Lower Bounds for Differential Privacy from Gaussian Width[*]

Assimakis Kattis[1] and Aleksandar Nikolov[2]

1 **Department of Computer Science, University of Toronto, Toronto, ON, Canada**
`kattis@cs.toronto.edu`

2 **Department of Computer Science, University of Toronto, Toronto, ON, Canada**
`anikolov@cs.toronto.edu`

Abstract

We study the optimal sample complexity of a given workload of linear queries under the constraints of differential privacy. The sample complexity of a query answering mechanism under error parameter α is the smallest n such that the mechanism answers the workload with error at most α on any database of size n. Following a line of research started by Hardt and Talwar [STOC 2010], we analyze sample complexity using the tools of asymptotic convex geometry. We study the sensitivity polytope, a natural convex body associated with a query workload that quantifies how query answers can change between neighboring databases. This is the information that, roughly speaking, is protected by a differentially private algorithm, and, for this reason, we expect that a "bigger" sensitivity polytope implies larger sample complexity. Our results identify the *mean Gaussian width* as an appropriate measure of the size of the polytope, and show sample complexity lower bounds in terms of this quantity. Our lower bounds completely characterize the workloads for which the Gaussian noise mechanism is optimal up to constants as those having asymptotically maximal Gaussian width.

Our techniques also yield an alternative proof of Pisier's Volume Number Theorem which also suggests an approach to improving the parameters of the theorem.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases differential privacy, convex geometry, lower bounds, sample complexity

Digital Object Identifier 10.4230/LIPIcs.SoCG.2017.45

1 Introduction

The main goal of private data analysis is to estimate aggregate statistics while preserving individual privacy guarantees. Intuitively, we expect that, for statistics that do not depend too strongly on any particular individual, a sufficiently large database allows computing an estimate that is both accurate and private. A natural question then is to characterize the *sample complexity* under privacy constraints: the smallest database size for which we can privately estimate the answers to a given collection of queries within some allowable error tolerance. Moreover, it is desirable to identify algorithms that are simple, efficient, and have close to the best possible sample complexity. In this work, we study these questions for collections of *linear queries* under the constraints of *approximate differential privacy.*

We model a *database* $\mathcal{D}$ of *size* n as a multiset of n elements (counted with repetition) from an arbitrary finite universe $\mathcal{U}$. Each element of the database corresponds to the data of a single individual. To define a privacy-preserving computation on $\mathcal{D}$, we use the strong

[*] A full version of the paper is available at `https://arxiv.org/abs/1612.02914`.

33rd International Symposium on Computational Geometry (SoCG 2017).
Editors: Boris Aronov and Matthew J. Katz; Article No. 45; pp. 45:1–45:16
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

notion of *differential privacy*. Informally, an algorithm is differentially private if it has almost identical behavior on any two databases $\mathcal{D}$ and $\mathcal{D}'$ that differ in the data of a single individual. To capture this concept formally, let us define two databases $\mathcal{D}$ and $\mathcal{D}'$ to be *neighboring* if we can get $\mathcal{D}'$ by replacing a single element of $\mathcal{D}$ with another element from $\mathcal{U}$. Then differential privacy is defined as follows:

▶ **Definition 1** ([5]). A randomized algorithm $\mathcal{A}$ that takes as input a database and outputs a random element from the set Y satisfies (ε, δ)-*differential privacy* if for all neighboring databases $\mathcal{D}, \mathcal{D}^{'}$ and all measurable $S \subseteq Y$ we have that:

$$\mathbb{P}[\mathcal{A}(\mathcal{D}) \in S] \leq e^{\varepsilon}\mathbb{P}[\mathcal{A}(\mathcal{D}^{'}) \in S] + \delta,$$

where probabilities are taken with respect to the randomness of $\mathcal{A}$.

One of the most basic primitives in private data analysis, and data analysis in general, are *counting queries* and, slightly more generally, *linear queries*. While interesting and natural in themselves, they are also quite powerful: any statistical query (SQ) learning algorithm can be implemented using noisy counting queries as a black box [10]. In our setting, we specify a linear query by a function $q\colon \mathcal{U} \to [0, 1]$ (given by a table of its values for each element of $\mathcal{U}$). Slightly abusing notation, we define the value of the query as $q(\mathcal{D}) = \frac{1}{n}\sum_{e\in\mathcal{D}} q(e)$, where the elements of $\mathcal{D}$ are counted with multiplicity, and n is the size of $\mathcal{D}$. For example, when $q\colon \mathcal{U} \to \{0, 1\}$, we can think of q as a property defined on $\mathcal{U}$ and $q(\mathcal{D})$ as the fraction of elements of $\mathcal{D}$ that satisfy the property: this is a *counting query*. We call a set $\mathcal{Q}$ of linear queries a *workload* and an algorithm that answers a query workload a *mechanism*. We denote by $\mathcal{Q}(\mathcal{D}) = (q(\mathcal{D}))_{q\in\mathcal{Q}}$ the vector of answers to the queries in $\mathcal{Q}$. Throughout the paper, we will use the letter m for the size of a workload $\mathcal{Q}$.

Starting from the work of Dinur and Nissim [4], it is known that we cannot hope to answer too many linear queries too accurately while preserving even a very weak notion of privacy. For this reason, we must allow our private mechanisms to make some error. We focus on *average error* (in an L_2 sense). We define the average error of an algorithm $\mathcal{A}$ on a query workload $\mathcal{Q}$ and databases of size at most n as:

$$\mathrm{err}(\mathcal{Q}, \mathcal{A}, n) = \max_{\mathcal{D}} \left(\mathbb{E} \sum_{q\in\mathcal{Q}} \frac{(\mathcal{A}(\mathcal{D})_q - q(D))^2}{|\mathcal{Q}|} \right)^{1/2} = \max_{\mathcal{D}} \left(\mathbb{E}\frac{1}{m} \|\mathcal{A}(\mathcal{D}) - \mathcal{Q}(\mathcal{D})\|_2^2 \right)^{1/2},$$

where the maximum is over all databases $\mathcal{D}$ of size at most n, $\mathcal{A}(\mathcal{D})_q$ is the answer to query q given by the algorithm $\mathcal{A}$ on input $\mathcal{D}$, and expectations are taken with respect to the random choices of $\mathcal{A}$. This is a natural notion of error that also works particularly well with the geometric tools that we use.

In this work we study *sample complexity*: the smallest database size which allows us to answer a given query workload with error at most α. The sample complexity of an algorithm $\mathcal{A}$ with error α is defined as:

$$\mathrm{sc}(\mathcal{Q}, \mathcal{A}, \alpha) = \min\{n : \mathrm{err}(\mathcal{Q}, \mathcal{A}, n) \leq \alpha\}.$$

The sample complexity of answering the linear queries $\mathcal{Q}$ with error α under (ε, δ)-differential privacy is defined by:

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q}, \alpha) = \inf\{\mathrm{sc}(\mathcal{Q}, \mathcal{A}, \alpha) : \mathcal{A} \text{ is } (\varepsilon, \delta)\text{-differentially private}\}.$$

The two main questions we are interested in are:

1. Can we characterize $\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha)$ in terms of a natural property of the workload $\mathcal{Q}$?
2. Can we identify conditions under which simple and efficient (ε,δ)-differentially private mechanisms have nearly optimal sample complexity?

We make progress on both questions. We identify a geometrically defined property of the workload that gives lower bounds on the sample complexity. The lower bounds also *characterize* when one of the simplest differentially private mechanisms, the Gaussian noise mechanism, has nearly optimal sample complexity in the regime of constant α.

Before we can state our results, we need to define a natural geometric object associated with a workload of linear queries. This object has been important in applying geometric techniques to differential privacy [9, 2, 16, 15].

▶ **Definition 2.** The *sensitivity polytope* K of a workload $\mathcal{Q}$ of m linear queries is equal to $K = \mathrm{conv}\{\pm\mathcal{Q}(\mathcal{D}) : \mathcal{D} \text{ is a database of size } 1\}$.

From the above definition, we see that K is a symmetric (i.e. $K = -K$) convex polytope in $\mathbb{R}^m$. The importance of K lies in the fact that it captures how query answers can change between neighboring databases: for any two neighboring databases $\mathcal{D}$ and $\mathcal{D}'$ of size n and n' respectively, $n\mathcal{Q}(\mathcal{D}) - n'\mathcal{Q}(\mathcal{D}') \in K$. This is exactly the information that a differentially private algorithm is supposed to hide. Intuitively, we expect that the larger K is, the larger $\mathrm{sc}_{\varepsilon,\delta}(K,\alpha)$ should be.

We give evidence for the above intuition, and propose the width of K in a random direction as a measure of its "size". Let h_K be the support function of K: $h_K(y) = \max_{x\in K}\langle x, y\rangle$. For a unit vector y, $h_K(y) + h_K(-y)$ is the width of K in the direction of y; for arbitrary y, $h_K(ty)$ scales linearly with t (and is, in fact, a norm). We define the ℓ^*-norm of K, also known as its *Gaussian mean width*, as $\ell^*(K) = \mathbb{E}[h_K(g)]$, where g is a standard Gaussian random vector in $\mathbb{R}^m$. The Gaussian mean width is closely related to the mean width $w(K) = \mathbb{E}\frac{h_K(y)+h_K(-y)}{2}$, where y is distributed according to the rotation invariant probability measure on the unit sphere. It's easy to show that $\ell^*(K) = \mathbb{E}\|g\|_2 w(K)$, where $g \sim N(0, I)$, and $\mathbb{E}\|g\|_2 = \Theta(\sqrt{m})$. The Gaussian mean width of the Euclidean unit ball B_2^m is $\Theta(\sqrt{m})$, and, since $h_K(y) \le h_{B_2^m}(y)$ for any $y \in \mathbb{R}^m$ and any $K \subseteq B_2^m$, we have $\ell^*(K) = O(\sqrt{m})$ for any such K.

The following theorem captures our main result.

▶ **Theorem 3.** *Let $\mathcal{Q}$ be a workload of m linear queries, and let K be its sensitivity polytope. The following holds for all $\varepsilon = O(1)$, $2^{-\Omega(n)} \le \delta \le 1/n^{1+\Omega(1)}$, and any $\alpha \le \frac{\ell^*(K)}{Cm(\log 2m)^2}$, where C is an absolute constant, and $\sigma(\varepsilon,\delta) = (0.5\sqrt{\varepsilon} + \sqrt{2\log(1/\delta)})/\varepsilon$:*

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha) = O\left(\min\left\{\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\sqrt{m}\alpha^2}, \frac{\sigma(\varepsilon,\delta)\sqrt{m}}{\alpha}\right\}\right);$$
$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)^2}{m^{3/2}(\log 2m)^4\alpha}\right).$$

The upper bound on the sample complexity is achieved by a mechanism running in time polynomial in m, n, and $|\mathcal{U}|$. Moreover, if $\ell^*(K) = \Omega(m)$, then $\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha) = \Theta\left(\frac{\sigma(\varepsilon,\delta)\sqrt{m}}{\alpha}\right)$ for any $\alpha \le 1/C$, where C is an absolute constant.

The sample complexity upper bounds in the theorem above are known from prior work: one is given by the projection mechanism from [16], with the sample complexity upper bound in terms of $\ell^*(K)$ shown in [6]; the other upper bound is given by the Gaussian noise mechanism [4, 7, 5]. The main new contribution in this work are the lower bounds on the sample complexity. The gap between upper and lower bounds is small when $\ell^*(K)$ is close to its maximal value of m. Indeed, when $\ell^*(K) = \Theta(m)$, our results imply that the Gaussian noise mechanism has optimal sample complexity up to constants. This is, to the best of our knowledge, the first example of a general geometric condition under which a simple and efficient mechanism has optimal sample complexity up to *constant* factors. Moreover, in the constant error regime this condition is also *necessary* for the Gaussian mechanism to be optimal up to constants: when $\ell^*(K) = o(m)$ and $\alpha = \Omega(1)$, the projection mechanism has asymptotically smaller sample complexity than the Gaussian mechanism.

We can prove somewhat stronger results for another natural problem in private data analysis, which we call the *mean point problem.* In this problem, we are given a closed convex set $K \subset \mathbb{R}^m$, and we are asked to approximate the mean $\overline{\mathcal{D}}$ of the database $\mathcal{D}$, where $\mathcal{D} = \{x_1, \dots, x_n\}$ is a multiset of points in K and $\overline{\mathcal{D}} = \frac{1}{n}\sum_{i=1}^n x_i$. This problem, which will be the focus for most of this paper, has a more geometric flavor, and is closely related to the query release problem for linear queries. In fact, Theorem 3 will essentially follow from a reduction from the results below for the mean point problem.

With respect to the mean point problem, we define the error of an algorithm $\mathcal{A}$ as:

$$\mathrm{err}(K, \mathcal{A}, n) = \sup_{\mathcal{D}} (\mathbb{E}\|\mathcal{A}(\mathcal{D}) - \overline{\mathcal{D}}\|_2^2)^{1/2},$$

where the supremum is over databases $\mathcal{D}$ consisting of at most n points from K, and the expectation is over the randomness of the algorithm. The sample complexity of an algorithm $\mathcal{A}$ with error α is defined as:

$$\mathrm{sc}(K, \mathcal{A}, \alpha) = \min\{n : \mathrm{err}(K, \mathcal{A}, n) \le \alpha\}.$$

The sample complexity of solving the mean point problem with error α over K is defined by:

$$\mathrm{sc}_{\varepsilon,\delta}(K, \alpha) = \min\{\mathrm{sc}(K, \mathcal{A}, \alpha) : \mathcal{A} \text{ is } (\varepsilon, \delta)\text{-differentially private}\}.$$

Our main result for the mean point problem is given in the following theorem:

▶ **Theorem 4.** *Let K be a symmetric convex body contained in the unit Euclidean ball B_2^m in $\mathbb{R}^m$. The following holds for all $\varepsilon = O(1)$, $2^{-\Omega(n)} \le \delta \le 1/n^{1+\Omega(1)}$, and any $\alpha \le \frac{\ell^*(K)}{C\sqrt{m}(\log 2m)^2}$, where C is an absolute constant, and $\sigma(\varepsilon, \delta) = (0.5\sqrt{\varepsilon} + \sqrt{2\log(1/\delta)})/\varepsilon$:*

$$\mathrm{sc}_{\varepsilon,\delta}(K, \alpha) = O\left(\min\left\{\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\alpha^2}, \frac{\sigma(\varepsilon,\delta)\sqrt{m}}{\alpha}\right\}\right);$$
$$\mathrm{sc}_{\varepsilon,\delta}(K, \alpha) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{(\log 2m)^2\alpha}\right).$$

The upper bound on the sample complexity is achieved by a mechanism running in time polynomial in m, n, and $|\mathcal{U}|$. Moreover, when $\ell^*(K) = \Omega(\sqrt{m})$, then $\mathrm{sc}_{\varepsilon,\delta}(K, \alpha) = \Theta\left(\frac{\sigma(\varepsilon,\delta)\sqrt{m}}{\alpha}\right)$ for any $\alpha \le 1/C$, where C is an absolute constant.

The upper bounds again follow from prior work, and in fact are also given by the projection mechanism and the Gaussian noise mechanism, which can be defined for the mean point problem as well. Notice that the gap between the upper and the lower bound is on the order $\frac{(\log 2m)^2}{\alpha}$. If the lower bound was valid for all values of the error parameter α less than a fixed constant, rather than for $\alpha \le \frac{\ell^*(K)}{C\sqrt{m}(\log 2m)^2}$, Theorem 4 would nearly characterize the optimal sample complexity for the mean point problem for all constant α. Unfortunately, the restriction on α is, in general, necessary (up to the logarithmic terms) for lower bounds on the sample complexity in terms of $\ell^*(K)$. For example, we can take $K = \gamma B_2^m$, i.e. a Euclidean ball in $\mathbb{R}^m$ with radius γ. Then, $\ell^*(K) = \Theta(\gamma\sqrt{m})$, but the sample complexity is 0 when $\alpha > \gamma$, since the trivial algorithm which ignores the database and outputs 0 achieves error γ. Thus, a more sensitive measure of the size of K is necessary to prove optimal lower bounds. We do, nevertheless, trust that the techniques introduced in this paper bring us closer to this goal.

We conclude this section with a high-level overview of our techniques. Our starting point is a recent tight lower bound on the sample complexity of a special class of linear queries: the 1-way marginal queries. These queries achieve the worst case sample complexity for a family of m linear queries: $\Omega(\sqrt{m}/\alpha)$ [3, 22]. The sensitivity polytope of the 1-way marginals is the cube $[-1,1]^m$, and it can be shown that the lower bound on the sample complexity of 1-way marginals implies an analogous lower bound on the sample complexity of the mean point problem with $K = Q^m = [-1/\sqrt{m}, 1/\sqrt{m}]^m$. For the mean point problem, it is easy to see that when $K' \subseteq K$, the sample complexity for K' is no larger than the sample complexity for K. Moreover, we can show that the sample complexity of any projection of K is no bigger than the sample complexity of K itself. So, our strategy then is to find a large scaled copy of $Q^{m'}$, $m' \leq m$, inside a projection of K onto a large dimensional subspace whenever $\ell^*(K)$ is large. We solve this geometric problem using deep results from asymptotic convex geometry, namely the Dvoretzky criterion, the low M^* estimate, and the MM^* estimate.

We note that in [3], the authors mention a similar idea of showing a lower bound on the sample complexity of an arbitrary family of *counting* queries $\mathcal{Q}$ by embedding the 1-way marginals into $\mathcal{Q}$. For average error, their approach gives a lower bound of $\Omega(\frac{\sqrt{d}}{\alpha})$ (ignoring the dependence on ε and δ) for any $\alpha \leq \frac{1}{10}\sqrt{\frac{d}{m}}$, where d is the the VC-dimension of the set system $\{S_e : e \in \mathcal{U}\}$ and $S_e = \{q \in \mathcal{Q} : q(e) = 1\}$. This lower bound is at least as strong as our lower bounds, since $\frac{\ell^*(K)}{\sqrt{m}} \leq C\sqrt{d}$ for a sufficiently large constant C. (This inequality is a well-known consequence of Dudley's chaining inequality and estimates on entropy numbers in terms of VC-dimension, e.g. Theorem 14.12. in [11].) Our results, however, hold for arbitrary linear queries, and not just counting queries. Moreover, our lower bound is in terms of the efficiently computable quantity $\ell^*(K)$, while there is evidence that computing VC-dimension is hard [18]. Thus, our lower bound can be seen as an efficiently computable relaxation of the VC-dimension lower bound, and also as a generalization of it to linear queries.

Our techniques also yield an alternative proof of the volume number theorem of Milman and Pisier [14]. Besides avoiding the quotient of subspace theorem, our proof yields an improvement in the volume number theorem, conditional on the well-known conjecture (see e.g. Chapter 6 in [1]) that any symmetric convex body K has a position (affine image) TK for which $\ell^*(TK)\ell(TK) = O(m\sqrt{\log 2m})$, where $\ell(K)$ is the expected K-norm of a standard Gaussian. More details about this connection are given in Section 6.

1.1 Prior Work

Most closely related to our work are the results of Nikolov, Talwar, and Zhang [16], who gave a private mechanism (also based on the projection mechanism, but more involved) which has nearly optimal sample complexity (with respect to average error), up to factors polynomial in $\log m$ and $\log|\mathcal{U}|$. This result was subsequently improved by Nikolov [15], who showed that the $\log m$ factors can be replaced by $\log n$. While these results are nearly optimal for subconstant values of the error parameter α, i.e. the optimality guarantees do not depend on $1/\alpha$, factors polynomial in $\log|\mathcal{U}|$ can be prohibitively large. Indeed, in many natural settings, such as that of marginal queries, $|\mathcal{U}|$ is exponential in the number of queries m, so the competitiveness ratio can be polynomial in m.

The line of work that applies techniques from convex geometry to differential privacy started with the beautiful paper of Hardt and Talwar [9], whose results were subsequently strengthened in [2]. These papers focused on the "large database" regime (or, in our language, the setting of subconstant error), and pure differential privacy ($\delta = 0$).

2 Preliminaries

We begin with the introduction of some notation. Throughout the paper we use C, C_1, etc., for absolute constants, whose value may change from line to line. We use $\langle \cdot, \cdot \rangle$ for the standard inner product on $\mathbb{R}^m$, $\|\cdot\|_2$ for the standard Euclidean norm, and $\|\cdot\|_1$ for the ℓ_1 norm in $\mathbb{R}^m$. We define B_1^m and B_2^m to be the ℓ_1 and ℓ_2 unit balls in $\mathbb{R}^m$ respectively, while $Q^m = [-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}]^m \subseteq \mathbb{R}^m$ will refer to the m-dimensional hypercube, normalized to be contained in the unit Euclidean ball. We use I_m for the identity operator on $\mathbb{R}^m$, as well as for the $m \times m$ identity matrix. For a given subspace E, we define $\Pi_E \colon \mathbb{R}^m \to \mathbb{R}^m$ as the orthogonal projection operator onto E. Moreover, when $T \colon E \to F$ is a linear operator between the subspaces $E, F \subseteq \mathbb{R}^m$, we define $\|T\| = \max\{\|Tx\|_2 : \|x\|_2 = 1\}$ as its operator norm, which is also equal to its largest singular value $\sigma_1(T)$. For the diameter of a set K we use the nonstandard, but convenient, definition $\operatorname{diam} K = \max\{\|x\|_2 : x \in K\}$. For sets symmetric around 0, this is equivalent to the standard definition, but scaled up by a factor of 2. We use $N(\mu, \Sigma)$ to refer to the Gaussian distribution with mean μ and covariance Σ, and we use the notation $x \sim N(\mu, \Sigma)$ to denote that x is distributed as a Gaussian random variable with mean μ and covariance Σ. For an $m \times m$ symmetric matrix (or equivalently a self-adjoint operator from ℓ_2^m to ℓ_2^m) A we use $A \succeq 0$ to denote that A is positive semidefinite, i.e. $\langle x, Ax \rangle \geq 0$ for any $x \in \mathbb{R}^m$. For positive semidefinite matrices/operators A, B, we use the notation $A \preceq B$ to denote $B - A \succeq 0$.

2.1 Convex Geometry

In this section, we outline the main geometric tools we use in later sections. For a more detailed treatment, we refer to the lecture notes by Vershynin [23] and the books by Pisier [21] and Artstein-Avidan, Giannopoulos, and Milman [1].

Throughout, we define a *convex body* K as a compact subset of $\mathbb{R}^m$ with non-empty interior. A convex body K is *(centrally) symmetric* if and only if $K = -K$. We define the *polar body* K° of K as: $K^\circ = \{y : \langle x, y \rangle \leq 1 \ \forall x \in K\}$. The following basic facts are easy to verify and very useful.

▶ **Fact 5.** *For convex bodies $K, L \subseteq \mathbb{R}^m$, $K \subseteq L \Leftrightarrow L^\circ \subseteq K^\circ$.*

▶ **Fact 6** (Section/Projection Duality). *For a convex body $K \subseteq \mathbb{R}^m$ and a subspace $E \subseteq \mathbb{R}^m$:*

1. $(K \cap E)^\circ = \Pi_E(K^\circ)$;
2. $(\Pi_E(K))^\circ = K^\circ \cap E$.

In both cases, the polar is taken in the subspace E.

▶ **Fact 7.** *For any invertible linear map T and any convex body K, $T(K)^\circ = T^{-*}(K^\circ)$, where T^{-*} is the inverse of the adjoint operator T^*.*

A simple special case of Fact 7 is that, for any convex body K, $(rK)^\circ = \frac{1}{r}K^\circ$. Using this property alongside Fact 5 and Fact 6, we have the following useful corollary.

▶ **Corollary 8.** *For a convex body $K \subseteq \mathbb{R}^m$ and $E \subseteq \mathbb{R}^m$ a subspace with $k = \dim E$, the following two statements are equivalent:*

1. $\Pi_E(rB_2^m) \subseteq \Pi_E(K)$;
2. $K^\circ \cap E \subseteq \frac{1}{r}(B_2^m \cap E)$,

where, as before, taking the polar set is considered in the subspace E. Notice that the second statement is also equivalent to $\mathrm{diam}(K^\circ \cap E) \leq \frac{1}{r}$.

Our work relies on appropriately quantifying the "size" of (projections and sections of) a convex body. It turns out that, for our purposes, the right measure of size is related to the notion of *width*, captured by the *support function*. Recall from the introduction that the support function of a convex body $K \subset \mathbb{R}^m$ is given by $h_K(y) = \max_{x \in K} \langle x, y \rangle$ for every $y \in \mathbb{R}^m$.

The support function is intimately related to the *Minkowski norm* $\|\cdot\|_K$, defined for a symmetric convex body $K \subseteq \mathbb{R}^m$ by $\|x\|_K = \min\{r \geq 0 : x \in rK\}$, for every $x \in \mathbb{R}^m$. It is easy to verify that $\|\cdot\|_K$ is indeed a norm. The support function h_K is identical to the Minkowski norm of the polar body K° (which is also the dual norm to $\|\cdot\|_K$): $h_K(y) = \|y\|_{K^\circ}$ for every $y \in \mathbb{R}^m$.

Now we come to the measure of the "size" of a convex body which will be central to our results: the Gaussian mean width of the body, defined next.

▶ **Definition 9.** The Gaussian mean width and Gaussian mean norm of a symmetric convex body $K \subseteq \mathbb{R}^m$ are defined respectively as:

$$\ell^*(K) = \mathbb{E}\|g\|_{K^\circ} = \mathbb{E}[h_K(g)], \qquad \ell(K) = \mathbb{E}\|g\|_K,$$

where $g \sim N(0, I_m)$ is a standard Gaussian random variable.

The next lemma gives an estimate of how the mean width changes when applying a linear transformation to K. The lemma is standard and the proof is deferred to the full version of the paper.

▶ **Lemma 10.** *For any symmetric convex body $K \subset \mathbb{R}^m$, and any linear operator $T \colon \ell_2^m \to \ell_2^m$:*

$$\ell^*(T(K)) \leq \|T\|\ell^*(K).$$

Similar to approaches in previous works ([9], [16]), we exploit properties inherent to a specific position of K to prove lower bounds on its sample complexity.

▶ **Definition 11** (ℓ-position)**.** A convex body $K \subseteq \mathbb{R}^m$ is in ℓ-position if for all linear operators $T \colon \ell_2^m \to \ell_2^m$:

$$\ell^*(K) \cdot \ell(K) \leq \ell^*(T(K)) \cdot \ell(T(K)).$$

Clearly, K is in ℓ-position if and only if K° is in ℓ-position, since $\ell^*(K) = \ell(K^\circ)$ for any convex body K. Note further that the product $\ell^*(K) \cdot \ell(K)$ is scale-invariant, in the sense that $\ell^*(rK) \cdot \ell(rK) = \ell^*(K) \cdot \ell(K)$ for any nonnegative real r. This is because, for any $x, y \in \mathbb{R}^m$, $\|x\|_{rK} = \frac{1}{r}\|x\|_K$, and $h_{rK}(y) = rh_K(y)$, so $\ell^*(rK) = r\ell^*(K)$ and $\ell(rK) = \frac{1}{r}\ell(K)$.

We will relate the Gaussian mean width of K to another measure of its size, and the size of its projections and sections, known as Gelfand width. A definition follows.

▶ **Definition 12** (Gelfand width)**.** For a symmetric convex body $K \subset \mathbb{R}^m$, the *Gelfand width of order k* of K (with respect to the ℓ_2 norm) is defined as:

$$c_k(K) = \inf_E \inf\{r : K \cap E \subseteq r(B_2^m \cap E)\} = \inf_E \sup\{\|x\|_2 : x \in K \cap E\},$$

where the first infimum is over subspaces $E \subseteq \mathbb{R}^m$ of co-dimension at most $k - 1$ (i.e. of dimension at least $m - k + 1$). When $k > m$, we define $c_k(K) = 0$.

Note that $c_k(K) = \inf_E \operatorname{diam}(K \cap E)$, where the infimum is over subspaces $E \subseteq \mathbb{R}^m$ of codimension at most $k-1$. Observe also that for any K, $c_k(K)$ is non-increasing in k. It is well-known that the infimum in the definition is actually achieved [19].

2.2 Known Bounds

In this section, we recall some known differentially private mechanisms, with bounds on their sample complexity, as well as a lower bound on the optimal sample complexity. We start with the lower bound:

▶ **Theorem 13** ([3, 22]). *For all $\varepsilon = O(1)$, $2^{-\Omega(n)} \le \delta \le 1/n^{1+\Omega(1)}$ and $\alpha \le 1/10$:*

$$\mathrm{sc}_{\varepsilon,\delta}\left(Q^m, \alpha\right) = \Omega\left(\frac{\sqrt{m \log 1/\delta}}{\alpha\varepsilon}\right). \tag{1}$$

Next we recall one of the most basic mechanisms in differential privacy, the Gaussian mechanism. A proof of the privacy guarantee, with the constants given below, can be found in [16].

▶ **Theorem 14** (Gaussian Mechanism [4, 7, 5]). *Let $\mathcal{D} = \{x_1, \ldots, x_n\}$ be such that $\forall i : \|x_i\|_2 \le \sigma$. If $w \sim N(0, \sigma(\varepsilon,\delta)^2\sigma^2 I_m)$, $\sigma(\varepsilon,\delta) = (\sqrt{\varepsilon} + 2\sqrt{2\log(1/\delta)})/\varepsilon$ and $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix, then the algorithm $\mathcal{A}_{GM}$ defined by $\mathcal{A}_{GM}(\mathcal{D}) = \overline{\mathcal{D}} + \frac{1}{n}w$ is (ε,δ)-differentially private.*

▶ **Corollary 15.** *For any symmetric convex $K \subseteq B_2^m$:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = O\left(\frac{\sqrt{m \log 1/\delta}}{\alpha\varepsilon}\right).$$

In the rest of the paper we will use the notation $\sigma(\varepsilon,\delta) = \frac{\sqrt{\varepsilon}+2\sqrt{2\log(1/\delta)}}{\varepsilon}$ from the theorem statement above.

Finally, we also present the projection mechanism from [16], which post-processes the output of the Gaussian mechanism by projecting onto K.

▶ **Theorem 16** (Projection Mechanism [16, 6]). *Let $K \subseteq B_2^m$ be a symmetric convex body, and define $\mathcal{A}_{PM}$ to be the algorithm that, on input $\mathcal{D} = \{x_1, \ldots, x_n\} \subset K$, outputs:*

$$\hat{y} = \arg\min\{\|\hat{y} - \tilde{y}\|_2^2 : \hat{y} \in K\},$$

where $\tilde{y} = \overline{\mathcal{D}} + \frac{1}{n}w$, $w \sim N(0, \sigma(\varepsilon,\delta)^2 I_m)$. Then $\mathcal{A}_{PM}$ satisfies (ε,δ)-differential privacy and has sample complexity:

$$\mathrm{sc}(K, \mathcal{A}_{PM}, \alpha) = O\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\alpha^2}\right).$$

▶ **Corollary 17.** *For any symmetric convex $K \subseteq B_2^m$:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = O\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\alpha^2}\right).$$

3 Basic Properties of Sample Complexity

In this section, we prove some fundamental properties of sample complexity that will be extensively used in later sections. The proofs Lemmas 18,20 and Theorem 23 are deferred to the full version of the paper.

▶ **Lemma 18.** *$L \subseteq K \Rightarrow \forall \alpha \in (0,1) : \mathrm{sc}_{\varepsilon,\delta}(L,\alpha) \leq \mathrm{sc}_{\varepsilon,\delta}(K,\alpha)$.*

▶ **Corollary 19.** *For all $\varepsilon = O(1)$, $2^{-\Omega(n)} \leq \delta \leq 1/n^{1+\Omega(1)}$ and $\alpha \leq 1/10$:*

$$\mathrm{sc}_{\varepsilon,\delta}(B_2^m,\alpha) = \Omega\left(\frac{\sqrt{m\log 1/\delta}}{\alpha\varepsilon}\right).$$

Proof. Since $Q^m \subseteq B_2^m$, this follows directly from Lemma 18 and Theorem 13. ◀

▶ **Lemma 20.** *For any $\alpha \in (0,1)$, any linear operator $T\colon \mathbb{R}^m \to \mathbb{R}^m$ and any symmetric convex body $K \subset \mathbb{R}^m$:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) \geq \mathrm{sc}_{\varepsilon,\delta}(T(K),\alpha \cdot \|T\|).$$

▶ **Corollary 21.** *For any $t > 0$:*

$$\mathrm{sc}_{\varepsilon,\delta}(tK,t\alpha) = \mathrm{sc}_{\varepsilon,\delta}(K,\alpha).$$

Proof. Taking $T = tI_m$ in Lemma 20, where I_m is the identity on $\mathbb{R}^m$, the lemma implies $\mathrm{sc}_{\varepsilon,\delta}(tK,t\alpha) \leq \mathrm{sc}_{\varepsilon,\delta}(K,\alpha)$. Since this inequality holds for any t and K, we may apply it to $K' = tK$ and $t' = 1/t$, and we get $\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \mathrm{sc}_{\varepsilon,\delta}((1/t)tK,(1/t)t\alpha) \leq \mathrm{sc}_{\varepsilon,\delta}(tK,t\alpha)$. ◀

Since for any subspace E of $\mathbb{R}^m$, the corresponding orthogonal projection Π_E has operator norm 1, we also immediately get the following corollary of Lemma 20:

▶ **Corollary 22.** *For any subspace E:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) \geq \mathrm{sc}_{\varepsilon,\delta}(\Pi_E(K),\alpha).$$

In the next theorem, we combine the lower bound in Theorem 19 and the properties we proved above in order to give a lower bound on the sample complexity of an arbitrary symmetric convex body K in terms of its geometric properties. In the following sections we will relate this geometric lower bound to the mean Gaussian width of K.

▶ **Theorem 23** (Geometric Lower Bound). *For all $\varepsilon = O(1)$, $2^{-\Omega(n)} \leq \delta \leq 1/n^{1+\Omega(1)}$, any convex symmetric body $K \subseteq \mathbb{R}^m$, any $1 \leq k \leq m$ and any $\alpha \leq 1/(10c_k(K^\circ))$:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \Omega\left(\frac{\sqrt{\log 1/\delta}}{\alpha\varepsilon} \cdot \frac{\sqrt{m-k+1}}{c_k(K^\circ)}\right).$$

4 Optimality of the Gaussian Mechanism

In this section, we present the result that the Gaussian mechanism is optimal, up to constant factors, when $K \subseteq B_2^m$ is sufficiently large. More specifically, if the Gaussian mean width of K is asymptotically maximal, then we can get a tight lower bound on the sample complexity of the Gaussian mechanism. This is summarized in the theorem below.

▶ **Theorem 24.** *For all $\varepsilon < O(1)$, $2^{-\Omega(n)} \leq \delta \leq 1/n^{1+\Omega(1)}$, sufficiently small constant α, and any symmetric convex body $K \subseteq B_2^m$, if*

$$\ell^*(K) = \Omega(\sqrt{m}),$$

then:

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \Theta\left(\frac{\sqrt{m \log 1/\delta}}{\alpha\varepsilon}\right),$$

and $\mathrm{sc}_{\varepsilon,\delta}(K,\alpha)$ is achieved, up to constants, by the Gaussian mechanism.

By Theorem 15 we have an upper bound for the Gaussian mechanism defined previously. To prove its optimality, we use a classical result from convex geometry, known as Dvoretzky's criterion, to show a matching lower bound for the sample complexity. This result relates the existence of a nearly-spherical section of a given convex body to the Gaussian mean norm. It was a key ingredient in Milman's probabilistic proof of Dvoretzky's theorem: see Matoušek's book [12] for an exposition.

▶ **Theorem 25** ([13]; Dvoretzky's Criterion). *For every symmetric convex body $K \subseteq \mathbb{R}^m$ such that $B_2^m \subseteq K$, and every $\beta < 1$, there exists a constant $c(\beta)$ and a subspace E with dimension $\dim E \geq c(\beta)\ell(K)^2$ for which:*

$$(1-\beta)\frac{\ell(K)}{\sqrt{m}} B_2^m \cap E \subseteq K \cap E \subseteq (1+\beta)\frac{\ell(K)}{\sqrt{m}} B_2^m \cap E.$$

Proof of Theorem 24. Given the matching upper bound on sample complexity in Theorem 15, it suffices to show the equivalent lower bound, namely that:

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \Omega\left(\frac{\sqrt{m \log 1/\delta}}{\alpha\varepsilon}\right).$$

To this end, we will show that there exists a $k \leq (1-c)m + 1$, for an absolute constant c, so that $c_k(K^\circ) = O(1)$. Then the lower bound will follow directly from Theorem 23.

We will prove the claim above by applying Dvoretzky's criterion to K°. By Theorem 5, $K \subseteq B_2^m \Rightarrow B_2^m \subseteq K^\circ$. We can then apply Dvoretzky's criterion with $\beta = 1/2$, ensuring that there exists a subspace E of dimension $\dim E \geq c(1/2)\ell(K^\circ)^2$ for which:

$$K^\circ \cap E \subseteq \frac{\ell(K^\circ)}{2\sqrt{m}} B_2^m \cap E.$$

Let us define $k = m - \dim E + 1$; then $k \leq m - c(1/2)\ell(K^\circ)^2 + 1 = m - c(1/2)\ell^*(K)^2 + 1$. Since, by assumption $\ell^*(K) = \Omega(m)$, there exists a constant c so that $k \leq (1-c)m + 1$. Finally, by the definition of Gelfand width, $c_k(K^\circ) \leq \frac{\ell(K^\circ)}{2\sqrt{m}} = O(1)$, as desired. This completes the proof. ◀

5 Gaussian Width Lower Bounds in ℓ-position

In Section 4 we showed that the Gaussian Mechanism is optimal when the Gaussian mean width of K is asymptotically as large possible. Our goal in this and the following section is to show general lower bounds on sample complexity in terms of $\ell^*(K)$. This is motivated by the sample complexity upper bound in terms of $\ell^*(K)$ provided by the projection mechanism.

It is natural to follow the strategy from Section 4: use Dvoretzky's criterion to find a nearly-spherical projection of K of appropriate radius and dimension. An inspection of the proof of Theorem 24 shows that the sample complexity lower bound we get this way is $\Omega\left(\frac{\ell^*(K)^2}{\sqrt{m}}\right)$ (ignoring the dependence on ε, δ, and α here, and in the rest of this informal discussion). Recall that we are aiming for a lower bound of $\Omega(\ell^*(K))$, so we are off by a factor of $\frac{\ell^*(K)}{\sqrt{m}}$. Roughly speaking, the problem is that Dvoretzky's criterion does too much: it guarantees a spherical section of K°, while we only need a bound on the diameter of the section. In order to circumvent this difficulty, we use a different result from asymptotic convex geometry, the low M^*-estimate, which bounds the diameter of a random section of K°, without also producing a large ball contained inside the section. A technical difficulty is that the resulting upper bound on the diameter is in terms of the Gaussian mean K-norm, rather than the (reciprocal of the) mean width. When K is in ℓ-position, this is not an issue, because results of Pisier, Figiel, and Tomczak-Jaegermann show that in that case $\ell(K)\ell^*(K) = O(\log m)$. In this section we assume that K is in ℓ-position, and we remove this requirement in the subsequent section.

The main result of this section is summarized below.

▶ **Theorem 26.** *For all $\varepsilon = O(1)$, $2^{-\Omega(n)} \leq \delta \leq 1/n^{1+\Omega(1)}$, all symmetric convex bodies $K \subseteq \mathbb{R}^m$ in ℓ-position, and for $\alpha \leq \frac{\ell^*(K)}{C\sqrt{m}\log 2m}$, where C is an absolute constant:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \Omega\left(\frac{\sqrt{\log 1/\delta}}{\alpha\varepsilon} \cdot \frac{\ell^*(K)}{\log 2m}\right).$$

The following two theorems are the main technical ingredients we need in the proof of Theorem 26.

▶ **Theorem 27** ([8], [20]; MM^* **Bound).** *There exists a constant C such that for every symmetric convex body $K \subset \mathbb{R}^m$ in ℓ-position:*

$$\ell(K) \cdot \ell^*(K) \leq C \cdot m\log 2m.$$

It is an open problem whether this bound can be improved to $m\sqrt{\log 2m}$. This would be tight for the cube Q^m. This improvement would lead to a corresponding improvement in our bounds.

▶ **Theorem 28** ([17]; Low M^* **estimate).** *' There exists a constant C such that for every symmetric convex body $K \subset \mathbb{R}^m$ there exists a subspace $E \subseteq \mathbb{R}^m$ with $\dim E = m - k$ for which:*

$$\mathrm{diam}\,(K \cap E) \leq C \cdot \frac{\ell^*(K)}{\sqrt{k}}.$$

Combining Theorems 27 and 28, we get the following key lemma.

▶ **Lemma 29.** *There exists a constant C such that for every symmetric convex body $K \subset \mathbb{R}^m$ in ℓ-position, and every $\beta \in (0, 1 - 1/m)$, there exists a subspace E of dimension at least βm satisfying:*

$$\mathrm{diam}(K \cap E) \leq C\frac{\sqrt{m}\log 2m}{\sqrt{1-\beta} \cdot \ell(K)}.$$

Proof. Let $k = \lfloor (1-\beta)m \rfloor \geq 1$. Using the low M^* estimate on K, there exists a subspace E with $\dim E = m - k = \lceil \beta m \rceil$ for which:

$$\operatorname{diam}(K \cap E) \leq C_1 \cdot \frac{\ell^*(K)}{\sqrt{k}}.$$

By the MM^* upper bound, since K is in ℓ-position, we have that:

$$\ell^*(K) \leq C_2 \cdot \frac{m \log 2m}{\ell(K)},$$

and, combining the two inequalities, we get that:

$$\operatorname{diam}(K \cap E) \leq C_1 C_2 \frac{m \log 2m}{\sqrt{k} \cdot \ell(K)} \leq C \frac{\sqrt{m} \log m}{\sqrt{1-\beta} \cdot \ell(K)},$$

for an appropriate constant C. This completes the proof. ◀

The proof of the desired lower bound now follows easily from this lemma.

Proof of Theorem 26. By Theorem 23, it suffices to show that

$$\max_{k=1}^{m} \frac{\sqrt{m-k+1}}{c_k(K^\circ)} = \Omega\left(\frac{\ell^*(K)}{\log 2m}\right). \tag{2}$$

Indeed, if k^* is the value of k for which the maximum on the left hand side is achieved, then $\frac{\sqrt{m-k^*+1}}{c_{k^*}(K^\circ)}$ is a lower bound on the sample complexity for all $\alpha \leq 1/(10 c_{k^*}(K^\circ))$, and by (2):

$$\frac{1}{10 c_{k^*}(K^\circ)} = \Omega\left(\frac{\ell^*(K)}{\sqrt{m-k^*+1} \cdot \log 2m}\right) = \Omega\left(\frac{\ell^*(K)}{\sqrt{m} \cdot \log 2m}\right).$$

In the rest of the proof, we establish (2).

Since K (and thus also K°) are in ℓ-position by assumption, from Lemma 29 applied to K° we have that there exists a subspace E such that $\dim E \geq m/2$ and:

$$\operatorname{diam}(K^\circ \cap E) = O\left(\frac{\sqrt{m} \log 2m}{\ell(K^\circ)}\right) = O\left(\frac{\sqrt{m} \log 2m}{\ell^*(K)}\right).$$

Setting $k_E = m - \dim E + 1 \leq m/2 + 1$, and because $c_{k_E}(K^\circ) \leq \operatorname{diam}(K^\circ \cap E)$ by definition, we get that:

$$\max_{k=1}^{m} \frac{\sqrt{m-k+1}}{c_k(K^\circ)} \geq \frac{\sqrt{\dim E}}{\operatorname{diam}(K^\circ \cap E)} = \Omega\left(\frac{\ell^*(K)}{\log 2m}\right),$$

as desired. ◀

6 Gaussian Width Lower Bounds for Arbitrary Bodies

In this section, we remove the assumption that K is in ℓ-position from the previous section. Instead, we use a recursive charging argument in order to reduce to the ℓ-position case. The resulting guarantee is worse than the one we proved for bodies in ℓ-position by a logarithmic factor.

The main lower bound result of this section is the following theorem.

▶ **Theorem 30.** *For all $\varepsilon = O(1)$, $2^{-\Omega(n)} \le \delta \le 1/n^{1+\Omega(1)}$, any symmetric convex body $K \subset \mathbb{R}^m$, and any $\alpha \le \frac{\ell^*(K)}{C\sqrt{m}(\log 2m)^2}$, where C is an absolute constant:*

$$\mathrm{sc}_{\varepsilon,\delta}(K,\alpha) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{(\log 2m)^2\alpha}\right).$$

The lower bound follows from the geometric lemma below, which may be of independent interest.

▶ **Lemma 31.** *There exists a constant C such that, for any symmetric convex body $K \subset \mathbb{R}^m$,*

$$\ell^*(K) \le C(\log 2m)\left(\sum_{i=1}^m \frac{1}{\sqrt{i}\cdot c_{m-i+1}(K^\circ)}\right). \tag{3}$$

Lemma 31 is closely related to the volume number theorem of Milman and Pisier [14], which states that the inequality (3) holds with $\frac{1}{c_{m-i+1}(K^\circ)}$ replaced by the volume number $v_i(K)$, defined as:

$$v_i(K) = \sup_{E:\dim E = i} \frac{\mathrm{vol}(\Pi_E(K))^{1/i}}{(\mathrm{vol}(\Pi_E(B_2^m))^{1/i}},$$

where the supremum is over subspaces E of $\mathbb{R}^m$. Inequality (3) is stronger than the volume number theorem, because $\frac{1}{c_{m-i+1}(K^\circ)} \le v_i(K)$. Indeed, setting $r = \frac{1}{c_{m-i+1}(K^\circ)}$, by Theorem 8 and the definition of Gelfand width we have that there exists a subspace E of dimension i such that $r\Pi_E(B_2^m) \subseteq \Pi_E K$. Therefore, $\mathrm{vol}(\Pi_E(K)) \ge r^i \,\mathrm{vol}(\Pi_E(B_2^m))$, which implies the desired inequality.

Even though the volume number theorem is weaker than (3), the proof given by Pisier in his book [21], with minor modifications, appears to yield the stronger inequality we need. In the full version of the paper we give an alternative proof, which only uses the low M^* estimate, the MM^* estimate, and elementary linear algebra.

Proof of Theorem 30. As in the proof of Theorem 26, it is sufficient to prove that:

$$\max_{k=1}^m \frac{\sqrt{m-k+1}}{c_k(K^\circ)} = \Omega\left(\frac{\ell^*(K)}{(\log 2m)^2}\right). \tag{4}$$

But this inequality follows easily from Lemma 31 and the trivial case of Hölder's inequality:

$$\begin{aligned}\ell^*(K) \le C(\log 2m)\left(\sum_{i=1}^m \frac{1}{\sqrt{i}\cdot c_{m-i+1}(K^\circ)}\right) &\le C(\log 2m)\left(\sum_{i=1}^m \frac{1}{i}\right)\cdot\left(\max_{i=1}^m \frac{\sqrt{i}}{c_{m-i+1}(K^\circ)}\right)\\ &= O((\log 2m)^2)\cdot\left(\max_{i=1}^m \frac{\sqrt{i}}{c_{m-i+1}(K^\circ)}\right).\end{aligned}$$

Then, the proof of the theorem follows from (4) analogously to the proof of Theorem 26. ◀

We now have everything in place to prove our main result for the mean point problem.

Proof of Theorem 4. The upper bounds on sample complexity follow from Theorem 14, Theorem 15, Theorem 16, and Theorem 17. The lower bounds follow from Theorem 30. The statement after "moreover" follows from Theorem 15 and Theorem 24. ◀

7 From Mean Point to Query Release

All the bounds we proved so far were for the mean point problem. In this section we show reductions between this problem, and the query release problem, which allow us to translate our lower bounds to the query release setting and prove Theorem 3. We will show that the problem of approximating $\mathcal{Q}(\mathcal{D})$ for a query workload $\mathcal{Q}$ under differential privacy is nearly equivalent to approximating the mean point problem with universe $K' = \frac{1}{\sqrt{m}}K$, where K is the sensitivity polytope of $\mathcal{Q}$. We state this reduction next, and defer its proof to the full version of the paper.

▶ **Lemma 32.** *Let $\mathcal{Q}$ be a workload of m linear queries over the universe $\mathcal{U}$ with sensitivity polytope K. Define $K' = \frac{1}{\sqrt{m}}K$. Then, we have the inequalities:*

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha) \le \mathrm{sc}_{\varepsilon,\delta}(K',\alpha); \tag{5}$$

$$\mathrm{sc}_{\varepsilon,\delta}(K',\alpha) \le \max\left\{\mathrm{sc}_{2\varepsilon,2\delta}(\mathcal{Q},\alpha/4), \frac{16\,\mathrm{diam}(K')}{\alpha^2}\right\}. \tag{6}$$

Moreover, we can use an (ε,δ)-differentially private algorithm $\mathcal{A}'$ as a black box to get an (ε,δ)-differentially private algorithm $\mathcal{A}$ such that $\mathrm{sc}(\mathcal{Q},\mathcal{A},\alpha) = \mathrm{sc}(K',\mathcal{A}',\alpha)$. $\mathcal{A}$ makes a single call to $\mathcal{A}'$, and performs additional computation of worst-case complexity $O(mn)$, where n is the size of the database.

We also use a simple lemma that relates the sample complexity at an error level α to the sample complexity at a lower error level $\alpha' < \alpha$. The proof is a padding argument and can be found in [22].

▶ **Lemma 33.** *For any workload $\mathcal{Q}$, any $0 < \alpha' < \alpha < 1$, and any privacy parameters ε, δ, we have*

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha') = \Omega\left(\frac{\alpha}{C\alpha'}\right)\cdot \mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha),$$

for an absolute constant C.

We are now ready to finish the proof of our main result for the query release problem.

Proof of Theorem 3. The upper bounds on sample complexity follow from the upper bounds in Theorem 4 together with Lemma 32.

Denote $K' = \frac{1}{\sqrt{m}}K$, and let $\alpha_0 = \frac{\ell^*(K')}{C\sqrt{m}(\log 2m)^2} = \frac{\ell^*(K)}{Cm(\log 2m)^2}$ be the smallest error parameter for which Theorem 30 holds. Then, by Theorem 30 and Lemma 32:

$$\max\left\{\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha_0), \frac{\mathrm{diam}(K)}{\sqrt{m}\alpha_0^2}\right\} = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\sqrt{m}(\log 2m)^2\alpha_0}\right).$$

It is easy to show that $\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha_0) = \Omega(\mathrm{diam}(K)/(\alpha_0\sqrt{m}))$ for all sufficiently small ε and δ. Therefore, we have:

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha_0) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)}{\sqrt{m}(\log 2m)^2}\right).$$

By Lemma 33, we get that for any $\alpha \le \alpha_0$ the sample complexity is at least:

$$\mathrm{sc}_{\varepsilon,\delta}(\mathcal{Q},\alpha) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)\alpha_0}{\sqrt{m}(\log 2m)^2\alpha}\right) = \Omega\left(\frac{\sigma(\varepsilon,\delta)\ell^*(K)^2}{m^{3/2}(\log 2m)^4\alpha}\right).$$

An analogous proof, with $\alpha_0 = 1/C$ set to the smallest error parameter for which Theorem 24 holds, establishes the statement after "moreover". ◀

References

1 Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D. Milman. *Asymptotic geometric analysis. Part I*, volume 202 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015. doi:10.1090/surv/202.

2 Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the 44th Symposium on Theory of Computing*, STOC'12, pages 1269–1284, New York, NY, USA, 2012. ACM. doi:10.1145/2213977.2214089.

3 Mark Bun, Jonathan Ullman, and Salil Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 1–10. ACM, 2014.

4 Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 202–210. ACM, 2003.

5 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer, 2006.

6 Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Using convex relaxations for efficiently and privately releasing marginals. In *30th Annual Symposium on Computational Geometry, SOCG'14, Kyoto, Japan, June 08-11, 2014*, page 261. ACM, 2014. doi:10.1145/2582112.2582123.

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

8 T. Figiel and Nicole Tomczak-Jaegermann. Projections onto Hilbertian subspaces of Banach spaces. *Israel J. Math.*, 33(2):155–171, 1979. doi:10.1007/BF02760556.

9 Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC'10, pages 705–714, New York, NY, USA, 2010. ACM. doi:10.1145/1806689.1806786.

10 Michael Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998. doi:10.1145/293347.293351.

11 Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*. Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

12 Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. doi:10.1007/978-1-4613-0039-7.

13 V. D. Milman. A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies. *Funkcional. Anal. i Priložen.*, 5(4):28–37, 1971.

14 V. D. Milman and G. Pisier. Gaussian processes and mixed volumes. *Ann. Probab.*, 15(1):292–304, 1987. URL: http://links.jstor.org/sici?sici=0091-1798(198701)15:1<292:GPAMV>2.0.CO;2-A&origin=MSN.

15 Aleksandar Nikolov. An improved private mechanism for small databases. In *Automata, Languages, and Programming – 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, volume 9134 of *Lecture Notes in Computer Science*, pages 1010–1021. Springer, 2015. doi:10.1007/978-3-662-47672-7_82.

16 Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 351–360. ACM, 2013. doi:10.1145/2488608.2488652.

17 Alain Pajor and Nicole Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional Banach spaces. *Proc. Amer. Math. Soc.*, 97(4):637–642, 1986. doi:10.2307/2045920.

18 Christos H. Papadimitriou and Mihalis Yannakakis. On limited nondeterminism and the complexity of the V-C dimension. *J. Comput. Syst. Sci.*, 53(2):161–170, 1996. doi:10.1006/jcss.1996.0058.

19 Allan Pinkus. *n-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985. doi:10.1007/978-3-642-69894-1.

20 G. Pisier. Sur les espaces de Banach K-convexes. In *Seminar on Functional Analysis, 1979–1980 (French)*, pages Exp. No. 11, 15. École Polytech., Palaiseau, 1980.

21 Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1989. doi:10.1017/CBO9780511662454.

22 Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *CoRR*, abs/1501.06095, 2015. URL: http://arxiv.org/abs/1501.06095.

23 R. Vershynin. Lectures in geometric functional analysis. 2009. URL: http://www-personal.umich.edu/~romanv/papers/GFA-book.pdf.