(\widetilde{\mu}_v) = \pi(S)q_v(\mu_v) \geq \frac{W\bullet U - \alpha/4}{O(\log(1/\alpha))},$$

$$\forall v \in [k]: \quad \widetilde{\mu}_v - \widetilde{\lambda}_v = \pi(S)(\mu_v - \lambda_v).$$

By the last of these facts, together with the definition of $\widetilde{\pi}$, it follows that the entries $\widetilde{u}_{v,x} = \widetilde{\pi}(v)(\widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x))$ of the matrix $\widetilde{U}$ satisfy

$$\widetilde{u}_{v,x} = \begin{cases} u_{v,x}, & \text{if } v \in S \\ 0, & \text{otherwise.} \end{cases}$$

In other words, $\widetilde{U}$ is obtained from U by replacing some of its rows with the zero-vector. It is easy to see from the definition of γ_2^* that this implies $\gamma_2^*(\widetilde{U}) \leq \gamma_2^*(U)$. □

Consider now the matrix $\widetilde{M} \in \mathbb{R}^{[k]\times\mathcal{X}}$ with entries $\widetilde{m}_{v,x} = \widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x)$. Since $\widetilde{M}$ is obtained from the matrix $\widetilde{U}$ of Lemma 19 by scaling each row v of $\widetilde{U}$ by $\frac{1}{2\pi(v)}$, it follows that

$$\|\widetilde{M}\|_{\ell_\infty\to L_1(\widetilde{\pi})} = \frac{1}{2}\|\widetilde{U}\|_{\infty\to 1} \leq \gamma_2^*(\widetilde{U}) \leq \gamma_2^*(U) = \frac{W\bullet U - \alpha}{\gamma_2(W,\alpha)}.$$

This is not quite the quantity

$$\|\widetilde{M}\|^2_{\ell_\infty\to L_2(\widetilde{\pi})} = \max_{f\in\mathbb{R}^{\mathcal{X}}:\|f\|_\infty\leq 1} \mathop{\mathbb{E}}_{V\sim\pi}\left[\left(\mathbb{E}_{x\sim\widetilde{\lambda}_V}[f_x] - \mathbb{E}_{x\sim\widetilde{\mu}_V}[f_x]\right)^2\right]$$

which Lemma 11 would have us bound. For comparison, note

$$\|\widetilde{M}\|_{\ell_\infty\to L_1(\widetilde{\pi})} = \max_{f\in\mathbb{R}^{\mathcal{X}}:\|f\|_\infty\leq 1} \mathop{\mathbb{E}}_{V\sim\pi}\left[\left|\mathop{\mathbb{E}}_{x\sim\widetilde{\lambda}_V}[f_x]] - \mathop{\mathbb{E}}_{x\sim\widetilde{\mu}_V}[f_x]\right|\right].$$

Since the trivial case of Holder's inequality implies that the $L_1(\widetilde{\pi})$-norm is always bounded above by the $L_2(\widetilde{\pi})$-norm, it holds that $\|\widetilde{M}\|_{\ell_\infty\to L_1(\widetilde{\pi})} \leq \|\widetilde{M}\|_{\ell_\infty\to L_2(\widetilde{\pi})}$. However, this inequality goes in the wrong direction for our requirements. This issue is remedied by taking advantage of Lemma 15.

Lemma 21. *Let Q be a collection of symmetric queries with workload matrix $W \in \mathbb{R}^{[k]\times\mathcal{X}}$. Let $U \in \mathbb{R}^{[k]\times\mathcal{X}}$ be the dual witness so that (3) is satisfied. Then there exist probability distributions $\widetilde{\lambda}_1,\dots,\widetilde{\lambda}_k$ and $\widetilde{\mu}_1,\dots,\widetilde{\mu}_k$ over $\mathcal{X}$, and a distribution $\widehat{\pi}$ over $[k]$ such that:*

1. *$\widetilde{\lambda}_1,\dots,\widetilde{\lambda}_k,\widetilde{\mu}_1,\dots,\widetilde{\mu}_k$ and $\widehat{\pi}$ satisfy criteria 1. and 2. of Lemma 19;*
2. *the matrix $\widetilde{M}$ with entries $\widetilde{m}_{v,x} = \widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x)$ satisfies*

$$\|\widetilde{M}\|_{\ell_\infty\to L_2(\widehat{\pi})} \leq 4\gamma_2^*(U) = \frac{4(W\bullet U - \alpha)}{\gamma_2(W,\alpha)}$$

Proof. Let $\widetilde{\lambda}_1,\dots,\widetilde{\lambda}_k,\widetilde{\mu}_1,\dots,\widetilde{\mu}_k$ and $\widetilde{\pi}$ be the distributions guaranteed to exist by Lemma 19, and let $\widetilde{U} \in \mathbb{R}^{[k]\times\mathcal{X}}$ be the corresponding matrix with entries $\widetilde{u}_{v,x} = \widetilde{\pi}(v)(\widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x))$. The entries of the matrix $\widetilde{M}$ satisfy $\pi(v)\widetilde{m}_{v,x} = \widetilde{u}_{v,x}$, so we may apply Lemma 15 to obtain a distribution $\widehat{\pi}$ such that

$$\|\widetilde{M}\|_{\ell_\infty\to L_2(\widehat{\pi})} \leq 4\gamma_2^*(\widetilde{U}) \leq 4\gamma_2^*(U) = \frac{4(W\bullet U - \alpha)}{\gamma_2(W,\alpha)}.$$

Lemma 15 further guarantees that the support of $\widehat{\pi}$ lies within the support of $\tilde{\pi}$, which together with the properties of the distributions $\widetilde{\lambda}_1,\dots,\widetilde{\lambda}_k,\widetilde{\mu}_1,\dots,\widetilde{\mu}_k$ and $\widehat{\pi}$ gives the first condition of our lemma. □

At last, we have all the components needed to prove our lower bounds for symmetric workloads.

Theorem 22. *Let $\alpha, \varepsilon \in (0,1]$. Let Q be a symmetric workload of statistical queries with workload matrix $W \in \mathbb{R}^{[k]\times\mathcal{X}}$. Then, for some $\alpha' = \Omega(\alpha/\log(1/\alpha))$, if $\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2} \geq \frac{C\log 2k}{(\alpha')^2}$ for a large enough constant C, we have*

$$\mathrm{sc}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,0}(Q,\alpha') = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}\right).$$

Proof. Let $\alpha' = \Omega(\alpha/\log(1/\alpha))$ be a value that will be decided shortly, and C' be a sufficiently large constant. If we run a ε-DP mechanism $\mathcal{M}$ on $n = \max\left\{\mathrm{sc}^{\ell_\infty}(\mathcal{M},Q,\alpha'), \frac{C'\log 2k}{(\alpha')^2}\right\}$ samples drawn i.i.d. from some distribution μ on $\mathcal{X}$, then, by classical uniform convergence results, $\mathop{\mathbb{E}}_{X\sim\mu^n}\left[\|Q(X)-Q(\mu)\|_\infty\right] \le \alpha'$, where $Q(\mu) = (q_1(\mu), \dots, q_k(\mu))$. Therefore, the mechanism will satisfy

$$\mathop{\mathbb{E}}_{X\sim\mu^n}\left[\|\mathcal{M}(X)-Q(\mu)\|_\infty\right] \le 2\alpha'. \tag{8}$$

We will show that for any ε-LDP mechanism $\mathcal{M}$ such that (8) holds for an arbitrary μ, we must have

$$n = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}\right). \tag{9}$$

Therefore, we get that $\max\left\{\mathrm{sc}^{\ell_\infty}(\mathcal{M}, Q, \alpha'), \frac{C'\log 2k}{(\alpha')^2}\right\} = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2},\right)$ which implies the theorem by the assumption on $\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}$.

Let $\widetilde{\lambda}_1, \dots, \widetilde{\lambda}_k$, $\widetilde{\mu}_1, \dots, \widetilde{\mu}_k$ and $\widehat{\pi}$ be the distributions, and $\widetilde{M} \in \mathbb{R}^{[k]\times\mathcal{X}}$ the matrix, guaranteed to exist by Lemma 21. The matrix $\widetilde{M}$ has entries $\widetilde{m}_{v,x} = \widetilde{\lambda}_v(x) - \widetilde{\mu}_v(x)$ and satisfies

$$\|\widetilde{M}\|_{\ell_\infty\to L_2(\widehat{\pi})} \le \frac{4(W\bullet U-\alpha)}{\gamma_2(W,\alpha)}.$$

Equivalently,

$$\max_{f\in\mathbb{R}^{\mathcal{X}}:\|f\|_\infty\le 1}\ \mathop{\mathbb{E}}_{V\sim\widehat{\pi}}\left[\left(\mathop{\mathbb{E}}_{x\sim\widetilde{\lambda}_V}[f_x] - \mathop{\mathbb{E}}_{x\sim\widetilde{\mu}_V}[f_x]\right)^2\right] \le \left(\frac{4(W\bullet U-\alpha)}{\gamma_2(W,\alpha)}\right)^2.$$

By Lemma 11, this implies

$$\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{\mathcal{M}}(\widetilde{\lambda}^n_{\widehat{\pi}})\|\mathcal{T}_{\mathcal{M}}(\widetilde{\mu}^n_{\widehat{\pi}})) \le O(n\varepsilon^2)\cdot\left(\frac{W\bullet U-\alpha}{\gamma_2(W,\alpha)}\right)^2 \tag{10}$$

Lemma 21 guarantees further that $q_i(\widetilde{\lambda}_v) = 0$ for all $i, v \in [k]$, while $q_v(\widetilde{\mu}_v) \ge \frac{W\bullet U-\alpha/4}{O(\log(1/\alpha))}$ for all v in the support of $\widehat{\pi}$. Let $\alpha' = \frac{1}{8}\min_{v\in[k]} q_v(\widetilde{\mu}_v)$. Then a mechanism $\mathcal{M}$ satisfying (8) can distinguish between the distributions $\widetilde{\lambda}^n_{\widehat{\pi}}$ and $\widetilde{\mu}^n_{\widehat{\pi}}$ with constant probability, and, by Pinsker's inequality, $\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{\mathcal{M}}(\widetilde{\lambda}^n_{\widehat{\pi}})\|\mathcal{T}_{\mathcal{M}}(\widetilde{\mu}^n_{\widehat{\pi}}))$ is bounded from below by some constant $C > 0$. By (10), this implies that

$$n = \Omega\left(\frac{\gamma_2(W,\alpha)}{\varepsilon\cdot(W\bullet U-\alpha)}\right)^2$$

samples are required to obtain accuracy $\alpha'/4$ and privacy ε.

Case 1: $W\bullet U \le 2\alpha$. Recall that $W\bullet U \ge \alpha$. Hence, if $W\bullet U \le 2\alpha$, then $n = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}\right)$ and furthermore

$$\alpha' \ge \frac{W\bullet U-\alpha/4}{O(\log(1/\alpha))} = \Omega\left(\frac{\alpha}{\log(1/\alpha)}\right)$$

Case 2: $W\bullet U > 2\alpha$. However, if $W\bullet U > 2\alpha$, then, for $\beta\in[0,1]$, we may instead consider the distributions $\widehat{\mu}_v = (1-\beta)\cdot\widetilde{\lambda}_v + \beta\cdot\widetilde{\mu}_v$ and $\widehat{\lambda}_v = \widetilde{\lambda}_v$, given for $v\in[k]$. We have

$$\begin{aligned}
&\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{\mathcal{M}}(\widehat{\lambda}^n_{\widehat{\pi}})\|\mathcal{T}_{\mathcal{M}}(\widehat{\mu}^n_{\widehat{\pi}}))\\
&\le O(\varepsilon^2 n)\cdot \max_{f\in\mathbb{R}^{\mathcal{X}}:\|f\|_\infty\le 1}\ \mathop{\mathbb{E}}_{V\sim\widehat{\pi}}\left[\left(\mathop{\mathbb{E}}_{x\sim\widehat{\lambda}_V}[f_x] - \mathop{\mathbb{E}}_{x\sim\widehat{\mu}_V}[f_x]\right)^2\right]\\
&= O(\varepsilon^2 n)\cdot\beta^2 \max_{f\in\mathbb{R}^{\mathcal{X}}:\|f\|_\infty\le 1}\ \mathop{\mathbb{E}}_{V\sim\widehat{\pi}}\left[\left(\mathop{\mathbb{E}}_{x\sim\widetilde{\lambda}_V}[f_x] - \mathop{\mathbb{E}}_{x\sim\widetilde{\mu}_V}[f_x]\right)^2\right]\\
&\le O(\varepsilon^2 n)\cdot\beta^2\cdot\left(\frac{W\bullet U-\alpha}{\gamma_2(W,\alpha)}\right)^2.
\end{aligned}$$

Also, $q_i(\widehat{\lambda}_v) = 0$ for all $i, v\in[k]$, while

$$q_v(\widehat{\mu}_v) = \beta\cdot q_v(\widetilde{\mu}_v) \ge \beta\cdot\left(\frac{W\bullet U-\alpha/4}{O(\log(1/\alpha))}\right)$$

for all i in the support of $\widehat{\pi}$. In particular, if we set

$$\alpha' = \frac{1}{8}\min_v q_v(\widehat{\mu}_v) \ge \frac{\beta(W\bullet U-\alpha/4)}{O(\log(1/\alpha))}$$

and (8) holds for $\mathcal{M}$ and this value of α', then $\mathcal{M}$ can distinguish between $\widehat{\lambda}^n_{\widehat{\pi}}$ and $\widehat{\mu}^n_{\widehat{\pi}}$. This implies $\mathrm{D}_{\mathrm{KL}}(\mathcal{T}_{\mathcal{M}}(\widehat{\lambda}^n_{\widehat{\pi}})\|\mathcal{T}_{\mathcal{M}}(\widehat{\mu}^n_{\widehat{\pi}}))$ is bounded below by a constant, from which we obtain that

$$n = \Omega\left(\frac{\gamma_2(W,\alpha)}{\varepsilon\beta\cdot(W\bullet U)}\right)^2$$

samples are required for privacy ε and accuracy α'. Indeed, by taking $\beta = \frac{U\bullet W}{\alpha}$, we get that

$$n = \Omega\left(\left(\frac{\gamma_2(W,\alpha)}{\varepsilon\alpha}\right)^2\right)$$

samples are required for privacy ε and accuracy α' which satisfies $\alpha' \ge \frac{\beta(W\bullet U-\alpha/4)}{O(\log(1/\alpha))} = \Omega\left(\frac{\alpha}{\log(1/\alpha)}\right)$.

In both cases, $n = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}\right)$ samples are required for privacy ε and accuracy α', where $\alpha' = \Omega\left(\frac{\alpha}{\log(1/\alpha)}\right)$ □

The symmetrization techniques of

Theorem 23 (Formal version of Theorem 1). *Let $\alpha, \varepsilon \in (0,1]$. Let Q be a collection of queries with workload matrix W. Then, for some $\alpha' = \Omega\left(\frac{\alpha}{\log(1/\alpha)}\right)$, if $\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2} \ge \frac{C\log 2k}{(\alpha')^2} + \frac{C}{\varepsilon^2(\alpha')^2}$ for a large enough constant C, we have*

$$\mathrm{sc}^{\ell_\infty,\mathrm{loc}}_{\varepsilon,0}(Q,\alpha') = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2\alpha^2}\right).$$

3.6 Applications of the Lower Bounds

In this subsection we apply Theorem 23 to several workloads of interest, and, using known bounds on the approximate γ_2 norm, prove new lower bounds on the sample complexity of these workloads.

We start with the threshold queries Q^{cdf}_T. Identifying q_t with t, we see that the corresponding workload matrix W is a lower triangular matrix, with entries equal to 1 on and below the main diagonal. Let us consider a different matrix $W' = 2W - J$, where J is the all-ones $T\times T$ matrix. Forster et al. [FSSS03] showed a lower bound on the margin complexity of W', which implies that for any $\widehat{W}$ such that $\widehat{w}_{t,x} w'_{t,x} \geq 1$ holds for all $t, x \in [T]$, we have

$$\gamma_2(\widehat{W}) = \Omega(\log T). \tag{11}$$

Note that, if $\widetilde{W}$ satisfies $\|\widetilde{W} - W'\|_{1\to\infty} \leq \frac{1}{2}$, then we can take $\widehat{W} = 2\widetilde{W}$, and (11) implies $\gamma_2(W', 1/2) = \Omega(\log T)$. Finally, homogeneity and the triangle inequality for γ_2, and $\gamma_2(J) = 1$ imply that $\gamma_2(W, 1/2) \geq \frac{1}{2}\gamma_2(W', 1/2) - \frac{1}{2} = \Omega(\log T)$. Together with Theorem 23, this gives Corollary 4.

Next, we consider the parity queries $Q^{\text{parity}}_{d,w}$. Note that the workload matrix W of these queries is a submatrix consisting of $\binom{d}{w}$ rows of the $2^d \times 2^d$ Hadamard matrix. Let $s = 2^d\binom{d}{w}$ be the number of entries in W. To prove a lower bound on $\gamma_2(W, \alpha)$, we can use Lemma 12 with $U = W$. The rows of a Hadamard matrix are pairwise orthogonal and have ℓ_2 norm $2^{d/2}$, and, so, Lemma 13, used with P and Q set to appropriately scaled copies of the identity matrices of the respective dimensions, implies that $\gamma_2^*(U) \leq \sqrt{s2^d}$. Moreover, $W \bullet U = \|U\|_1 = s$, and, by Lemma 12, we have

$$\gamma_2(W, 1/2) \geq \frac{\sqrt{s}}{2^{(d/2)+1}} = \Omega\left(\binom{d}{w}^{1/2}\right).$$

This gives Corollary 5.

Finally, we treat marginal queries. Let us define these queries slightly more generally than we did in the introduction, by allowing for negation. We define $Q^{\text{marginal}}_{d,w}$ to consist of the queries $q_{S,y}(X) = \frac{1}{n}\sum_{i=1}^n \prod_{j\in S} \mathbb{I}[x_{i,j} = y_j]$, with S ranging over subsets of $[d]$ of size at most w, and y ranging over $\{0,1\}^d$. These queries can be expressed in terms of the q_S queries defined in the introduction by doubling the dimension d.

To prove a lower bound for $Q^{\text{marginal}}_{d,w}$, we use the pattern matrix method of Sherstov [She11]. We will omit a full definition of a pattern matrix here, and refer the reader to Sherstov's paper. Instead, we remark that, denoting by f the AND function on w bits, a (d, w, f)-pattern matrix W' is a $\frac{(2d)^w}{w^w} \times 2^d$ submatrix of the workload matrix W for $Q^{\text{marginal}}_{d,w}$. Let $s = 2^d \frac{(2d)^w}{w^w}$ be the number of entries in W'. By Theorem 8.1. in [She11], we have that, for any $\alpha \leq \frac{1}{6}$,

$$\min\left\{\frac{1}{\sqrt{s}}\|\widetilde{W}\|_{tr} : \|\widetilde{W} - W'\|_{1\to\infty} \leq \alpha\right\} = \Omega\left(\frac{d}{w}\right)^{\deg_{1/3}(f)/2},$$

where $\|\widetilde{W}\|_{tr}$ is the trace-norm, i.e., the sum of singular values of $\widetilde{W}$, and $\deg_{1/3}(f)$ is the $(1/3)$-approximate degree of f, which is known to be $\Omega(\sqrt{w})$ [NS94]. Since $\frac{1}{\sqrt{s}}\|\widetilde{W}\|_{tr}$ is a lower bound on $\gamma_2(\widetilde{W})$ (see [LMSS07, Lemma 3.4]), this implies

$$\gamma_2(W, 1/6) \geq \frac{1}{\sqrt{s}}\|\widetilde{W}\|_{tr} = \Omega\left(\frac{d}{w}\right)^{\Omega(\sqrt{w})},$$

giving us Corollary 6.

4 NON-INTERACTIVE LOCAL DP: PAC LEARNING

It turns out that we are able to translate our algorithm and lower bound for answering linear queries in the local model into an algorithm and lower bound for *probably approximately correct learning* in the local model.

A concept $c : \mathcal{X}^+ \to \{-1,+1\}$ from a concept class C identifies each sample x of $\mathcal{X}^+$ with a label $c(x)$. The labelled pair $(x, c(x)) = (x, 1)$ may be identified with the sample x of $\mathcal{X}^+$, while the labelled pair $(x, c(x)) = (x, -1)$ may be identified with the sample $-x$ of $\mathcal{X}^-$. Let $q : \mathcal{X} \to \{-1,+1\}$ be given by

$$q(x) = \begin{cases} c(x), & \text{if } x \in \mathcal{X}^+ \\ -c(-x), & \text{if } x \in \mathcal{X}^- \end{cases}$$

Then the *loss* of the concept c on a dataset $\overline{X} = ((x_1, y_1), \ldots, (x_n, y_n))$, denoted $\Lambda_{\overline{X}}(c)$, is

$$\begin{aligned}\Lambda_{\overline{X}}(c) &= \frac{1}{n}\sum_{i=1}^n (1 - \mathbb{I}[f(x_i) = y_i]) \\ &= \frac{1}{2} - \frac{1}{2n}\sum_{i=1}^n f(x_i)y_i = \frac{1}{2} - \frac{1}{2n}\sum_{i=1}^n q(x_i \cdot y_i) = \frac{1}{2} - \frac{1}{2}q(X)\end{aligned}$$

where X is the dataset $(x_1 \cdot y_1, \ldots, x_n \cdot y_n)$. In this way, estimating $\Lambda_{\overline{X}}(c)$ given the dataset $\overline{X}$ is equivalent to estimating $q(X)$ given the dataset X. More generally, if we consider the query workload Q consisting of all such queries q obtained from some concept c of C in this way, then estimating $Q(X)$ is equivalent to estimating $\Lambda_{\overline{X}}(C) = (\Lambda_{\overline{X}}(c))_{c\in C}$. This idea allows us to adapt the algorithm of Theorem 10 for estimating linear queries to an algorithm for learning. The result is stated in terms of the *concept matrix* $D \in \mathbb{R}^{C\times\mathcal{X}^+}$ of C with entries given by

$$d_{c,x} = c(x)$$

and takes advantage of the fact that the workload matrix W of the corresponding query workload Q is obtained by extending D to $C \times \mathcal{X}$ in the usual way with $w_{q,x} = d_{c,x}$ and $w_{q,-x} = -d_{c,x}$ for $q \in Q$ and $x \in \mathcal{X}^+$ when c is the concept that corresponds to q. In particular, the queries Q are symmetric, and, by Lemma 18, $\gamma_2(D, \alpha) = \gamma_2(W, \alpha)$.

In order to state our results for agnostic learning, we need to define notation for population loss, in addition to the empirical loss defined above. For a distribution μ over $\mathcal{X}^+ \times \{-1,+1\}$, we will use $\Lambda_\mu(c)$ to denote the loss of the concept c on μ, given by

$$\Lambda_\mu(c) = \Pr_{(x,y)\sim\mu}[c(x) \neq y].$$

For $\alpha, \beta > 0$ we will say that the mechanism $\mathcal{M}$ *(α,β)-learns* C with n samples if, for all distributions μ over $\mathcal{X}^+ \times \{-1,+1\}$, given as input a dataset $\overline{X} = ((x_1, y_1), \ldots, (x_n, y_n))$ of n samples drawn IID from μ, $\mathcal{M}$ outputs a concept $c \in C$ and an estimate $\overline{\Lambda}$ such that

$$\Pr_{\mathcal{M},\overline{X}}[\Lambda_\mu(c) \leq \min_{c'\in C}\Lambda_\mu(c') + \alpha \text{ and } |\overline{\Lambda} - \Lambda_\mu(c)| \leq \alpha] \geq 1 - \beta.$$

Typically, the learning problem does not require outputting an estimate of the loss $\Lambda_\mu(c)$, since it is usually easy to compute such an estimate with few additional samples, once a concept c has been computed. In the local model, however, this would require an additional round of interactivity. Since we focus on the non-interactive local model, it is natural to make this additional requirement on the learning algorithm.

Since we wish to bound population loss, it is necessary to assume that there are sufficiently many samples to guarantee uniform convergence. It suffices to assume, for some constant C, that the number of samples is at least $n \geq \frac{C \log 2|C|}{\alpha^2}$ to guarantee

$$\Pr_{\overline{X}}[\forall c \in C,\ |\Lambda_{\overline{X}}(c) - \Lambda_{\mu}(c)| \leq \alpha] \geq 1 - \frac{\beta}{2}$$

when $\overline{X}$ consists of n IID samples drawn from μ.

Theorem 24. *Let $\alpha, \beta \in (0,1)$, and let $\varepsilon > 0$. There exists an ε-LDP mechanism $\mathcal{M}$ such that, for any concept class C of size $|C| = k$ with corresponding concept matrix $D \in \mathbb{R}^{C \times \mathcal{X}^+}$, it suffices to have a dataset $\overline{X} = ((x_1, y_1), \ldots, (x_n, y_n))$ of*

$$n = \max\left\{O\left(\frac{\gamma_2(D,\alpha)^2 \log k}{\varepsilon^2 \alpha^2}\right), O\left(\frac{\log k}{\alpha^2}\right)\right\}$$

samples to guarantee that $\mathcal{M}$ (α, β)-learns C.

Applying the same ideas, we know that if we estimate the quantity $\min_{c \in C} \Lambda_{\overline{X}}(c)$, then we can estimate $\max_{q \in Q} q(X)$. Similarly, estimating $\min_{c \in C} \Lambda_{\mu}(c)$ is equivalent to estimating $\max_{q \in Q} q(\mu')$ where μ' is the distribution on $\mathcal{X}$ obtained from μ by associating samples of the forms $(x, 1)$ and $(x, -1)$ with x and $-x$, respectively. Since the matrix $W \in \mathbb{R}^{C \times \mathcal{X}}$ obtained from D is symmetric, and estimating $\max_{q \in Q} q(\mu')$ is precisely what is required for the lower bound of Theorem 22, we the following lower bound for agnostic learning.

Theorem 25. *Let $\beta \in (0,1)$ be a small enough constant, and let $\varepsilon > 0$. Let C be a concept class with concept matrix $D \in \mathbb{R}^{C \times \mathcal{X}^+}$. For some $\alpha' = \Omega\left(\frac{\alpha}{\log(1/\alpha)}\right)$, if $\frac{\gamma_2(W,\alpha)}{\varepsilon^2 \alpha^2} \geq \frac{C \log 2k}{\alpha'}$ for a large enough constant $C > 0$, then any ε-LDP mechanism $\mathcal{M}$ which (α', β)-learns C requires*

$$n = \Omega\left(\frac{\gamma_2(W,\alpha)^2}{\varepsilon^2 \alpha^2}\right)$$

samples as input.

5 CHARACTERIZING CENTRAL DP FOR LARGE DATASETS

The goal of this section is to show that the sample complexity of releasing a given set of linear queries with workload matrix W is

$$\mathrm{sc}^{\ell_2^2}(W, \alpha, \varepsilon, \delta) = \Theta\left(\frac{\gamma_F(W)}{\alpha \varepsilon}\right)$$

when α is sufficiently small (smaller than some $\alpha^*(Q, \varepsilon)$). Or, equivalently, we show that $\mathrm{err}^{\ell_2^2}(W, n, \varepsilon, \delta) = \Theta(\frac{\gamma_F(W)}{\varepsilon n})$, when n is sufficiently large (larger than some $n^*(Q, \varepsilon)$).

The proof consists of two steps. First, we argue that error

$$\mathrm{err}(W, n, \varepsilon, \delta) = \Theta(\frac{\gamma_F(W)}{\varepsilon n})$$

is necessary for *every n* if we restrict attention only to mechanisms that are *data-independent*. That is, mechanisms that perturb the output with noise from a fixed distribution independent of the dataset. Then, we apply a lemma of Bhaskara et al. [BDKT12] that says, when n is sufficiently large, any instance-dependent mechanism can be replaced with an instance-independent mechanism with the same error and similar privacy parameters.

5.1 Data-Independent Mechanisms

Let Q be a workload of linear queries over data universe $\mathcal{X}$ and let $W \in \mathbb{R}^{Q \times \mathcal{X}}$ be the matrix form of this workload. An *instance-independent* mechanism $\mathcal{M}$ can be written (as a function of the histogram of the dataset) as,

$$\mathcal{M}(h) = \frac{1}{n}(Wh + Z)$$

where Z is a random variable over $\mathbb{R}^Q$ whose distribution does not depend on h. Without loss of generality, we assume $\mathbb{E}[Z] = 0$. Let $\Sigma = \mathbb{E}\left[ZZ^T\right]$ be the covariance matrix of Z. Then the ℓ_2 error of such a mechanism is

$$\begin{aligned}\mathrm{err}^{\ell_2^2}(\mathcal{M}, W, n) &= \max_{h : \|h\|_1 = n} \sqrt{\mathbb{E}\left[\frac{\|\mathcal{M}(h) - \frac{1}{n}Wh\|_2^2}{|Q|}\right]} \\ &= \sqrt{\mathbb{E}\left[\frac{\|Z\|_2^2}{n|Q|}\right]} = \sqrt{\frac{\mathrm{Tr}(\Sigma)}{n|Q|}}\end{aligned}$$

In this section, we will show that, if $\mathcal{M}$ is (ε, δ)-differentially private (for ε, δ smaller than some absolute constants), then $\mathrm{Tr}(\Sigma) = \Omega(\frac{|Q|\gamma_F(W)^2}{\varepsilon^2})$, and thus $\mathrm{err}(\mathcal{M}, W, n) = \Omega(\frac{\gamma_F(W)}{\varepsilon n})$.

We start with the following basic lemma about differential privacy, which says that the variance of any differentially private algorithm for answering a single query w must be proportional to the sensitivity of the query.

Lemma 26 ([KRSU10]). *For any single-query workload $w \in \mathbb{R}^{|\mathcal{X}|}$, and any data-independent mechanism $\mathcal{M}(h) = \frac{1}{n} w^\top h + \frac{1}{n} z$ that is (ε, δ)-differentially private for ε, δ smaller than some absolute constants, $\mathbb{E}\left[z^2\right] \geq \frac{1}{C\varepsilon}\|w\|_\infty$ for some absolute constant $C > 0$.*

Next, we define the *sensitivity polytope* $K = WB_1^{|\mathcal{X}|}$, where $B_1^{|\mathcal{X}|} = \{h \in \mathbb{R}^{|\mathcal{X}|} : \|h\|_1 \leq 1\}$. With this definition, we have that for any pair of neighboring datasets X, X' with associated histograms h, h', we have $W(h - h') \in K$. The next lemma says that the covariance matrix Σ defines an ellipsoid that contains at least a constant multiple of the sensitivity polytope.

Lemma 27. *Let W be a workload matrix such that the sensitivity polytope K is full dimensional. Let $\mathcal{M}$ be an (ε, δ)-differentially private data-independent mechanism for W that has covariance matrix Σ, for ε, δ smaller than some absolute constants. Then Σ is invertible, and*

$$\max_{y \in K} \|\Sigma^{-1/2} y\|_2^2 = \max_{y \in K} y^\top \Sigma^{-1} y \leq C^2 \varepsilon^2$$

for some absolute constant $C > 0$.

Proof. By post-processing, for any $u \in \mathbb{R}^{|\mathcal{X}|}$,

$$u^\top \mathcal{M}(h) = \frac{1}{n} u^\top Wh + \frac{1}{n} u^\top Z$$

is an (ε, δ)-DP mechanism for the single query $u^\top W$. The sensitivity polytope of the workload $u^\top W$ is the line $[-h_K(u), h_K(u)]$, where $h_K(u) = \max_{y \in K} u^\top y$ is the support function. By Lemma 26, if $\mathcal{M}$ is an (ε, δ)-differentially private mechanism, then for some constant C,

$$\|\Sigma^{1/2} u\|_2 = \sqrt{u^\top \Sigma u} \geq \frac{h_K(u)}{C\varepsilon}. \tag{12}$$

If K is full dimensional, then in particular we have $h_K(e_i) > 0$ for any standard basis vector e_i, which implies that the matrix Σ is positive definite and invertible.

By change of variables we can write $v = \Sigma^{1/2}u$ and rewrite (12) as

$$\|v\|_2 \ge \frac{1}{C\varepsilon}\cdot h_K(\Sigma^{-1/2}v) = \frac{1}{C\varepsilon}\cdot\max_{y\in K}(\Sigma^{-1/2}v)^\top y = \frac{1}{C\varepsilon}\cdot\max_{y\in K} v^\top\Sigma^{-1/2}y$$

Since the above holds for any unit vector $v \in \mathbb{S}^{|\mathcal{X}|-1}$, we have

$$\begin{aligned}\max_{y\in K}\|\Sigma^{-1/2}y\|_2 &= \max_{y\in K}\max_{v\in\mathbb{S}^{|\mathcal{X}|-1}} v^\top\Sigma^{-1/2}y\\ &= \max_{v\in\mathbb{S}^{|\mathcal{X}|-1}}\max_{y\in K} v^\top\Sigma^{-1/2}y \le C\varepsilon\end{aligned}$$

where the first equality is the equality-case of Cauchy-Schwarz. □

Recall that for a matrix $W \in \mathbb{R}^{\mathcal{Q}\times\mathcal{X}}$,

$$\gamma_F(W) = \inf\left\{\tfrac{1}{|\mathcal{Q}|^{1/2}}\|R\|_F\|A\|_{1\to 2} : RA = W\right\}.$$

We now prove our main result, which shows that the error of data-independent private mechanisms must be proportional to $\gamma_F(W)$.

Theorem 28. *Let W be a workload matrix. Let $\mathcal{M}$ is a (ε,δ)-differentially private data-independent mechanism for W with covariance matrix Σ, for ε,δ smaller than some absolute constants. Then*

$$\mathrm{err}^{\ell_2^2}(\mathcal{M}, W, n) = \Omega\left(\frac{\gamma_F(W)}{C\varepsilon n}\right).$$

Proof. Let $w_1,\dots,w_{|\mathcal{X}|}$ be the columns of the workload matrix W. Let $A = \Sigma^{-1/2}W$ with columns $a_1,\dots,a_{|\mathcal{X}|}$ and let $R = \Sigma^{1/2}$ so that $RA = W$. By Lemma 27, the matrix A is well defined, and for every i, $\|a_i\| = \|\Sigma^{-1/2}w_i\|_2 \le C\varepsilon$. Hence $\|A\|_{1\to2} \le C\varepsilon$. We also have

$$\|R\|_F = \mathrm{Tr}(R^\top R)^{1/2} = \mathrm{Tr}(\Sigma)^{1/2} = |\mathcal{Q}|^{1/2}\cdot n\cdot \mathrm{err}^{\ell_2^2}(\mathcal{M}, W, n).$$

Combining the inequalities, we get

$$\gamma_F(W) \le \frac{1}{|\mathcal{Q}|^{1/2}}\|R\|_F\|A\|_{1\to2} \le C\varepsilon\cdot n\cdot\mathrm{err}^{\ell_2^2}(\mathcal{M}, W, n).$$

The theorem follows from rearranging this inequality. □

5.2 From Data-Dependent to Data-Independent Mechanisms

In this section we describe a reduction of Bhaskara et al. [BDKT12] showing, in the case of symmetric workloads, that any data-dependent mechanism with small error for datasets of arbitrary size can be converted into a data-independent mechanism with approximately the same error.

Lemma 29 ([BDKT12]). *Let $W \in \mathbb{R}^{\mathcal{Q}\times\mathcal{X}}$ be a symmetric workload matrix. For every (ε,δ)-differentially private mechanism $\mathcal{M}$, there exists a $(2\varepsilon, 2e^{\varepsilon}\delta)$-differentially private data-independent mechanism $\mathcal{M}'$ such that*

$$\mathrm{err}^{\ell_2^2}(\mathcal{M}', W, n) \le \frac{1}{n}\max_{m\in\mathbb{N}}(m\cdot\mathrm{err}^{\ell_2^2}(\mathcal{M}, W, m))$$

As an immediate, corollary, lower bounds for data-independent mechanisms imply lower bounds for arbitrary data-dependent mechanisms for some dataset size n^*. Thus we obtain the following theorem by combining Theorem 28 with Lemma 29.

Theorem 30. *Let Q be linear queries with symmetric workload matrix $W \in \mathbb{R}^{\mathcal{Q}\times\mathcal{X}}$. Then for every ε,δ smaller than some absolute constants, there exists $n^* \in \mathbb{N}$ such that*

$$\forall n \le n^*\ \mathrm{err}^{\ell_2^2}_{\varepsilon,\delta}(Q, n) \ge \frac{\gamma_F(W)}{C\varepsilon n}.$$

By standard transformations (see e.g. [BUV14]), we can convert this to the following sample complexity lower bound,

Corollary 31. *Let Q be linear queries with symmetric workload matrix $W \in \mathbb{R}^{\mathcal{Q}\times\mathcal{X}}$. Then for every ε,δ smaller than some absolute constants, there exists $\alpha^* > 0$ such that*

$$\forall\alpha \le \alpha^*\ \mathrm{sc}^{\ell_2^2}_{\varepsilon,\delta}(Q, \alpha) \ge \frac{\gamma_F(W)}{C\varepsilon\alpha}$$

We remark that our lower bounds may be extended to case of non-symmetric workloads by using the same technique which we used to obtain Lemma 17. An advantage of performing this reduction in the central model is that we may take advantage of the central model version of the Laplace mechanism, which will use only $n = O\left(\frac{1}{\alpha\varepsilon}\right)$ rather than $n = O\left(\frac{1}{\alpha^2\varepsilon^2}\right)$ samples. In this way, Theorem 28, Theorem 30, and Corollary 31 may be obtained under the additional assumption that $\gamma_F(W) > D$ for a sufficiently large constant $D > 0$.

We also note that Theorem 28, Theorem 30, and Corollary 31 may be extended to ℓ_∞^2-error, defined by

$$\mathrm{err}^{\ell_\infty^2}(\mathcal{M}, Q, n) = \max_{X\in\mathcal{X}^n}\mathop{\mathbb{E}}_{\mathcal{M}}\left[\|\mathcal{M}(X) - Q(X)\|_\infty^2\right]^{1/2},$$

with $\mathrm{err}^{\ell_\infty^2}_{\varepsilon,\delta}(Q, n)$, and $\mathrm{sc}^{\ell_\infty^2}_{\varepsilon,\delta}(Q, \alpha)$ defined analogously, and $\gamma_F(W)$ replaced by $\gamma_2(W)$ in the lower bounds.

ACKNOWLEDGMENTS

Part of this work was done while the authors were visiting the Simons Institute for Theory of Computing. We are grateful to Toniann Pitassi for many helpful discussions about local differential privacy. AE and AN were supported by an Ontario Early Researcher Award, and an NSERC Discovery Grant. JU was supported by NSF grants CNS-1718088, CCF-1750640 and CNS-1816028, and a Google Faculty Research Award.

REFERENCES

[App17] Apple Differential Privacy Team. Learning with privacy at scale. Apple Machine Learning Journal, 2017.

[BBNS19] Jaroslaw Blasiok, Mark Bun, Aleksandar Nikolov, and Thomas Steinke. Towards instance-optimal private query release. In *SODA*, pages 2480–2497. SIAM, 2019.

[BCD+07] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the 26th ACM Symposium on Principles of Database Systems*, PODS '07, pages 273–282. ACM, 2007.

[BDKT12] Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*, STOC '12, pages 1269–1284, 2012.

[BEM+17] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, pages 441–459. ACM, 2017.

[BFJ+94] Avrim Blum, Merrick L. Furst, Jeffrey C. Jackson, Michael J. Kearns, Yishay Mansour, and Steven Rudich. Weakly learning DNF and characterizing statistical query learning using fourier analysis. In *Proceedings* *of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 253–262, 1994.

[BNS18] Mark Bun, Jelani Nelson, and Uri Stemmer. Heavy hitters and the structure of local privacy. In *Proceedings of the 37th ACM Symposium on Principles of Database Systems*, PODS'18, pages 435–447. ACM, 2018.

[BUV14] Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *46th Annual ACM Symposium on the Theory of Computing*, STOC '14, pages 1–10, New York, NY, USA, 2014.

[CSS11] T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):26, 2011.

[CTUW14] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. In *Proceedings of the 5th ACM Conference on Innovations in Theoretical Computer Science*, ITCS '14, pages 287–402, Princeton, NJ, 2014. ACM.

[DJW18] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.*, 113(521):182–201, 2018.

[DLS+17] Aref N. Dajani, Amy D. Lauger, Phyllis E. Singer, Daniel Kifer, Jerome P. Reiter, Ashwin Machanavajjhala, Simson L. Garfinkel, Scot A. Dahl, Matthew Graham, Vishesh Karwa, Hang Kim, Philip Lelerc, Ian M. Schmutte, William N. Sexton, Lars Vilhuber, and John M. Abowd. The modernization of statistical disclosure limitation at the U.S. census bureau, 2017. Presented at the September 2017 meeting of the Census Scientific Advisory Committee.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, Berlin, Heidelberg, 2006. Springer.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the 22nd ACM Symposium on Principles of Database Systems*, PODS '03, pages 202–210. ACM, 2003.

[DN04] Cynthia Dwork and Kobbi Nissim. Privacy-preserving datamining on vertically partitioned databases. In *Annual International Cryptology Conference*, pages 528–544. Springer, 2004.

[DNPR10] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *Symposium on Theory of Computing (STOC)*, pages 715–724. ACM, 2010.

[DNT15] Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. *Discrete & Computational Geometry*, 53(3):650–673, 2015.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[DR18] John C. Duchi and Feng Ruan. The right complexity measure in locally private estimation: It is not the fisher information. *arXiv preprint arXiv:1806.05756*, 2018.

[EGS03] Alexandre V. Evfimievski, Johannes Gehrke, and Ramakrishnan Srikant. Limiting privacy breaches in privacy preserving data mining. In *PODS*, pages 211–222. ACM, 2003.

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, CCS'14. ACM, 2014.

[FSSS03] Jürgen Forster, Niels Schmitt, Hans Ulrich Simon, and Thorsten Suttorp. Estimating the optimal margins of embeddings in euclidean half spaces. *Machine Learning*, 51(3):263–281, 2003.

[GHRU11] Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan Ullman. Privately releasing conjunctions and the statistical query barrier. In *Proceedings of the 43rd ACM Symposium on Theory of Computing*, STOC '11, pages 803–812, San Jose, CA, 2011.

[Gro53] A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

[HRS12] Moritz Hardt, Guy N. Rothblum, and Rocco A. Servedio. Private data release via learning thresholds. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 168–187, 2012.

[HT10] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC*, 2010.

[JNS18] Noah Johnson, Joseph P Near, and Dawn Song. Towards practical differential privacy for sql queries. *Proceedings of the VLDB Endowment*, 11(5):526–539, 2018.

[Kea93] Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. In *STOC*, pages 392–401. ACM, May 16-18 1993.

[KLN+08] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? In *FOCS*, pages 531–540. IEEE, Oct 25–28 2008.

[KN12] Subhash Khot and Assaf Naor. Grothendieck-type inequalities in combinatorial optimization. *Comm. Pure Appl. Math.*, 65(7):992–1035, 2012.

[KRSU10] Shiva Prasad Kasiviswanathan, Mark Rudelson, Adam Smith, and Jonathan Ullman. The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 775–784. ACM, 2010.

[KSS94] Michael J Kearns, Robert E Schapire, and Linda M Sellie. Toward efficient agnostic learning. *Machine Learning*, 17(2-3):115–141, 1994.

[LHR+10] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. Optimizing linear counting queries under differential privacy. In *Proceedings of the 29th ACM Symposium on Principles of Database Systems*, PODS'10, pages 123–134. ACM, 2010.

[LMSS07] Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, 2007.

[LS09] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Struct. Algorithms*, 34(3):368–394, 2009.

[MMHM18] Ryan McKenna, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. Optimizing error of high-dimensional statistical queries under differential privacy. *Proceedings of the VLDB Endowment*, 11(10):1206–1219, 2018.

[Nik14] Aleksandar Nikolov. *New Computational Aspects of Discrepancy Theory*. PhD thesis, Rutgers, The State University of New Jersey, 2014.

[Nik15] Aleksandar Nikolov. An improved private mechanism for small databases. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP*, pages 1010–1021, 2015.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[NT15] Aleksandar Nikolov and Kunal Talwar. Approximating hereditary discrepancy via small width ellipsoids. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA'15, pages 324–336. SIAM, 2015.

[NTZ16] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: The small database and approximate cases. *SIAM J. Comput.*, 45(2):575–616, 2016.

[Pis12] Gilles Pisier. Grothendieck's theorem, past and present. *Bull. Amer. Math. Soc. (N.S.)*, 49(2):237–323, 2012.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[Szö09] Balázs Szörényi. Characterizing statistical query learning: Simplified notions and proofs. In *ALT*, volume 5809 of *Lecture Notes in Computer Science*, pages 186–200. Springer, 2009.

[TUV12] Justin Thaler, Jonathan Ullman, and Salil P. Vadhan. Faster algorithms for privately releasing marginals. In *39th International Colloquium on Automata, Languages, and Programming -*, ICALP '12, pages 810–821, Warwick, UK, 2012. Springer.

[WZL+19] Royce J Wilson, Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel Simmons-Marengo, and Bryant Gipson. Differentially private sql with bounded user contribution. *arXiv preprint arXiv:1909.01917*, 2019.