LA-UR-09-04593

Approved for public release; distribution is unlimited.

Title:	The Price of Privately Releasing Contingency Tables, and the Spectra of Random Matrices with Correlated Rows
Author(s):	Shiva Kasiviswanathan, 209013, CCS-3 Mark Rudelson, University of Missouri Adam Smith, Pennsylvania State University
Intended for:	Symposium on Theory of Computing 2010 Boston, MA June 6 - 8, 2010

Los Alamos
NATIONAL LABORATORY
EST.1943

Form 836 (7/06)

The Price of Privately Releasing Contingency Tables, and the Spectra of Random Matrices with Correlated Rows

Shiva Kasiviswanathan* Mark Rudelson† Adam Smith‡

Abstract

Contingency tables are the method of choice of government agencies for releasing statistical summaries of categorical data. In this paper, we consider lower bounds on how much distortion (noise) is necessary in these tables to provide privacy guarantees when the data being summarized is sensitive. We extend a line of recent work on lower bounds on noise for private data analysis [10, 13, 14, 15] to a natural and important class of functionalities. Our investigation also leads to new results on the spectra of random matrices with correlated rows.

Consider a database D consisting of n rows (one per individual), each row comprising d binary attributes. For any subset of T attributes of size $|T| = k$, the marginal table for T has 2^k entries; each entry counts how many times in the database a particular setting of these attributes occurs. Imagine an agency that wishes to release all $\binom{d}{k}$ contingency tables for a given database.

For constant k, previous work showed that distortion $\tilde{O}(\min\{n, (n^2 d)^{1/3}, \sqrt{d^k}\})$ is *sufficient* for satisfying differential privacy, a rigorous definition of privacy that has received extensive recent study. Our main contributions are:

- For ϵ- and (ϵ, δ)-differential privacy (with ϵ constant and $\delta = 1/poly(n)$), we give a lower bound of $\tilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$, which is tight for $n = \tilde{\Omega}(d^k)$. Moreover, for a natural and popular class of mechanisms based on additive noise, our bound can be strengthened to $\Omega(\sqrt{d^k})$, which is tight for all n. Our bounds extend even to non-constant k, losing roughly a factor of $\sqrt{2^k}$ compared to the best known upper bounds for large n.
- We give efficient polynomial time attacks which allow an adversary to reconstruct sensitive information given insufficiently perturbed contingency table releases. For constant k, we obtain a lower bound of $\tilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$ that applies to a large class of privacy notions, including K-anonymity (along with its variants) and differential privacy. In contrast to our bounds for differential privacy, this bound (a) is shown only for constant k, but (b) is tight for all values of n when k is constant.
- Our reconstruction-based attacks require a new lower bound on the least singular values of random matrices with correlated rows. For a constant k, consider a matrix M with $\binom{d}{k}$ rows which are formed by taking all possible k-way entry-wise products of an underlying set of d random vectors. We show that even for nearly square matrices with $d^k / \log d$ columns, the least singular value is $\Omega(\sqrt{d^k})$ with high probability— asymptotically, the same bound as one gets for a matrix with *independent* rows. The proof requires several new ideas for analyzing random matrices and could be of independent interest.

*CCS-3, Los Alamos National Laboratories, kasivisw@gmail.com.

†Department of Mathematics, University of Missouri, rudelson@math.missouri.edu.

‡Department of Computer Science and Engineering, Pennsylvania State University, asmith@cse.psu.edu.

1 Introduction

The goal of *private data analysis* is to provide global, statistical properties of a data set of sensitive information while protecting the privacy of the individuals whose records the data set contains. There is a vast body of work on this problem in statistics and computer science. However, until recently, most schemes proposed in the literature lacked rigor: typically, the schemes had either no formal privacy guarantees or ensured security only against a specific suite of attacks.

The seminal results of Dinur and Nissim [10] and Dinur, Dwork and Nissim [9] initiated a rigorous study of the tradeoff between privacy and utility. The notion of *differential privacy* [13] that emerged from this line of work provides rigorous guarantees even in the presence of a malicious adversary with access to arbitrary side information. Differential privacy requires, roughly, that any single individual's data have little effect on the outcome of the analysis. Recently, many techniques have been developed for designing differentially private algorithms [4, 13, 31, 29, 2, 28, 20, 5, 17, 19, 39, 8, 12, 41, 18]. A typical objective is to release as accurate an approximation as possible to some function f evaluated on the database D.

A complementary line of work seeks to establish lower bounds on how much distortion is necessary for particular functions f. Some of these bounds apply only to differential privacy (e.g. [13, 18]); other bounds rule out *any* reasonable notion of privacy by giving algorithms to reconstruct almost all of the data D given sufficiently accurate approximations to $f(D)$ [10, 14, 15]. We refer to the latter works as lower bounds for *minimal* privacy.

In this paper, we investigate lower bounds on the distortion necessary for releasing a set of *contingency tables*, or marginal tables, under both differential and minimal privacy. A database D in our setting consists of n rows, each row comprising values for d binary attributes $x_1, \ldots, x_d$. For any subset of T attributes of size $|T| = k$, the marginal table for T has 2^k entries; each entry counts how many times in the database a particular setting of these attributes occurs. Alternatively, we may think of the table as counting the number of rows in the database that satisfy each of the 2^k possible *conjunctions* on the attributes in T. Contingency tables are important summary statistics for categorical data: in addition to being easy to interpret, they are sufficient statistics for popular classes of probabilistic models [3]. Because of this, they are a format of choice for data release by government statistical bureaus [2].

Barak *et al.* [2] investigated upper bounds on the noise needed to release contingency tables differentially privately. One can also derive incomparable upper bounds from the techniques of Blum *et al.* [4, 5]. These bounds are described in Tables 1 and 3. For the remainder of the introduction we identify the two notions, and treat ϵ and δ as constants.

1.1 Our Contributions

Let $\mathcal{C}_k(D)$ be the set of all k-way contingency tables (equivalently, the frequencies of all possible k-attribute conjunctions) for a database $D \in (\{0,1\}^d)^n$. One can think of $\mathcal{C}_k(D)$ as a single real vector of length $2^k\binom{d}{k}$.

(1) Lower bounds for Differential Privacy: We show that algorithms which do not sufficiently distort the contingency tables of D cannot be differentially private. Specifically, we give lower bounds on the (square root of the) average *mean squared error* (MSE) per entry of differentially private estimates of $\mathcal{C}_k$. The upper and lower bounds are stated in Table 3, and discussed in Section 2. Table 1 restates the bounds of Table 3 for the special case where k, ϵ are constants and $\delta = 1/poly(n)$.

For constant k, the best known algorithms yield distortion $\tilde{O}(\min\{n, (n^2 d)^{1/3}, \sqrt{d^k}\})$, while our lower bound is $\widetilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$. Our bounds imply that adding carefully calibrated Gaussian noise to each entry in $\mathcal{C}_k$ (as proposed in [4, 31]) is optimal for large databases (when $n = \tilde{\Omega}(d^k)$). Moreover, for a natural and popular class of algorithms based on adding *instance-independent* noise, our bound can be

Mechanism	U. B. (ϵ, δ)-diff privacy	L. B. (ϵ, δ)-diff privacy (This Paper)
Instance-Independent	$\tilde{O}\left(\sqrt{d^k}\right)$ [4]	$\Omega\left(\sqrt{d^k}\right)$
General	$\tilde{O}\left(\min\left\{n, \left(n^2 d\right)^{\frac{1}{3}}, \sqrt{d^k}\right\}\right)$ [4, 5]	$\tilde{\Omega}\left(\min\left\{\sqrt{n}, \sqrt{d^k}\right\}\right)$

Table 1: *Upper and lower bounds on the root average mean squared error per cell entry for releasing all k-way contingency tables with $k = const$, $\epsilon = const$, and $\delta = 1/poly(n)$. The diagonal entries of the mean squared error matrix are the mean squared error of the estimates, and taking the square root of the average of the diagonal entries gives the root average mean squared error. The n term in the upper bound for the general case comes from an algorithm that releases a vector of $n/2$'s for all D's. The $\tilde{O}(\cdot)$ and $\tilde{\Omega}(\cdot)$ notation hides polylogarithmic factors in the parameters of the problem. The full version of this table (Table 3) is in Section 2.*

strengthened to $\Omega(\sqrt{d^k})$, which is tight for all n. Our bounds extend even to non-constant k, losing a factor of $\sqrt{2^k}$ compared to the best known upper bounds (again, for large n in the general case).

The rough idea behind these lower bounds is to bound the projection of the mean squared error matrix of some database D along a large set of orthogonal directions. Combined with concentration inequalities for matrix-valued random variables, this allows us to bound the trace of the MSE matrix and hence the average MSE. This line of argument is quite different from the indistinguishability arguments used to bound the accuracy of parity queries in [13].

(2) Lower Bounds for Minimal Privacy: Using a disjoint set of techniques, we also show (slightly weaker) lower bounds that apply to a large class of "privacy" definitions for statistical databases, including differential privacy. Roughly, we construct distributions on databases D for which releasing too good an approximation to $\mathcal{C}_k(D)$ allows an adversary to efficiently recover almost all of D, even though the adversary's *a priori* chance of guessing any row of D is small, and the rows of D are statistically independent. The bounds are stated in Table 2 and discussed in Section 3.

We give two types of reconstruction results, corresponding to two violations of "minimal" privacy: we call schemes that allow these violations *strongly non-private* and *attribute non-private*, respectively. As a point of comparison, for constant k, our bound for strong non-privacy allows the same conclusion as do the bounds on differential privacy, namely average distortion $\tilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$ per entry is necessary. In contrast to our bounds for differential privacy, however, this bound (a) is shown only for constant k, but (b) is tight for all values of n when k is constant (that is, there is a non-differentially private algorithm, based on sampling, with distortion $\tilde{O}(\sqrt{n})$).

(3) The Least Singular Value of Random Matrices with Correlated Rows: For $k > 1$, the bounds on minimal privacy **(2)** above require significantly different techniques from previous work. Previous lower bounds [10, 14, 15] were based on variants of the following reconstruction problem: given a real-valued matrix M, and a corrupted "codeword" $Ms + e$, the goal is to compute an approximation $\hat{s}$ to s such that the "reconstruction error" $\hat{s} - s$ is somehow bounded in terms of the noise vector e. Typically, assuming some norm $\|e\|_p$ is small, one can bound a related norm of $\hat{s} - s$.

The connection to data privacy is that, if $s \in \mathbb{R}^n$ is a database with one number assigned per person, we can think of $y = Ms + e$ as a vector of (distorted) estimates of the quantities $\langle M_i, s\rangle$, where M_i is the ith row of M. Thus, any private data release that allows a user to estimate $\langle M_i, s\rangle$, allows an attacker to obtain y. Therefore, an algorithm for approximating s from y can be used to infer sensitive data from the release.

Previous lower bounds rely heavily on the freedom to design M by selecting the rows of M independently (either at random [10, 14, 15] or from an algebraic code [15]). They are closely related to techniques used to analyze the performance of random matrices in compressed sensing schemes and LP decoding (see [7, 6], and references therein).

When $k = 1$ a similar flexibility is available in our lower bounds. However, for $k > 1$ the rows of the matrices M that arise in our lower bounds are highly correlated: the matrix M has $\binom{d}{k}$ rows which are formed by taking all possible k-way *entry-wise products*[1] of an underlying set of d random vectors. The techniques of previous work, from the literature on both privacy and random matrices, break down.

We show that reconstruction procedures using these matrices can in fact be analyzed, by showing for any constant k that a *random* rectangular 0-1 correlated matrix has approximately same the high-probability bound on its least singular value as would a random 0-1 matrix with *independent* rows. Tight bounds are known on the least singular value bound for various types of matrices (e.g., square, rectangular) with independent random entries (see, e.g., [33, 34, 32, 37] and references therein). However, to deal with the dependencies, we develop several new tools, which may be of independent interest. We show that if M has less than $d^k/\log(d)$ columns, then it's least singular value $\Omega(\sqrt{d^k})$ with probability exponentially large in d. For comparison, a uniformly random $N \times n$ matrix with 0-1 entries has least singular value at least $\sqrt{N} - \sqrt{n-1}$ with exponentially high probability (Rudelson and Vershynin [34]). The basic idea is to decompose the unit sphere into different regions and to argue using epsilon-net arguments that for each region and every vector z from that region, $||Mz||$ is large with high probability. Our spectral bound allows for a reconstruction algorithm of the form $\hat{s} = \text{round}(M' \cdot (Ms + e))$, where s is a 0-1 vector and M' is an appropriate pseudoinverse of M.

1.2 Comparison to Previous Lower Bounds

Dinur and Nissim [10] showed that a mechanism which answers (or allows the user to compute) $O(n \log n)$ arbitrary inner product queries on a database $s \in \mathbb{R}^n$ with noise $o(\sqrt{n})$ is not private. Their attack was subsequently extended to use a linear number of queries [14], allow a small fraction of answers to be arbitrarily distorted [14], and run significantly more quickly [15].

In their simplest form, such inner product queries require the adversary to be able to "name rows", that is, specify a coefficient for each component of the vector s. Thus, the lower bound does not seem to apply to any functionality that is symmetric in the rows of the data set (such as, for example, "counting queries"). It was pointed out in [9] that in databases with more than one entry per row, random inner product queries (on, say, attribute x_d) can be simulated via hashing: for example, the adversary could ask for the sum the function $H(x_1, ..., x_{d-1}) \cdot x_d$ over the whole database, where $H : \{0,1\}^{d-1} \to \{0,1\}$ is an appropriate hash function. This is a symmetric query, but it might seem odd to a statistician (with, e.g., a 2-wise independent hash function).

Using a more algebraic approach, Dwork *et al.* [13] gave a lower bound for differentially private mechanisms based on counting the number of rows that satisfy parity functions. Their attacks also require either that the adversary "name rows", or be able to index individual entries via hashing.

The lower bounds we give for contingency table releases are the first for symmetric functions regularly released by official statistics agencies. As with previous bounds based on reconstruction, we show a lower bound of roughly $\sqrt{n}$ on the average distortion per entry (as long as $n = o(\binom{d}{k})$). This $\sqrt{n}$ behavior is tight, since a small random sample of the database allows counting queries to be answered with about this accuracy, yet clearly precludes reconstruction of the entire database.

1.3 Relating Reconstruction Problems to Privacy Lower Bounds

Our results suggest a general connection between reconstruction problems and lower bounds for private data release. The lower bounds for relaxed privacy definitions proceed by "embedding" an instance of the reconstruction problem for a matrix with correlated rows into a contingency table release problem. We give two such embeddings (or reductions), leading to two differently flavored results.

[1] The entry-wise product of k vectors $u_1, \ldots, u_k \in \mathbb{R}^d$ is the vector in $v \in \mathbb{R}^d$ with entries $v(i) = \prod_j u_j(i)$.

Problem	Upper Bound	Lower Bound (This Paper)
$\mathcal{C}_1$	$\tilde{O}(\min\{\sqrt{n}, \sqrt{d}\})$	$\Omega(\min\{\sqrt{n}, \sqrt{d}\})$
$\mathcal{C}_k$ $(k \geq 2)$	$\tilde{O}(\min\{\sqrt{n}, \sqrt{d^k}\})$	$\tilde{\Omega}\left(\min\left\{\sqrt{n}, \sqrt{d^k}\right\}\right)$

Table 2: *Upper and lower bounds on root mean squared error per cell entry for releasing all k-way contingency tables under (not) strong non-privacy with $k = const$. The upper bound is proved in Proposition 3.3. The lower bounds are proved in Theorems 3.9 and 3.14.*

Let $k \geq 1$ be a constant. Consider a 0-1 matrix $S \in \{0,1\}^{d \times n}$, and let $M^{(k)}$ be the $\binom{d}{k} \times n$ matrix whose rows consist of all the k-way entry-wise products of the rows of S. Let $s \in \{0,1\}^n$.

- **First reduction (Attribute Non-Privacy):** Consider the database $D = (S^\top | s) \in (\{0,1\}^{d+1})^n$. That is, the first d columns of D are given by the rows in the set S, and the last column is s. Suppose that the adversary knows S but wants to learn s — this corresponds to the model, common in the data privacy literature, e.g., [35, 27, 24], of d *nonsensitive* attributes (e.g., demographic information), which can be learned from other sources, and one *sensitive* attribute (e.g., disease). Then the k-way contingency tables of D contain the vector $M^{(k-1)}s$. We show (using our result **(3)** on least singular value) that with high probability over random S, for every $s \in \{0,1\}^n$, any mechanism that approximates k-way contingency tables on D with distortion $\approx \sqrt{n}$ per entry allows the adversary to compute $n - o(n)$ bits of s (that is, to find $\hat{s}$ that agrees in almost all entries with s), as long as $n = o(d^{k-1})$. One can extend the reduction to get a lower bound of $\tilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^{k-1}}\})$ for all n.

  The lower bound applies to any model of privacy which purports to protect individual values of the sensitive attribute (in particular, to differential privacy but also, e.g., the notion of privacy implicit in "K-anonymity" [35] and its variants). Another interesting interpretation of this result is that conjunctions form a "good enough" family of hash functions H for the purposes of the Dinur-Nissim style attack described as above.

- **Second reduction (Strong Non-Privacy):** Consider now a database D given by $diag(s) \cdot S^\top$, where $diag(s)$ is an $n \times n$ diagonal matrix with diagonal s. Because s is a 0-1 vector, this corresponds to a world where person i's data is either S_i or 0^d, according to the ith bit of s (where S_i is the ith column of S). As before, assume that the adversary knows S, but not s. Then the k-way (as opposed to $k+1$ above) contingency tables of D contain the vector $M^{(k)}s$. We show (using again **(3)**) that with high probability over random S, for every $s \in \{0,1\}^n$, any mechanism that approximates k-way contingency tables on D with distortion $\approx \sqrt{n}$ per entry allows the adversary to compute $n - o(n)$ bits of s, as long as $n = o(d^k)$. One can extend the reduction to get a lower bound of $\tilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$ for all n.

  The distribution on databases generated by this reduction is somewhat less natural than in the first reduction, but yields a stronger lower bound. It applies, roughly, to any notion of privacy that seeks to protect any complete row of the database (as opposed to only individual entries). This includes differential privacy (see Lemma 3.4), as well as its relaxations to metrics on probability distributions such as total variation distance or KL divergence.

The schemata above for reducing lower bounds on privacy to reconstruction problems are quite general, and they raise the question: for what types of (natural) correlations on the rows of a random matrix can we bound the least singular values (or Lipschitz coefficients for other norms)? More generally, what properties of a function determine how accurately it can be released privately?

1.4 Preliminaries

We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. $d_H(\cdot, \cdot)$ measures the Hamming distance, and $negl(n)$ denotes a function that is asymptotically smaller than $1/n^c$ for all $c > 0$. $\Pr[\cdot]$, $\mathbb{E}[\cdot]$, $Var[\cdot]$, and $\text{supp}(\cdot)$, denotes probability, expectation, variance, and support of a random variable, respectively. We often add subscripts to $\Pr[\cdot]$ and $\mathbb{E}[\cdot]$ to emphasize the source of randomness.

Vectors are always column vectors. For a vector v, $v^\top$ denotes its transpose (row vector) and $||v||$ denotes its Euclidean norm. v_i denotes the ith entry of the vector v. We use u_v be the unit vector corresponding to v (i.e., $u_v = v/||v||$). For two vectors v_1 and v_2, $\langle v_1, v_2 \rangle$ denotes the inner product of v_1 and v_2. For a matrix M, $tr(M)$ denotes the trace and $||M||_\infty$ denotes the operator norm. Operator norm of M, $||M||_\infty$ equals the maximum eigenvalue of M. Let $\mathbb{I}_d$ denote the identity matrix of dimension d. Let M be an $N \times n$ real matrix with $N \geq n$. The singular values $\sigma_j(M)$ are the eigenvalues of $\sqrt{M^\top M}$ arranged in non-increasing order. Of particular importance in this paper is the smallest singular value $\sigma_n(M) = \inf_{z:||z||=1} ||Mz||$.

Differential Privacy. A database D' is said to be a neighbor of D if it differs from D in exactly one row. A (randomized) algorithm is *differentially private* if neighbor databases induce nearby distributions on the outputs.

Definition 1.1 (ϵ-Differential Privacy [13]). *A randomized algorithm $\mathcal{A}$ is ϵ-differentially private if for all neighboring databases D, D', and for all sets $\mathcal{S}$ of possible outputs, $\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(\epsilon) \cdot \Pr[\mathcal{A}(D') \in \mathcal{S}]$. The probability is taken over the random coins of the algorithm $\mathcal{A}$.*

Let X and Y be random variables taking values in a set $\mathcal{O}$. We use $X \approx_\epsilon Y$ to indicate that random variables X and Y are ϵ-indistinguishable, i.e., $\forall \mathcal{S} \subseteq \mathcal{O}$, $\exp(-\epsilon) \cdot \Pr[Y \in \mathcal{S}] \leq \Pr[X \in \mathcal{S}] \leq \exp(\epsilon) \cdot \Pr[Y \in \mathcal{S}]$. All our results are symmetric, that is, they do not depend on the order of entries in the database D. The size of a database is the number of rows in it. Some of our results are independent of the number of the rows in the database (i.e., they hold even when database is a vector).

In our analysis, we assume that a (private) algorithm $\mathcal{A}$ for a function class $\mathcal{F}$ on input D releases a vector $\mathcal{A}(D) = (\mathcal{A}_1(D), \ldots, \mathcal{A}_{|\mathcal{F}|}(D))$, where each entry in the vector is an estimate of one of predicates in $\mathcal{F}$. This assumption is without loss of generality, because if $\mathcal{A}$ on input D releases some other sanitized structure $\widehat{D}$ then we can define a new private algorithm $\widehat{\mathcal{A}}$ that first runs $\mathcal{A}$ on D and then releases the vector $\mathcal{F}(\widehat{D})$. The perturbation introduced doesn't change by this second step, and therefore, we can think of $\mathcal{A}$ as directly releasing the sanitized vector.

Boolean Conjunctions. The domain for Boolean conjunctions is $\{0,1\}^d$. Each $x \in \{0,1\}^d$ is interpreted as an assignment to d Boolean variables $x_1, \ldots, x_d$. A conjunction predicate $c_v : \{0,1\}^d \to \{0,1\}$ for $v \in \{-1, 0, 1\}^d$ is defined as $c_v(x) = 1$ iff $x_i = 1$ if $v_i = 1$ and $x_i = 0$ if $v_i = -1$ for all $i \in [d]$. The value of v_i indicates whether the variable x_i appears as not negated (if $v_i = 1$), negated (if $v_i = -1$), or absent (if $v_i = 0$). The length of a conjunction predicate is the number of coordinates of v that are non-zero. We will refer to a conjunction predicate of length k as a k-way conjunction. Let $\mathcal{C}_k$ be the function class of all k-way conjunction predicates on variables $x_1, \ldots, x_d$. The size of $\mathcal{C}_k$, $|\mathcal{C}_k| = 2^k \binom{d}{k}$. Let $D \in (\{0,1\}^d)^n$ be a database. Each row is represents information contributed by one individual. The ith column of D contains the assignments to variable x_i. For a predicate $c_v \in \mathcal{C}_k$ define, $c_v(D) = \sum_{x \in D} c_v(x)$. We use $\mathcal{C}_k(D)$ to represent the vector of all the $c_v(D)$'s.

1.5 Known Upper Bounds for Differentially Private Releases

In [4, 13] it was shown that addition of carefully calibrated noise to functions satisfying a Lipschitz condition is enough to ensure differential privacy. Blum *et al.* [4] showed that adding instance-independent random noise

drawn from a normal distribution with mean 0 and standard deviation $\sqrt{2m_k \log(1/\delta)}/\epsilon$ to each entry in $\mathcal{C}_k(D)$ guarantees (ϵ, δ)-differential privacy, and adding instance-independent random noise drawn from a Laplacian distribution with mean 0 and standard deviation $2m_k/\epsilon$ to each entry in $\mathcal{C}_k(D)$ guarantees ϵ-differential privacy. The mean squared (covariance) matrix of the SuLQ mechanism is a simple diagonal matrix with each entry along the diagonal being equal to the variance of the additive noise distribution. Recently, Blum, Ligett, and Roth [5] presented an elegant algorithm that instead of adding direct noise, preserves privacy by adapting the exponential sampling technique of McSherry and Talwar [29]. They show that for every class of predicates the exponential mechanism of McSherry and Talwar can be used to generate a synthetic database that maintains *usefulness* in that with high probability the L_∞ distance between the vector of answers output by the mechanism and the true vector of answers is small. For comparison with our bounds, we show in Corollary 1.5 that for k-way conjunctions each diagonal entry in their mean squared matrix is $\tilde{O}((n^2dk/\epsilon)^{2/3})$.

Blum *et al.* Upper Bound. Blum, Ligett, and Roth designed an ϵ-differentially private algorithm that, given a database D, outputs a new "synthetic" database $\widehat{D}$. Their work provides a high-probability bound on the L_∞ distance between the vector of answers output by the mechanism and the true vector of answers. For comparison with our bounds, we state their result in terms of mean squared error. We start by describing their result. To measure how well $\widehat{D}$ represents D with respect to a specific function class $\mathcal{F}$, they introduce the following notion:

Definition 1.2 ((α, β)-usefulness [5]). *An algorithm $\mathcal{A}$ is* (α, β)-useful *for class of predicates $\mathcal{F}$ and database D if, with probability at least $1 - \beta$, $\mathcal{A}(D)$ outputs a database $\widehat{D}$ that satisfies $|f_j(\widehat{D})/|\widehat{D}| - f_j(D)/|D|| \leq \alpha$ for every $f_j \in \mathcal{F}$.*

Theorem 1.3 ([5]). *Let $\alpha, \beta, \epsilon > 0$. For every class $\mathcal{F}$ of predicates from $\{0,1\}^d$ to $\{0,1\}$, there exists an ϵ-differentially private algorithm $\mathcal{A}$ that is (α, β)-useful for $\mathcal{F}$ and all databases $D \in (\{0,1\}^d)^n$ with*

$$n \geq C \cdot \left(\frac{VCDIM(\mathcal{F}) d \log(1/\alpha)}{\alpha^3 \epsilon} + \frac{\log(1/\beta)}{\epsilon\alpha} \right)$$

entries, where C is a sufficiently large constant. (The algorithm may not be efficient.)

Proposition 1.4 (BLR [5] upper bound). *For a class $\mathcal{F}$ of predicates with VC-dimension equal to $VCDIM(\mathcal{F})$, the Blum, Ligett, and Roth mechanism produces a synthetic database such that for each predicate $f_j \in \mathcal{F}$, the mean squared error of the estimated integer count of f_j is $\tilde{O}(n^2 \cdot VCDIM(\mathcal{F}) \cdot d/\epsilon)^{2/3}$.*

Proof. Let D be a database. Theorem 1.3 shows that if n is large enough, then with probability $1 - \beta$, the mechanism returns a synthetic database $\widehat{D}$ such that the fractional count of every predicate on $\widehat{D}$ is within α of the corresponding fractional count on the real database D.

This translates to an expected square error in the estimated integer counts of at most $(1-\beta)(\alpha n)^2 + \beta n^2 \leq (\alpha n)^2 + \beta n^2$, since a count can be off by at most n. Setting $\beta = \alpha^2$, and assuming that $d \geq 2$, we get that

$$n \geq \frac{Cd \cdot VCDIM(\mathcal{F}) \cdot \log(1/\alpha)}{\alpha^3 \epsilon} \tag{1}$$

gives an expected square error $2\alpha^2 n^2$. Isolating α from Equation 1, and substituting it in $2\alpha^2 n^2$, we get that the expected error for each count is $\tilde{O}(n^2 \cdot VCDIM(\mathcal{F}) \cdot d/\epsilon)^{2/3}$. □

Observing that k-way conjunctions have VC dimension at most $k \log d$, we obtain:

Corollary 1.5. *For k-way conjunctions, the Blum, Ligett, and Roth mechanism produces a synthetic database such that for each conjunction predicate, the mean squared error of the estimated integer count is $\tilde{O}(n^2 \cdot d \cdot k/\epsilon)^{2/3}$.*

2 Lower Bounds on Noise for Differential Privacy

In this section, we prove lower bounds on noise needed to ϵ-differentially privately release all k-way contingency tables. Our results also extend to a relaxation of ϵ-differential privacy (called (ϵ, δ)-differential privacy) that satisfies similar semantics [21]. This extension is simple and we discuss it in Appendix C. Table 3 distinguishes between ϵ-differential privacy and (ϵ, δ)-differential privacy.

We divide differentially private algorithms into broadly three (not necessarily disjoint) categories based on the noise introduction process.

- **Instance-independent additive noise case:** An algorithm $\mathcal{A}$ is in this category if the noise it adds has a fixed distribution (is independent of the database) and is additive. Let Z be some noise distribution, then for all D, $\mathcal{A}(D) = \mathcal{F}(D) + Z$. Therefore, for D' a neighbor of D, $\mathcal{A}(D') = \mathcal{A}(D) + \mathcal{F}(D') - \mathcal{F}(D)$. The SuLQ algorithm of Blum, Dwork, McSherry, and Nissim [4] falls into this category.
- **Unbiased noise case:** An algorithm $\mathcal{A}$ is in this category if for all D, $\mathbb{E}[\mathcal{A}(D)] = \mathcal{F}(D)$. Therefore, for D' a neighbor of D, $\mathbb{E}[\mathcal{A}(D')] = \mathbb{E}[\mathcal{A}(D)] + \mathcal{F}(D') - \mathcal{F}(D)$. Here, there is a noise distribution for every database, but the algorithm is unbiased (i.e., expected value equals the true value).
- **General case:** Unlike the previous two cases, here, we make no assumptions about the algorithm. The algorithm of Blum, Ligett, and Roth [5] falls into this category.

Remark: If an algorithm $\mathcal{A}$ adds instance-independent additive noise, then $\mathbb{E}[\mathcal{A}(D)] = \mathcal{F}(D) + \mathbb{E}[Z]$ (where Z is a random variable independent of D). Note that but for this displacement by $\mathbb{E}[Z]$, algorithms that add instance-independent additive noise case are also unbiased (e.g., if $\mathbb{E}[Z] = 0$ as in the case of SuLQ [4], then an algorithm that adds instance-independent additive noise is also unbiased). The results for instance-independent additive case follow directly from the unbiased case, and we state them in Appendix A.

In the general case, we measure the perturbation introduced by a randomized algorithm $\mathcal{A}$ using the mean squared error matrix $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{F}(D))(\mathcal{A}(D) - \mathcal{F}(D))^\top]$. This is necessary because an algorithm could always add noise such that the output released is always a 0 vector for every database. This clearly satisfies all the privacy requirements and also the variance for each of the $\mathcal{A}_j(D)$'s is 0, but the deviation from the true answer is big. For an unbiased algorithm $\mathcal{A}$ the mean squared error matrix is same as the covariance matrix $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])^\top]$. Table 3 summarizes our results.

In Section 2.1, we consider unbiased ϵ-differentially private algorithms. Let $m_k = \binom{d}{k}$. We show that any unbiased algorithm $\mathcal{A}$ for $\mathcal{C}_k$ that for every database D has an average variance (i.e., average mean squared error) of $o(m_k/(2^k\epsilon^2))$ for $\mathcal{A}(D)$ is not ϵ-differentially private. The idea is to show that for any two neighboring databases D and D', with $\mathcal{C}_k(D') - \mathcal{C}_k(D) = \Delta$ and unit vector $u_\Delta = \Delta/||\Delta||$, the indistinguishability requirement of differential privacy along with the unbiasedness of the algorithm forces both $u_\Delta^\top \Sigma(\mathcal{A}(D))u_\Delta$ (which is the expected squared length of the projection of $\mathcal{A}(D) - \mathcal{C}_k(D)$ on Δ) and $u_\Delta^\top \Sigma(\mathcal{A}(D'))u_\Delta$ to be at least $||\Delta||^2$ (square of the length of Δ). Of particular interest to us are the Δ vectors with large lengths (close to $\sqrt{m_k}$). Then using a careful argument involving geometries of these Δ vectors we show that there exists a database D^* such that the trace of $\Sigma(\mathcal{A}(D^*))$ is at least $m_k^2/(2^k\epsilon^2)$. The result follows by an averaging argument (since, there covariance matrix has m_k diagonal entries, at least one of the diagonal entry is greater than $m_k/(2^k\epsilon^2)$).

In Section 2.2, we consider general ϵ-differentially private algorithms. We show that any algorithm $\mathcal{A}$ for $\mathcal{C}_k$ that for every database D has an average mean squared error of $\min\{o(m_k/(2^k\epsilon^2)), o(n/(2^k\epsilon \log m_k))\}$ for $\mathcal{A}(D)$ is not ϵ-differentially private. The analysis of the general case is trickier, because it is no longer necessary that both $u_\Delta^\top \Sigma(\mathcal{A}(D))u_\Delta$ and $u_\Delta^\top \Sigma(\mathcal{A}(D'))u_\Delta$ be greater than $||\Delta||^2$, but we show that the indistinguishability requirement forces at least one of them to be greater than $||\Delta||^2$. We then look at databases picked uniformly at

Mechanism	U.B. ϵ-diff privacy	L.B. ϵ-diff privacy (This Paper)
Instance-Independent	$O\left(\frac{m_k}{\epsilon}\right)$ [4]	$\Omega\left(\frac{\sqrt{m_k}}{\sqrt{2^k}\epsilon}\right)$
Unbiased	$O\left(\frac{m_k}{\epsilon}\right)$ [4]	$\Omega\left(\frac{\sqrt{m_k}}{\sqrt{2^k}\epsilon}\right)$
General	$\tilde{O}\left(\min\left\{n, \left(\frac{n^2dk}{\epsilon}\right)^{\frac{1}{3}}, \frac{m_k}{\epsilon}\right\}\right)$ [4, 5]	$\Omega\left(\min\left\{\frac{\sqrt{n}}{\sqrt{2^k\epsilon\log m_k}}, \frac{\sqrt{m_k}}{\sqrt{2^k}\epsilon}\right\}\right)$
	U.B. (ϵ,δ)-diff privacy	**L.B. (ϵ,δ)-diff privacy (This Paper)**
Instance-Independent	$O\left(\frac{\sqrt{m_k\log 1/\delta}}{\epsilon}\right)$ [4]	$\Omega\left(\frac{\sqrt{m_k}(1-\delta)}{\sqrt{2^k}\epsilon}\right)$
Unbiased	$O\left(\min\left\{\frac{n}{\sqrt{\delta}}, \left(\frac{n^2dk}{\epsilon}\right)^{\frac{1}{3}} + \frac{n}{\sqrt{\delta}}, \frac{\sqrt{m_k\log 1/\delta}}{\epsilon}\right\}\right)$ [4]	$\Omega\left(\min\left\{\frac{\sqrt{n}(1-\delta)}{\sqrt{2^k\epsilon\log m_k}}, \frac{\sqrt{m_k}(1-\delta)}{\sqrt{2^k}\epsilon}\right\}\right)$
General	$\tilde{O}\left(\min\left\{n, \left(\frac{n^2dk}{\epsilon}\right)^{\frac{1}{3}}, \frac{\sqrt{m_k\log 1/\delta}}{\epsilon}\right\}\right)$ [4, 5]	$\Omega\left(\min\left\{\frac{\sqrt{n}(1-\delta)}{\sqrt{2^k\epsilon\log m_k}}, \frac{\sqrt{m_k}(1-\delta)}{\sqrt{2^k}\epsilon}\right\}\right)$

Table 3: Upper and lower bounds on the root average mean squared error per cell entry for releasing all k-way contingency tables (for $1 \le k \le d$). Theorems A.5, 2.9, and 2.26 prove our results for the ϵ-differential privacy case. The results for (ϵ,δ)-differential privacy case are proved in Theorems C.2 and C.8. Here, $m_k = \binom{d}{k}$.

random, and show by an application of matrix-valued Chernoff bound that with high probability the trace of the mean squared matrix of a random database is at least $\min\{m_k^2/(2^k\epsilon^2), nm_k/(2^k\epsilon\log m_k)\}$.

2.1 Lower Bounds on Noise for Unbiased ϵ-differential Privacy

In the analysis instead of conjunctions we consider inner products over the domain $\{-1,1\}^d$. In Appendix B, we provide the relation between these two problems. An inner product predicate $i_v : \{-1,1\}^d \to \{-1,1\}$ is defined as $i_v(x) = \prod_i x_i \cdot v_i$, where the value of v_i indicates whether x_i is present (if $v_i = 1$) or not (if $v_i = 0$). We call the corresponding function class $\mathcal{I}_k$. The size of $\mathcal{I}_k$, $|\mathcal{I}_k| = m_k$. Let D be a database from $(\{-1,1\}^d)^n$. Define, $i_v(D) = \sum_{x\in D} i_v(x)$. Let $\mathcal{I}_k(D)$ be the vector of all the $i_v(D)$'s. We analyze 1-way inner products in Section 2.1.1 and higher way inner products in Section 2.1.2. The analysis of the 1-way case is simple and direct, but it provides key insights that would prove useful for analyzing higher way inner products.

2.1.1 1-way Inner Products - Lower Bounds for Unbiased ϵ-differentially Privacy

Let $D \in (\{-1,1\}^d)^n$ be a database. Consider the problem of privately releasing $\mathcal{I}_1(D)$. Notice that $\mathcal{I}_1(D)$ is a vector whose ith entry is just the sum of the ith column of D. Consider a neighbor D' of D, and let $\Delta = \mathcal{I}_1(D') - \mathcal{I}_1(D)$. Now, $\Delta \in \{-2,0,2\}^d$. However, instead of working over Δ's that are from $\{-2,0,2\}^d$, we restrict ourselves to Δ's that are from $\{-2,2\}^d$. In other words, we consider only those neighbors D'' of D such that $\mathcal{I}_1(D'') - \mathcal{I}_1(D) \in \{-2,2\}^d$. In the remainder of this section, $\Delta \in \{-2,2\}^d$.

Let $\Sigma(\mathcal{A}(D))$ be the covariance matrix of $\mathcal{A}(D)$. We start by proving that for any unbiased ϵ-differentially private algorithm $\mathcal{A}$, both $\Delta^\top\Sigma(\mathcal{A}(D))\Delta$ and $\Delta^\top\Sigma(\mathcal{A}(D'))\Delta$ are $\Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2)$, where $\Delta = \mathbb{E}[\mathcal{A}(D')] - \mathbb{E}[\mathcal{A}(D)]$. The proof uses the fact that projections onto Δ need to be ϵ-indistinguishable for $\mathcal{A}(D)$ and $\mathcal{A}(D')$.

Lemma 2.1. *Let $\mathcal{A}$ be any unbiased ϵ-differentially private algorithm for $\mathcal{I}_1$. Let $\mathcal{A}(D) \approx_\epsilon \mathcal{A}(D')$. Let $\mathcal{I}_1(D') = \mathcal{I}_1(D) + \Delta$ for some $\Delta \in \{-2,2\}^d$. Then, $\mathbb{E}[\langle\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)], \Delta\rangle^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2) = \Omega(d^2/\epsilon^2)$ and $\mathbb{E}[\langle\mathcal{A}(D') - \mathbb{E}[\mathcal{A}(D')], \Delta\rangle^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2) = \Omega(d^2/\epsilon^2)$.*

Proof. We first prove the one-dimensional version of this lemma. That is we show that if $\mathcal{A}$ is unbiased ϵ-differentially private such that for any two neighboring databases D and D', $|\mathbb{E}[\mathcal{A}(D)] - \mathbb{E}[\mathcal{A}(D')]| \le 1$, then both $Var[\mathcal{A}(D)]$ and $Var[\mathcal{A}(D')]$ are $\Omega(1/\epsilon^2)$.

Lemma 2.2. *Let $\mathcal{A}(D)$ and $\mathcal{A}(D')$ be distributions over $\mathbb{R}$. Let $\mathcal{A}(D) \approx_\epsilon \mathcal{A}(D')$. Let $\mathbb{E}[\mathcal{A}(D)] = p$ and $\mathbb{E}[\mathcal{A}(D')] = p+1$. Then, $\mathbb{E}[\mathcal{A}(D)^2] = \Omega(1/\epsilon^2)+p^2$ and $\mathbb{E}[\mathcal{A}(D')^2] = \Omega(1/\epsilon^2)+p^2$. Therefore, $Var[\mathcal{A}(D)] = \Omega(1/\epsilon^2)$ and $Var[\mathcal{A}(D')] = \Omega(1/\epsilon^2)$. The expectation is over the randomness of the algorithm.*

Proof. Define $X = \mathcal{A}(D) - p$ and $Y = \mathcal{A}(D') - p$. Lets assume that X and Y are continuous random variables with support over $\mathbb{R}$ (proofs for other situations are quite similar). Define $a_i = \Pr[X \in [i, i+1)]$ and $b_i = \Pr[Y \in [i, i+1)]$. Note that

$$\sum_{i \in \mathbb{Z}} a_i = \sum_{i \in \mathbb{Z}} b_i = 1.$$

By requirements of differential privacy, we have $e^{-\epsilon} b_i \leq a_i \leq e^{\epsilon} a_i$ for all $i \in \mathbb{Z}$. Let $\mathrm{Int}_i = [i, i+1)$. Now,

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\mathbb{R}} z \Pr[X = z] dz = \sum_{i=-\infty}^{\infty} \int_{\mathrm{Int}_i} z \Pr[X = z] dz \geq \sum_{i=0}^{\infty} i a_i + \sum_{i=-\infty}^{0} i a_i \\
&= \sum_{i=0}^{\infty} i a_i - \sum_{i=0}^{\infty} i a_{-i} \geq -\sum_{i=0}^{\infty} e^{\epsilon} i b_{-i} + \sum_{i=0}^{\infty} e^{-\epsilon} i b_i = -\sum_{i=0}^{\infty} (e^{-\epsilon} + e^{\epsilon} - e^{-\epsilon}) i b_{-i} + \sum_{i=0}^{\infty} e^{-\epsilon} i b_i \\
&= e^{-\epsilon} \left(\sum_{i=0}^{\infty} i b_i - \sum_{i=0}^{\infty} i b_{-i} \right) - (e^{\epsilon} - e^{-\epsilon}) \sum_{i=0}^{\infty} i b_{-i} = e^{-\epsilon} \left(\sum_{i=0}^{\infty} i b_i + \sum_{i=-\infty}^{-1} i b_i \right) - (e^{\epsilon} - e^{-\epsilon}) \sum_{i=0}^{\infty} i b_{-i} \\
&= e^{-\epsilon} \mathbb{E}[Y] - (e^{\epsilon} - e^{-\epsilon}) \sum_{i=0}^{\infty} i b_{-i}.
\end{aligned}$$

Since $\mathbb{E}[X] = 0$ and $\mathbb{E}[Y] = 1$, therefore, for small ϵ,

$$\sum_{i=0}^{\infty} i b_{-i} \geq (e^{-\epsilon})/(e^{\epsilon} - e^{-\epsilon}) = \Omega(1/\epsilon).$$

Define, a new random variable Y_- as

$$\Pr[Y_- = z] = \Pr[Y = -z \mid Y \leq 0].$$

Then $\mathbb{E}[Y_-] \geq \frac{1}{\Pr[Y \leq 0]} \sum_{i=0}^{\infty} i b_{-i}$. Rearranging the terms and using the fact that $\mathbb{E}[Y_-] \leq \sqrt{\mathbb{E}[Y_-^2]}$,

$$\sum_{i=0}^{\infty} i b_{-i} \leq \mathbb{E}[Y_-] \Pr[Y \leq 0] \leq \sqrt{\mathbb{E}[Y_-^2]} \Pr[Y \leq 0].$$

Using the bound for $\sum_{i=0}^{\infty} i b_{-i}$, we get that

$$\mathbb{E}[Y_-^2] = \Omega\left(\left(\frac{1}{\Pr[Y \leq 0]} \right)^2 \right).$$

Now, define $\Pr[X_+ = z] = \Pr[X = z \mid X \geq 1]$. Using similar analysis as above gives that

$$\mathbb{E}[X_+^2] = \Omega\left(\left(\frac{1}{\epsilon \Pr[X \geq 1]} \right)^2 \right).$$

Now,

$$\mathbb{E}[X^2] = \sum_{i=-\infty}^{\infty} \int_{\text{Int}_i} z^2 \Pr[X = z] dz.$$

In particular,

$$\mathbb{E}[X_+^2] = \sum_{i=1}^{\infty} \int_{\text{Int}_i} \frac{z^2 \Pr[X = z] dz}{\Pr[X \geq 1]} \leq \frac{\mathbb{E}[X^2]}{\Pr[X \geq 1]}.$$

Similarly, we get that

$$\mathbb{E}[Y_-^2] = \sum_{i=-\infty}^{-1} \int_{\text{Int}_i} \frac{z^2 \Pr[Y = z] dz}{\Pr[Y \leq 0]} \leq \frac{\mathbb{E}[Y^2]}{\Pr[Y \leq 0]}.$$

Now substituting the lower bound for $\mathbb{E}[X_+^2]$ and $\mathbb{E}[Y_-^2]$, we get that

$$\Omega\left(\frac{1}{\epsilon^2}\right) = \mathbb{E}[X^2] \Pr[X \geq 1] \quad \text{and} \quad \Omega\left(\frac{1}{\epsilon^2}\right) = \mathbb{E}[Y^2] \Pr[Y \leq 0].$$

Hence, $\mathbb{E}[X^2] = \Omega(1/\epsilon^2)$ and $\mathbb{E}[Y^2] = \Omega(1/\epsilon^2)$. Re-substituting X and Y in terms of $\mathcal{A}(D)$ and $\mathcal{A}(D')$ completes the proof. □

We now extend Lemma 2.2 to the higher dimensional case. The proof has similar structure to Lemma 2.2. We will just present the main differences. The idea is as follows: Define, $X = \langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta \rangle$ and $Y = \langle \mathcal{A}(D') - \mathcal{I}_1(D), \Delta \rangle$. Repeating same arguments as in the previous lemma, we get that

$$\mathbb{E}[X] \geq e^{-\epsilon} \mathbb{E}[Y] - (e^{\epsilon} - e^{-\epsilon}) \sum_{i=1}^{\infty} i b_{-i}.$$

Since $\mathbb{E}[Y] = \Delta^\top \Delta = 4d$. Therefore, for small ϵ,

$$\sum_{i=1}^{\infty} i b_{-i} = \Omega\left(\frac{\Delta^\top \Delta}{\epsilon}\right) = \Omega\left(\frac{d}{\epsilon}\right).$$

As in Lemma 2.2 we define random variables Y_- and X_+. By arguments similar to Lemma 2.2 we can show that

$$\mathbb{E}[X_+^2] = \Omega\left(\left(\frac{d}{\epsilon \Pr[X \geq 1]}\right)^2\right) \quad \text{and} \quad \mathbb{E}[Y_-^2] = \Omega\left(\left(\frac{d}{\epsilon \Pr[Y \leq 0]}\right)^2\right).$$

As in Lemma 2.2,

$$\mathbb{E}[X_+^2] \leq \frac{\mathbb{E}[X^2]}{\Pr[X \geq 1]} \quad \text{and} \quad \mathbb{E}[Y_-^2] \leq \frac{\mathbb{E}[Y^2]}{\Pr[Y \leq 0]}.$$

Therefore, we now get that

$$\Omega\left(\frac{d^2}{\epsilon^2}\right) = \mathbb{E}[X^2] \Pr[X \geq 1] \quad \text{and} \quad \Omega\left(\frac{d^2}{\epsilon^2}\right) = \mathbb{E}[Y^2] \Pr[Y < 0].$$

As in Lemma 2.2, we can argue that $\mathbb{E}[X^2] = \Omega(d^2/\epsilon^2)$ and $\mathbb{E}[Y^2] = \Omega(d^2/\epsilon^2)$. Therefore, $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta \rangle^2] = \Omega(d^2/\epsilon^2)$ and $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_1(D), \Delta \rangle^2] = \Omega(d^2/\epsilon^2)$. Also,

$$\begin{aligned}
\mathbb{E}[\langle \mathcal{A}(D') - \mathbb{E}[\mathcal{A}(D')], \Delta \rangle^2] &= \mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_1(D) - \Delta, \Delta \rangle^2] \\
&= \mathbb{E}[(\Delta^\top (\mathcal{A}(D') - \mathcal{I}_1(D)) - d)^2] \\
&= \mathbb{E}[(Y - d)^2] = \mathbb{E}[Y^2] + d^2 - 2d\,\mathbb{E}[Y] \\
&\geq \mathbb{E}[Y^2] + d^2 - 2d\sqrt{\mathbb{E}[Y^2]} = \Omega(d^2/\epsilon^2).
\end{aligned}$$

The last line follows because $\mathbb{E}[Y^2] = \Omega(d^2/\epsilon^2)$. $\square$

The proof of the following proposition relies on the fact that there exists a database D such that for every $\Delta \in \{-2, 2\}^d$, D has a close neighbor D_Δ such that $\mathcal{I}_1(D_\Delta) - \mathcal{I}_1(D) = \Delta$. Let D be any database which has at least a row of both -1^d and 1^d. We now show that there exists a database D_Δ at Hamming distance 2 from D such that for every $\Delta = \{-2, 2\}^d$, $\mathcal{I}_1(D_\Delta) - \mathcal{I}_1(D) = \Delta$. Consider any vector $\Delta \in \{-2, 2\}^d$. Let D_Δ be a database which is identical to D except that the row -1^d is replaced by $\Delta/2$ and the row 1^d is replaced by $\Delta/2$. The Hamming distance between D and D_Δ is 2 and $\mathcal{I}_1(D_\Delta) - \mathcal{I}_1(D) = \Delta$. The idea now is to use the fact that the set of Δ's (which contains every vector from $\{-2, 2\}^d$) contains an orthonormal (Hadamard) basis.

Proposition 2.3 (Unbiased case: 1-way inner products). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^d$ be any unbiased ϵ-differentially private algorithm for $\mathcal{I}_1$. Let D be any database which has at least a row of both -1^d and 1^d. Then, $tr(\Sigma(\mathcal{A}(D))) = \Omega(d^2/\epsilon^2)$.*

Proof. The covariance matrix $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])^\top]$ (i.e., the (i, j)th entry looks like $\mathbb{E}[(\mathcal{A}_i(D) - \mathbb{E}[\mathcal{A}_i(D)])(\mathcal{A}_j(D) - \mathbb{E}[\mathcal{A}_j(D)])]$). Consider any vector $\Delta \in \{-2, 2\}^d$. Since $\mathcal{A}$ is differentially private, by Claim 2.23, $\mathcal{A}(D) \approx_{2\epsilon} \mathcal{A}(D_\Delta)$ (where D_Δ is as defined above). From Lemma 2.1[2], we know that $\Delta^\top \Sigma(\mathcal{A}(D))\Delta = \Omega(d^2/\epsilon^2)$ and therefore $u_\Delta^\top \Sigma(\mathcal{A}(D))u_\Delta = \Omega(d/\epsilon^2)$ (where u_Δ is the unit vector corresponding to Δ). This holds for every $\Delta \in \{-2, 2\}^d$. Consider an orthonormal basis $u_1, \ldots, u_d$ such that $u_i^\top \Sigma(\mathcal{A}(D))u_i = \Omega(d/\epsilon^2)$ for all $i \in [d]$ (one such example is the Hadamard basis). Identity matrix $\mathbb{I}_d = \sum_{i=1}^d u_i u_i^\top$. Now,

$$\sum_{i=1}^{d} tr(u_i^\top \Sigma(\mathcal{A}(D))u_i) = \sum_{i=1}^{d} tr(\Sigma(\mathcal{A}(D))u_i u_i^\top) = tr(\Sigma(\mathcal{A}(D)) \cdot \mathbb{I}_d) = tr(\Sigma(\mathcal{A}(D))).$$

Note that $tr(u_i^\top \Sigma(\mathcal{A}(D))u_i) = tr(\Sigma(\mathcal{A}(D))u_i u_i^\top)$ as trace is cyclically invariant. This implies that $tr(\Sigma(\mathcal{A}(D)))$ $= \Omega(d^2/\epsilon^2)$. $\square$

2.1.2 k-way Inner Products - Lower Bounds for Unbiased ϵ-differential Privacy

Since the results in this subsection will be independent of n, for simplicity we assume $n = 1$ (see the proof of Theorem 2.9 for the simple extension to $n > 1$).

Inner-Products Over the Domain $\{-1, 0, 1\}^d$. First lets consider inner products over the domain $\{-1, 0, 1\}^d$. Let $\mathcal{K}_k$ be the function class of all k-way inner product predicates over the domain $\{-1, 0, 1\}^d$. Let $r \in \{-1, 1\}^d$ be a random vector with independent entries taking values -1 and 1 with probability $1/2$. Let $m_k = \binom{d}{k}$. Define a random vector z of length m_k as $z_r = \mathcal{I}_k(r)$. Each entry in z_r is set to 1 with probability $1/2$ and -1 with probability $1/2$ (but the entries are not independent of each other). Define a matrix $B = \mathbb{E}_{z_r}[z_r z_r^\top]$, where the randomness is over z_r.

Lemma 2.4. *Let $z_r = \mathcal{I}_k(r)$. Then, $\mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_k}$, where $\mathbb{I}_{m_k}$ is an identity matrix of dimension m_k.*

Proof. We prove the lemma for $k = 2$ (2-way case). The proofs for higher k's follow similarly. Consider 2-way inner products. Let r_i denote the ith entry in r. Now $z_r = (z_{1,1}, z_{1,2}, \ldots, z_{d-1,d})$, where $z_{i,j} = r_i r_j$. Then, for $a \neq b$ and $c \neq d$,

$$\mathbb{E}_{z_r}[z_{a,b} z_{c,d}] = \begin{cases} 1 \text{ if } \{a, b\} = \{c, d\}, \\ 0 \text{ otherwise.} \end{cases}$$

[2]Substitute D_Δ for D' in Lemma 2.1. As the Hamming distance between D_Δ and D is two ϵ gets replaced by 2ϵ.

Note that if $\{a,b\} = \{c,d\}$, then $\mathbb{E}_{z_r}[z_{a,b}z_{c,d}] = \mathbb{E}_{z_r}[z_{a,b}^2] = 1$. If $\{a,b\} \neq \{c,d\}$, then there are three cases: if a,b,c,d are all disjoint then $\mathbb{E}_{z_r}[z_{a,b}z_{c,d}] = \mathbb{E}_{z_r}[z_{a,b}]\,\mathbb{E}_{z_r}[z_{c,d}] = 0$, if $a = c$ then $\mathbb{E}_{z_r}[z_{a,b}z_{c,d}] = \mathbb{E}_{z_r}[(r_a^2)(r_b)(r_d)] = \mathbb{E}_{z_r}[r_a^2]\,\mathbb{E}_{z_r}[r_b]\,\mathbb{E}_{z_r}[r_d] = 0$, and if $b = d$ then $\mathbb{E}_{z_r}[z_{a,b}z_{c,d}] = \mathbb{E}_{z_r}[r_a]\,\mathbb{E}_{z_r}[r_c]\,\mathbb{E}_{z_r}[r_d^2] = 0$. Therefore, $\mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_2}$. □

The following lemma proves an extension of Lemma 2.1 to the k-way case.

Lemma 2.5. *Let $\mathcal{A}$ be any unbiased ϵ-differentially private algorithm for $\mathcal{K}_k$. Let $\mathcal{A}(D_b) \approx_\epsilon \mathcal{A}(D_b')$. Let $\mathcal{K}_k(D_b') = \mathcal{K}_k(D_b) + z$ for some $z \in \text{supp}(z_r)$. Then, $\mathbb{E}[\langle \mathcal{A}(D_b) - \mathbb{E}[\mathcal{A}(D_b)], z\rangle^2] = \Omega(\langle z,z\rangle^2/\epsilon^2)$ and $\mathbb{E}[\langle \mathcal{A}(D_b') - \mathbb{E}[\mathcal{A}(D_b')], z\rangle^2] = \Omega(\langle z,z\rangle^2/\epsilon^2)$.*

Proof. Same as Lemma 2.1. We want to show $\mathbb{E}[\langle \mathcal{A}(D_b) - \mathcal{K}_k(D_b), z\rangle^2] = \Omega(\langle z,z\rangle^2/\epsilon^2)$ and $\mathbb{E}[\langle \mathcal{A}(D_b') - \mathcal{K}_k(D_b'), z\rangle^2] = \Omega(\langle z,z\rangle^2/\epsilon^2)$. Compared to Lemma 2.1, z plays the role of Δ. □

The following simple proposition lower bounds the trace of the covariance matrix. The proof follows by combining the above two lemmas.

Proposition 2.6 (Unbiased case: k-way inner products over $\{-1,0,1\}^d$). *Let $\mathcal{A} : \{-1,0,1\}^d \to \mathbb{R}^{m_k}$ be any unbiased ϵ-differentially private algorithm for $\mathcal{K}_k$. Let $D_b = 0^d$. Then, $tr(\Sigma(\mathcal{A}(D_b))) = \Omega(m_k^2/\epsilon^2)$.*

Proof. Let $D_b = 0^d$. Let $m_k = \binom{d}{k}$. Let $\mathcal{T}$ be the set of all $z \in \{-1,1\}^{m_k}$ such that there exists a neighbor D_b' of D_b with $\mathcal{K}_k(D_b') - \mathcal{K}_k(D_b) = z$. Note that $\mathcal{T} = \text{supp}(z_r)$. Let $\Sigma(\mathcal{A}(D_b)) = \mathbb{E}[(\mathcal{A}(D_b) - \mathcal{K}_k(D_b))(\mathcal{A}(D_b) - \mathcal{K}_k(D_b))^\top]$. Then, expected value (expectation over random z_r) of $z_r^\top \Sigma(\mathcal{A}(D_b))z_r$ is

$$\begin{aligned}
\mathop{\mathbb{E}}_{z_r}[z_r^\top \Sigma(\mathcal{A}(D_b))z_r] &= \mathop{\mathbb{E}}_{z_r}[tr(z_r^\top \Sigma(\mathcal{A}(D_b))z_r)] \\
&= \mathop{\mathbb{E}}_{z_r}[tr(\Sigma(\mathcal{A}(D_b))z_r z_r^\top)] = tr(\Sigma(\mathcal{A}(D_b))B) \\
&\leq tr(\Sigma(\mathcal{A}(D_b)))||B||_\infty = tr(\Sigma(\mathcal{A}(D_b))).
\end{aligned}$$

The last equality follows because $B = \mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_k}$, and thus $||B||_\infty = 1$. Since for $\forall z \in \text{supp}(z_r)$, $\langle z,z\rangle = m_k$, from Lemma 2.5 we get that $\forall z \in \text{supp}(z_r)$,

$$\mathbb{E}[\langle \mathcal{A}(D_b) - \mathcal{K}_k(D_b), z\rangle^2] = z^\top \Sigma(\mathcal{A}(D_b))z = \Omega(m_k^2/\epsilon^2).$$

Since the previous statement is true for all $z \in \text{supp}(z_r)$, therefore,

$$\mathop{\mathbb{E}}_{z_r}[z_r^\top \Sigma(\mathcal{A}(D_b))z_r] = \mathop{\mathbb{E}}_{z_r}[\mathbb{E}[\langle \mathcal{A}(D_b) - \mathcal{K}_k(D_b), z_r\rangle^2]] = \Omega(m_k^2/\epsilon^2).$$

Therefore, $tr(\Sigma(\mathcal{A}(D_b))) = \Omega(m_k^2/\epsilon^2)$. □

Going from Domain $\{-1,0,1\}^d$ to $\{-1,1\}^d$. The previous argument relied on starting from a database $D_b = 0^d$. We now consider databases from the domain $\{-1,1\}^d$.

Lemma 2.7. *Let $\mathcal{A}$ be any unbiased ϵ-differentially private algorithm for $\mathcal{I}_k$. Let $D_c, D_c', \widehat{D}$ be the databases as defined above. Let $z = \mathcal{I}_k(D_c') - \mathcal{I}_k(\widehat{D})$ and $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o,o\rangle$. Then, $\mathbb{E}[\langle \mathcal{A}(D_c) - \mathbb{E}[\mathcal{A}(D_c)], \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2/\epsilon^2)$.*

Proof. Using arguments similar to Lemma 2.1 shows that $(\pi\tilde{z})^\top \Sigma(\mathcal{A}(D_c))(\pi\tilde{z}) = \Omega(\langle \pi\tilde{z}, \pi\tilde{z}\rangle^2/\epsilon^2)$. Since $\pi\tilde{z} = \pi z$, we get the desired result. □

Let $D_c = 1^d$. Let $D'_c \in \{-1,1\}^d$ be a neighbor of D_c. Let $\widehat{D} = 0^d$ be an *imaginary* database. Then,

$$\mathcal{I}_k(D'_c) - \mathcal{I}_k(D_c) = \mathcal{I}_k(D'_c) - \mathcal{I}_k(\widehat{D}) + \mathcal{I}_k(\widehat{D}) - \mathcal{I}_k(D_c).$$

Let $\tilde{z} = \mathcal{I}_k(D'_c) - \mathcal{I}_k(D_c)$ and $z = \mathcal{I}_k(D'_c) - \mathcal{I}_k(\widehat{D})$. $\mathcal{I}_k(D_c) = 1^{m_k}$, and $\mathcal{I}_k(\widehat{D}) = 0^{m_k}$. Let $o = 1^{m_k}$. Then, $\tilde{z} = z - o$. Let $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o\rangle$ be the orthogonal projection matrix onto the orthogonal complement of o. Note that $oo^\top$ is a $m_k \times m_k$ matrix of all 1's. Let $\Sigma(\mathcal{A}(D_c)) = \mathbb{E}[(\mathcal{A}(D_c) - \mathcal{I}_k(D_c))(\mathcal{A}(D_c) - \mathcal{I}_k(D_c))^\top]$. The idea now is to extend Lemma 2.1 to show that $\mathbb{E}[\langle \mathcal{A}(D_c) - \mathbb{E}[\mathcal{A}(D_c)], \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2/\epsilon^2)$ (Lemma 2.7), which in turn can be used to show the following.

Proposition 2.8 (Unbiased case: k-way inner products over $\{-1,1\}^d$). *Let $\mathcal{A} : \{-1,1\}^d \to \mathbb{R}^{m_k}$ be any unbiased ϵ-differentially private algorithm for $\mathcal{I}_k$. Let $D_c = 1^d$. Then, $tr(\Sigma(\mathcal{A}(D_c))) = \Omega(m_k^2/\epsilon^2)$.*

Proof. Let $D_c = 1^d$. Since D'_c can be any vector from $\{-1,1\}^d$ and $\mathcal{I}_k(\widehat{D}) = 0^{m_k}$, the set of vectors $z = \mathcal{I}_k(D'_c) - \mathcal{I}_k(\widehat{D})$ is exactly $\text{supp}(z_r)$. As in Proposition 2.6 (with expectation over random z_r),

$$\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r)] \le tr(\pi^\top \Sigma(\mathcal{A}(D_c))\pi)||B||_\infty = tr(\pi^\top \Sigma(\mathcal{A}(D_c))\pi) \le tr(\Sigma(\mathcal{A}(D_c))).$$

The last inequality follows because π is an orthogonal projection matrix. Note that from Lemma 2.4, $B = \mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_k}$. Also $\forall z \in \text{supp}(z_r)$ from Lemma 2.5,

$$(\pi z)^\top \Sigma(\mathcal{A}(D_c))(\pi z) = \Omega(\langle \pi z, \pi z\rangle^2/\epsilon^2).$$

Now,

$$\langle z, z\rangle = z^\top \pi z + z^\top(\mathbb{I}_{m_k} - \pi)z = z^\top \pi^\top \pi z + \langle z, o\rangle^2/m_k = \langle \pi z, \pi z\rangle + \langle z, o\rangle^2/m_k.$$

If you look at the expected value of $z_r^\top \pi z_r$,

$$\mathop{\mathbb{E}}_{z_r}[z_r^\top \pi z_r] = \mathop{\mathbb{E}}_{z_r}[tr(\pi z_r z_r^\top)] = tr(\pi \mathbb{I}_{m_k}) = tr(\pi) = m_k - 1.$$

Now for all $z \in \text{supp}(z_r)$, $z^\top \pi z \le z^\top z = m_k$. Therefore, $z_r^\top \pi z_r$ is a random variable whose range is between $[0, m_k]$ and with expectation of $m_k - 1$. If p is the probability that $z_r^\top \pi z_r$ takes a value greater than $m_k - 2$, then

$$m_k - 1 = \mathop{\mathbb{E}}_{z_r}[z_r^\top \pi z_r] \le p m_k + (1-p)(m_k - 2) \Rightarrow 1/2 \le p.$$

We can expand $\mathbb{E}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r)]$ as

$$\begin{aligned}\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r)] = \mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r) \mid z_r^\top \pi z_r \ge m_k - 2] \Pr[z_r^\top \pi z_r \ge m_k - 2] + \\ \mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r) \mid z_r^\top \pi z_r \le m_k - 2] \Pr[z_r^\top \pi z_r \le m_k - 2].\end{aligned}$$

From above arguments we get that $\Pr_{z_r}[z_r^\top \pi z_r \ge m_k - 2] \ge 1/2$, therefore

$$\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r)] \ge \mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r) \mid z_r^\top \pi z_r \ge m_k - 2]\frac{1}{2} + 0 = \Omega((m_k - 2)^2/(2\epsilon^2)).$$

Since $\mathbb{E}_{z_r}[(\pi z_r)^\top \Sigma(\mathcal{A}(D_c))(\pi z_r)] \le tr(\Sigma(\mathcal{A}(D_c)))$, we get that $tr(\Sigma(\mathcal{A}(D_c))) = \Omega(m_k^2/\epsilon^2)$. □

The result for $\mathcal{I}_k$ can be extended to $\mathcal{C}_k$ (see Corollary B.2). We immediately get the following result.

Theorem 2.9 (Unbiased case). *Let $m_k = \binom{d}{k}$. Any unbiased algorithm $\mathcal{A}$ for releasing all k-way inner products that for every database $D \in (\{-1,1\}^d)^n$ has an average variance of $o(m_k/\epsilon^2)$ for $\mathcal{A}(D)$ is not ϵ-differentially private. Also, any unbiased algorithm $\mathcal{A}$ for releasing all k-way conjunctions that for every database $D \in (\{0,1\}^d)^n$ has an average variance of $o(m_k/(2^k\epsilon^2))$ for $\mathcal{A}(D)$ is not ϵ-differentially private.*

Proof. In the above discussion we considered databases where $n = 1$. However, there is an easy extension to databases where $n > 1$. Let $D \in (\{-1,1\}^d)^n$ be a database with a row of 1^d. We can repeat the above arguments to show $tr(\Sigma(\mathcal{A}(D))) = \Omega(m_k^2/\epsilon^2)$. For the proof, we restrict our attention to those neighbors of D which are obtained by replacing the row of 1^d in D by a vector from $\{-1,1\}^d$. □

2.2 Lower Bounds on Noise for General ϵ-differential Privacy

In this section, we prove lower bounds on the perturbation introduced by any differentially private algorithm for $\mathcal{I}_k$ and $\mathcal{C}_k$. Again our analysis looks at the related problem of releasing inner products.

2.2.1 1-way Inner Products - Lower Bounds for General ϵ-differential Privacy

We initially prove the lower bound by setting $\epsilon = 1/2$. In Section 2.3, we strengthen the lower bound by introducing ϵ into it. As in Section 2.1.1, we only consider Δ's from $\{-2,2\}^d$. Let $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{I}_1(D))(\mathcal{A}(D) - \mathcal{I}_1(D))^\top]$ be the mean squared error matrix. Let $\mathcal{A}(D) \approx_{1/2} \mathcal{A}(D')$ and $\Delta = \mathcal{I}(D') - \mathcal{I}(D)$. Unlike in the unbiased case (Lemma 2.1) it is not necessarily that both $\Delta^T\Sigma(\mathcal{A}(D))\Delta$ and $\Delta^T\Sigma(\mathcal{A}(D'))\Delta$ be $\Omega(d^2)$, but the following lemma shows that at least one of them is $\Omega(d^2)$.

Lemma 2.10. *Let $\mathcal{A}$ be any $1/2$-differentially private algorithm for $\mathcal{I}_1$. Let $\mathcal{A}(D) \approx_{1/2} \mathcal{A}(D')$. Let $\mathcal{I}_1(D') = \mathcal{I}_1(D) + \Delta$ for some $\Delta \in \{-2,2\}^d$. Then $\min\{\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta\rangle^2], \mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_1(D'), \Delta\rangle^2]\} = \Omega(d^2)$.*

Proof. We prove the lemma in a slightly general setting (under the notion of (ϵ,δ)-privacy from Appendix C). Lemma 2.10 follows from setting $X = \langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta\rangle$, $Y = \langle \mathcal{A}(D') - \mathcal{I}_1(D'), \Delta\rangle$, $\delta = 0$, and $a = \Delta^\top\Delta = 4d$ in the following lemma. If two random variables, X and Y are (ϵ,δ)-indistinguishable, then the statistical difference[3] between X and Y is at most $e^{1/2} - 1 + \delta$.

Lemma 2.11 (Lemma 2.10, restated). *Suppose X, Y are real-valued random variables with statistical difference at most $e^{1/2} - 1 + \delta$. Then, for all real numbers a, at least one of $\mathbb{E}[X^2]$ or $\mathbb{E}[(Y-a)^2]$ is $\Omega(a^2(1-\delta)^2)$.*

Proof. Since X and Y have statistical difference at most $e^{1/2} - 1 + \delta$, we can find random variables X', Y', U such that X' and Y' have the same marginal distributions as X and Y respectively, and $X' = Y' = U$ with probability at least $2 - e^{1/2} - \delta$. Moreover, if E is the event that $X' = Y' = U$, we may choose U so that it is independent of the event E. (See, for example, the proof Lemma 3.1.8 in Vadhan's thesis [38] for a proof of this.)

We can bound the expectation of X in terms of the expectation of U:

$$\mathbb{E}[X] = \mathbb{E}[X'] = \mathbb{E}[X'|E]\Pr[E] + \mathbb{E}[X'|\overline{E}]\Pr[\overline{E}] \geq (2 - e^{1/2} - \delta)\,\mathbb{E}[X'|E] = (2 - e^{1/2} - \delta)\,\mathbb{E}[U]\,.$$

Now suppose that $a > 0$, and that the expectation $\mathbb{E}[U]$ is at least $a/2$. Then,

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2 \geq (2 - e^{1/2} - \delta)^2\,\mathbb{E}[U]^2 \geq a^2(2 - e^{1/2} - \delta)^2/4 = \Omega(a^2(1-\delta)^2).$$

Similarly, if $a > 0$ and $\mathbb{E}[U]$ is less than $a/2$, we have $\mathbb{E}[(Y-a)^2] = \Omega(a^2(1-\delta)^2)$. The cases in which $a < 0$ are symmetric to the cases where $a > 0$, and the statement is trivially true when $a = 0$. □

[3] The statistical difference between random variables X and Y on a discrete space $\mathcal{X}$ is $\max_{S \subset \mathcal{X}} |\Pr[X \in S] - \Pr[Y \in S]|$.

This concludes the proof of Lemma 2.10. □

For a database D, define $T_1(D) = \{\Delta_1, \ldots, \Delta_n\}$ as the (multi) set of Δ's from $\{-2, 2\}^d$ such that for each Δ_i there exists a neighbor D' of D such that $\mathcal{I}_1(D') - \mathcal{I}_1(D) = \Delta_i$. In other words, consider the n neighbors $D_1, \ldots, D_n$ of D, where D_i is obtained by replacing all the -1's to 1's and all the 1's to -1's in the ith row of D, and set $\Delta_i = \mathcal{I}_1(D_i) - \mathcal{I}_1(D)$. Define $S_1(D) \subseteq T_1(D)$, $M_1(D)$, and $N_1(D)$ as

$$S_1(D) = \{\Delta \in T_1(D) \mid \mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta\rangle^2] = \Omega(d^2)\},$$
$$M_1(D) = \textstyle\sum_{\Delta \in S_1(D)} u_\Delta u_\Delta^\top \quad \text{and} \quad N_1(D) = \sum_{\Delta \in T_1(D)} u_\Delta u_\Delta^\top.$$

Let $\mathcal{D}$ be set of all databases from $(\{-1, 1\}^d)^n$. The proof considers databases D_r drawn at random from $\mathcal{D}$. The following lemma bounds the largest eigenvalue of $M_1(D_r)$. The proof uses an application of matrix-valued Chernoff bound from Ahlswede and Winter [1].

Some facts used in the proof Lemma 2.13. We let $M \geq 0$ to denote that M is positive semidefinite. This gives an ordering of matrices namely, $M_1 \leq M_2$ iff $M_2 - M_1 \geq 0$. For two matrices $M_1 \leq M_2$, we will let $[M_1, M_2]$ denote the set of all matrices M_3 such that $M_1 \leq M_3 \leq M_2$. The matrix exponential is define as:

$$\exp(M) = \sum_{i=0}^{\infty} \frac{M^i}{i!}.$$

$\exp(M)$ is diagonalizable in the same basis as M, and if λ is an eigenvalue of M, then e^λ is an eigenvalue for $\exp(M)$.

For a database D define,

$$N_1(D) = \sum_{\Delta \in T_1(D)} u_\Delta u_\Delta^\top.$$

Claim 2.12. $||M_1(D)||_\infty \leq ||N_1(D)||_\infty$.

Proof. Consider any vector $v \in \mathbb{R}^d$. Since $M_1(D)$ and $N_1(D)$ are positive semidefinite, $v^\top M_1(D) v \leq v^\top N_1(D) v$. Since, the previous inequality holds for every vector $v \in \mathbb{R}^d$, we get that $||M_1(D)||_\infty \leq ||N_1(D)||_\infty$. □

Lemma 2.13. *With probability greater than* $1 - 1/2n$ *over* D_r *chosen uniformly at random from* $\mathcal{D}$, $||M_1(D_r)||_\infty = O(\max\{n/d, \log d\})$.

Proof. To prove the lemma we will show that with high probability $||N_1(D_r)||_\infty = O(n/d)$. Then by using Claim 2.12 we get the desired result. To prove the bound on $||N_1(D_r)||_\infty$ we use the following matrix-valued Chernoff bound of Ahlswede and Winter [1].

Theorem 2.14 ([1, 40]). *Suppose* $f : [\ell] \to [-\mathbb{I}_d, \mathbb{I}_d]$ *and let* $X_1, \ldots, X_k$ *be arbitrary independent random variables distributed over* $[\ell]$. *Then, for all* $\gamma \in \mathbb{R}$ *and* $t > 0$:

$$\Pr\left[\frac{1}{k}\sum_{j=1}^{k} f(X_j) \not\leq \gamma \mathbb{I}_d\right] \leq d \exp(-t\gamma k) \prod_{j=1}^{k} ||\mathbb{E}[\exp(t f(X_j))]||_\infty.$$

Let $T_1(D_r) = \{\Delta_1, \ldots, \Delta_n\}$. Now, $u_\Delta u_\Delta^\top \in [-\mathbb{I}_d, \mathbb{I}_d]$ for $\Delta \in \{-2, 2\}^d$. Restating the above theorem:

Corollary 2.15. *Let $\Delta_j \in \{-2,2\}^d$ for $j \in [n]$. For all $\gamma \in \mathbb{R}$ and $t > 0$,*

$$\Pr_{D_r}\left[\frac{1}{n}\sum_{j=1}^{n} u_{\Delta_j}u_{\Delta_j}^\top - \frac{\mathbb{I}_d}{d} \not\leq \gamma \mathbb{I}_d\right] \leq d\exp(-t\gamma k)\prod_{j=1}^{n}\left|\left|\mathop{\mathbb{E}}_{D_r}[\exp(T_1(u_{\Delta_j}u_{\Delta_j}^\top) - \frac{\mathbb{I}_d}{d})]\right|\right|_\infty.$$

Note that $\frac{1}{n}\sum_{j=1}^{n} u_{\Delta_j}u_{\Delta_j}^\top - \frac{\mathbb{I}_d}{d} \not\leq \gamma\mathbb{I}_d \equiv ||\frac{1}{n}N_1(D_r)||_\infty \geq \frac{1}{d}+\gamma$. Also, since $u_{\Delta_1}u_{\Delta_1}^\top, \ldots, u_{\Delta_n}u_{\Delta_n}^\top$ are all independent and identically distributed we can restate the corollary in a more useful form as (where $\Delta = \Delta_1$):

$$\Pr_{D_r}\left[\left|\left|\tfrac{1}{n}N_1(D_r)\right|\right|_\infty \geq \tfrac{1}{d}+\gamma\right] \leq d\exp(-t\gamma n)\left(\left|\left|\mathbb{E}_\Delta[\exp(T_1(u_\Delta u_\Delta^\top - \tfrac{\mathbb{I}_d}{d}))]\right|\right|_\infty\right)^n. \tag{2}$$

Let $diag(c_1,\ldots,c_n)$ be an $n \times n$ diagonal matrix, with $c_1,\ldots,c_n$ as the entries in the diagonal. Consider, $||\,\mathbb{E}_\Delta[\exp(T_1(u_\Delta u_\Delta^\top - \frac{\mathbb{I}_d}{d}))]||_\infty$,

$$\begin{aligned}
\left|\left|\mathbb{E}_\Delta[\exp(T_1(u_\Delta u_\Delta^\top - \tfrac{\mathbb{I}_d}{d}))]\right|\right|_\infty &= ||\,\mathbb{E}_\Delta[\exp(t\cdot diag(1-1/d,-1/d,\ldots,-1/d))]||_\infty \\
&= ||\,\mathbb{E}_\Delta[\exp(diag(t-t/d,-t/d,\ldots,-t/d))]||_\infty \\
&= ||\,\mathbb{E}_\Delta[diag(\exp(t-t/d),\exp(-t/d),\ldots,\exp(-t/d))]||_\infty \\
&= ||\,\mathbb{E}_\Delta[(e^{t-t/d}-e^{-t/d})u_\Delta u_\Delta^\top + e^{-t/d}\mathbb{I}_d]||_\infty \\
&= \left|\left|\left(\frac{(e^{t-t/d}-e^{-t/d})}{d} + e^{-t/d}\right)\mathbb{I}_d\right|\right|_\infty = \frac{(e^{t-t/d}-e^{-t/d})}{d} + e^{-t/d}.
\end{aligned}$$

The second last equality follows because $\mathbb{E}_\Delta[u_\Delta u_\Delta^\top] = \mathbb{I}_d/d$. Setting $t = 1$, the right hand of Equation 2 simplifies to

$$\begin{aligned}
d\exp(-\gamma n)\left(\left|\left|\mathbb{E}_\Delta[\exp(u_\Delta u_\Delta^\top - \tfrac{\mathbb{I}_d}{d})]\right|\right|_\infty\right)^n &= d\exp(-\gamma n)\left(\frac{(e^{1-1/d}-e^{-1/d})}{d} + e^{-1/d}\right)^n \\
&= d\exp(-\gamma n)\exp(-n/d)\left(\frac{e-1}{d}+1\right)^n \\
&\leq d\exp(-\gamma n)\exp(-n/d)\exp((e-1)n/d) \\
&= d\exp(-n(\gamma+(2-e)/d)).
\end{aligned}$$

The last inequality uses the fact that $1 + x \leq e^x$. We consider two cases:

Case 1: $\mathbf{n \geq 8d\log d}$. Setting $\gamma = 1/d$, implies that

$$\Pr_{D_r}\left[\left|\left|\tfrac{1}{n}N_1(D_r)\right|\right|_\infty \geq \tfrac{2}{d}\right] \leq d\exp\left(\tfrac{-n(3-e)}{d}\right).$$

which simplifies to $\Pr_{D_r}[(1/n)||N_1(D_r)||_\infty \geq 2/d] \leq 1/(2n)$.

Case 2: $\mathbf{n < 8d\log d}$. Setting $\gamma = (8\log d)/n$, implies that $\Pr_{D_r}[(1/n)||N_1(D_r)||_\infty \geq (16\log d)/n] \leq 1/(2n)$. Rewriting, the above inequalities proves the desired statement. □

Proposition 2.17 follows by using the lemma and the fact that the expected size of $S_1(D_r)$ is at least $n/2$ (Claim 2.16).

Claim 2.16. $\mathbb{E}_{D_r}[|S_1(D_r)|] \geq n/2$.

Proof. Consider two neighboring databases, D and D_i such that $\mathcal{I}_1(D_i) - \mathcal{I}_1(D) = \Delta \in \{-2, 2\}^d$. From Lemma 2.10, we know that Δ is present in at least one of $S_1(D)$ or $S_1(D_i)$. Since every D has n such neighbors D_i's, the average size of $S_1(D)$ is at least $n/2$. Therefore, for a random database D_r, $\mathbb{E}_{D_r}[|S_1(D_r)|] \geq n/2$. □

Proposition 2.17 (General case: 1-way inner products). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^d$ be any $1/2$-differentially private algorithm for $\mathcal{I}_1$. Then, with probability at least $1/n$ over D_r chosen uniformly at random from $\mathcal{D}$, $tr(\Sigma(\mathcal{A}(D_r))) = \Omega(\min\{d^2, nd/\log d\})$.*

Proof. Let $H_1(D_r) = \frac{1}{|S_1(D_r)|} M_1(D_r)$. Consider $tr(\Sigma(\mathcal{A}(D_r))H_1(D_r))$,

$$\begin{aligned} tr(\Sigma(\mathcal{A}(D_r))H_1(D_r)) &= \frac{1}{|S_1(D_r)|} \sum_{\Delta \in S_1(D_r)} tr(\Sigma(\mathcal{A}(D_r))u_\Delta u_\Delta^\top) \\ &= \frac{1}{|S_1(D_r)|} \sum_{\Delta \in S_1(D_r)} tr(u_\Delta^\top \Sigma(\mathcal{A}(D_r))u_\Delta) = \Omega(d). \end{aligned}$$

The last equality follows by the definition of $S_1(D_r)$.

Also, $tr(\Sigma(\mathcal{A}(D_r))H_1(D_r)) \leq tr(\Sigma(\mathcal{A}(D_r)))||H_1(D_r)||_\infty$. From Claim 2.16, we know that $\mathbb{E}_{D_r}[|S_1(D_r)|] \geq n/2$. Let E_1 be the event that $|S_1(D_r)| < n/2$. Since, $\mathbb{E}_{D_r}[|S_1(D_r)|] \geq n/2$, the $\Pr[E_1] \leq 1 - 1/n$. Let E_2 be the event that

$$||H_1(D_r)||_\infty \geq \max\{cn/(d|S_1(D_r)|), (c\log d)/|S_1(D_r)|\}$$

for some constant c. From Lemma 2.13, we know that, with probability at least $1 - 1/(2n)$ over D_r,

$$||H_1(D_r)||_\infty \leq \max\left\{\frac{cn}{d|S_1(D_r)|}, \frac{c\log d}{|S_1(D_r)|}\right\}.$$

Since, $||H_1(D_r)||_\infty tr(\Sigma(\mathcal{A}(D_r))) = \Omega(d)$, it implies that with probability at least $1 - 1/(2n)$ over D_r, (for some constant c'),

$$tr(\Sigma(\mathcal{A}(D_r))) \geq \min\left\{\frac{d^2}{c'n}|S_1(D_r)|, \frac{d}{c'\log d}|S_1(D_r)|\right\}.$$

Since with probability at least $1 - \Pr[E_1]$, $|S_1(D_r)| \geq n/2$, we get that with probability at least $1 - \Pr[E_1] - \Pr[E_2]$, $\Sigma(\mathcal{A}(D_r)) = \Omega(d^2)$. Substituting, these probabilities implies that with probability at least $1/n$, $\Sigma(\mathcal{A}(D_r)) = \Omega(\min\{d^2, (nd)/\log d\})$. □

2.2.2 k-way Inner Products - Lower Bounds for General ϵ-differential Privacy

The analysis uses ideas from Sections 2.1.2 and Sections 2.2.1. Let $\widetilde{\mathcal{D}} \subset \mathcal{D}$ be set of all databases from $(\{-1, 1\}^d)^n$ who have at least one row of 1^d. The idea is to show that with probability at least $1/n$, either a random database from $\mathcal{D}$ or a random database from $\widetilde{\mathcal{D}}$ has big trace for its mean squared matrix.

Consider the random vector $z_r = \mathcal{I}_k(r)$ from Section 2.1.2. Define another random vector $\tilde{z}_r$ as $\tilde{z}_r = \mathcal{I}_k(1^d) - z_r = o - z_r$, where $o = 1^{m_k}$. The following lemma is a generalization of the Lemmas 2.7 and 2.10.

Lemma 2.18. *Let $\mathcal{A}$ be any $1/2$-differentially private algorithm for $\mathcal{I}_k$. Let $\mathcal{A}(D) \approx_{1/2} \mathcal{A}(D')$. Let $\mathcal{I}_k(D') = \mathcal{I}_k(D) + \tilde{z}$ for some $\tilde{z} \in \text{supp}(\tilde{z}_r)$ and $z = o - \tilde{z}$. Then, at least one of $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z\rangle^2]$ or $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_k(D'), \pi z\rangle^2]$ is $\Omega(\langle \pi z, \pi z\rangle^2)$.*

We consider random databases for the proof. Let $\widetilde{\mathcal{D}} \subset \mathcal{D}$ be set of all databases from $(\{-1, 1\}^d)^n$ who have at least one row of 1^d. For a database $D \in \mathcal{D}$, consider the n neighboring databases[4] $\widetilde{D}_1, \widetilde{D}_2, \ldots, \widetilde{D}_n$, where $\widetilde{D}_i$

[4] If D has a row of 1^d (say the ith), then $D = \widetilde{D}_i$ and $\tilde{z}_i = 0^{m_k}$. For uniformity, we will still treat D and $\widetilde{D}_i$ as neighbors.

is obtained by replacing ith row of D by 1^d. The databases $\widetilde{D}_1, \ldots, \widetilde{D}_n$ belong to $\widetilde{\mathcal{D}}$. Let $T_k(D) = \{z_1, \ldots, z_n\}$ denote the (multi) set such that $\mathcal{I}_k(\widetilde{D}_i) - \mathcal{I}_k(D) = \tilde{z}_i$ and $z_i = o - \tilde{z}_i$. Define,

$$S_k(D) = \{z \in T_k(D) \mid \mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z\rangle^2] = \Omega(m_k^2)\}$$
$$M_k(D) = \textstyle\sum_{z \in S_k(D)} u_z u_z^\top \quad \text{and} \quad N_k(D) = \sum_{z \in T_k(D)} u_z u_z^\top.$$

Every database in $\widetilde{\mathcal{D}}$ has at least one row of 1^d. For $\widetilde{D} \in \widetilde{\mathcal{D}}$, define $Neig(\widetilde{D})$ to be the set of all neighbors of $\widetilde{D}$ obtained by replacing a row of 1^d in $\widetilde{D}$ by a vector from $\{-1, 1\}^d$. Consider a set of n databases $D_1, \ldots, D_n$ drawn independently at random from $Neig(\widetilde{D})$. The databases $D_1, \ldots, D_n$ belong to $\mathcal{D}$. Let $\widetilde{T}_k(\widetilde{D}) = \{z_1, \ldots, z_n\}$ denote the (multi) set such that $\mathcal{I}_k(D_i) - \mathcal{I}_k(\widetilde{D}) = \tilde{z}_i$ and $z_i = \tilde{z}_i + o$. Define, $\widetilde{S}_k(\widetilde{D}) \subseteq \widetilde{T}_k(\widetilde{D})$, $\widetilde{M}_k(\widetilde{D})$, and $\widetilde{N}_k(\widetilde{D})$ as

$$\widetilde{S}_k(\widetilde{D}) = \{z \in \widetilde{T}_k(\widetilde{D}) \mid \mathbb{E}[\langle \mathcal{A}(\widetilde{D}) - \mathcal{I}_k(\widetilde{D}), \pi z\rangle^2] = \Omega(m_k^2)\},$$
$$\widetilde{M}_k(\widetilde{D}) = \textstyle\sum_{z \in \widetilde{S}_k(\widetilde{D})} u_z u_z^\top \quad \text{and} \quad \widetilde{N}_k(\widetilde{D}) = \sum_{z \in \widetilde{T}_k(\widetilde{D})} u_z u_z^\top.$$

The following lemma follows by modifying the parameters of Lemma 2.13.

Lemma 2.19. *With probability greater than* $1 - 1/2n$ *over* D_r *chosen uniformly at random from* $\mathcal{D}$, $||M_k(D_r)||_\infty \leq ||N_k(D_r)||_\infty = O(\max\{n/m_k, \log m_k\})$. *Also for every* $\widetilde{D} \in \widetilde{\mathcal{D}}$ *with probability (over the random choices of* $D_1, \ldots, D_n$*) greater than* $1 - 1/2n$, $||\widetilde{M}_k(\widetilde{D})||_\infty \leq ||\widetilde{N}_k(\widetilde{D})||_\infty = O(\max\{n/m_k, \log m_k\})$.

Proof. Replace the parameter d in the proof of Lemma 2.19 by m_k. The proof requires that $\mathbb{E}_{z_r}[u_{z_r} u_{z_r}^\top] = \mathbb{I}_{m_k}/m_k$ (or $\mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_k}$), which can be established as in Lemma 2.4. □

Claim 2.20. *Let* D_r *be a database chosen uniformly at random from* $\mathcal{D}$ *and* $\widetilde{D}_r$ *be a database chosen uniformly at random from* $\widetilde{\mathcal{D}}$. *Then, at least one of* $\mathbb{E}_{D_r}[|S_k(D_r)|]$ *or* $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{S}_k(\widetilde{D}_r)|]$ *is at least* $n/4$.

Proof. Firstly $\forall z \in \text{supp}(z_r)$,

$$\langle z, z\rangle = \langle \pi z, \pi z\rangle + \langle z, o\rangle^2/m_k.$$

Now consider the random vector $z_r = \mathcal{I}_k(r)$ (where $r \in \{-1, 1\}^d$ is random). With probability at least $1 - c^{-d}$, $\langle z_r, o\rangle \leq C m_k$ (where $c > 1$ and $C < 1$ are constants). Therefore, with probability at least $1 - c^{-d}$, $\langle \pi z_r, \pi z_r\rangle \geq m_k(1 - C^2)$ (as $\forall z \in \text{supp}(z_r)$, $\langle z, z\rangle = m_k$).

For a database $D \in \mathcal{D}$ and $\widetilde{D} \in \widetilde{\mathcal{D}}$ define,

$$R_k(D) = \{z \in T_k(D) \mid \mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2)\}$$
$$\widetilde{R}_k(\widetilde{D}) = \{z \in \widetilde{T}_k(\widetilde{D}) \mid \mathbb{E}[\langle \mathcal{A}(\widetilde{D}) - \mathcal{I}_k(\widetilde{D}), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2)\}.$$

Let D_r and $\widetilde{D}_r$ be random databases from $\mathcal{D}$ and $\widetilde{\mathcal{D}}$, respectively. Now every $z \in S_k(D_r)$ is an independent copy of z_r. Therefore, each $z \in S_k(D_r)$ independently satisfies $\langle \pi z, \pi z\rangle = \Omega(m_k)$ with probability at least $1 - c^{-d}$. By using this along with the definitions of $S_k(D_r)$ and $R_k(D_r)$ implies $\mathbb{E}_{D_r}[|S_k(D_r)|] \geq (1 - c^{-d})\, \mathbb{E}_{D_r}[|R_k(D_r)|]$. Similarly, each $z \in \widetilde{S}_k(\widetilde{D}_r)$ independently satisfies $\langle \pi z, \pi z\rangle = \Omega(m_k)$ with probability at least $1 - c^{-d}$, and therefore,

$$\mathop{\mathbb{E}}_{\widetilde{D}_r}[|\widetilde{S}_k(\widetilde{D}_r)|] \geq (1 - c^{-d}) \mathop{\mathbb{E}}_{\widetilde{D}_r}[|\widetilde{R}_k(\widetilde{D}_r)|].$$

We show that if $\mathbb{E}_{D_r}[|R_k(D_r)|] < n/2$, then $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{R}_k(\widetilde{D}_r)|] \geq n/2$. Let $Neig(\widetilde{D})$ to be the set of all neighbors of $\widetilde{D}$ obtained by replacing a row of 1^d in $\widetilde{D}$ by a vector from $\{-1, 1\}^d$. Let $o = 1^{m_k}$.

For any database $D' \in Neig(\widetilde{D})$, we know (from Lemma 2.18) that at least one of $\mathbb{E}[\langle \mathcal{A}(\widetilde{D}) - \mathcal{I}_k(\widetilde{D}), \pi z, \rangle^2]$ or $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_k(D'), \pi z \rangle^2]$ is $\Omega(\langle \pi z, \pi z \rangle^2)$ (where $z = \mathcal{I}_k(D') - \mathcal{I}_k(\widetilde{D}) + o$). Now if $\mathbb{E}_{D_r}[|R_k(D_r)|] < n/2$, then

$$\mathop{\mathbb{E}}_{\widetilde{D}_r}[|D' \in Neig(\widetilde{D}_r) \,:\, \mathbb{E}[\langle \pi z, \mathcal{A}(\widetilde{D}_r) - \mathcal{I}_k(\widetilde{D}_r) \rangle^2] = \Omega(\langle \pi z, \pi z \rangle^2)|] \geq \frac{|Neig(\widetilde{D}_r)|}{2}.$$

Therefore, for a database D'' chosen uniformly at random $Neig(\widetilde{D}_r)$, with probability at least $1/2$,

$$\mathbb{E}[\langle \mathcal{A}(\widetilde{D}_r) - \mathcal{I}_k(\widetilde{D}_r), \pi z \rangle^2] = \Omega(\langle \pi z, \pi z \rangle^2) \text{ where } z = \mathcal{I}_k(D'') - \mathcal{I}_k(\widetilde{D}_r) + o.$$

Therefore, if $\mathbb{E}_{D_r}[|R_k(D_r)|] < n/2$, then $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{R}_k(\widetilde{D}_r)|] \geq n/2$.

Therefore, at least one of $\mathbb{E}_{D_r}[|R_k(D_r)|]$ or $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{R}_k(\widetilde{D}_r)|]$ is at least $n/2$. Hence, at least one of $\mathbb{E}_{D_r}[|S_k(D_r)|]$ or $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{S}_k(\widetilde{D}_r)|]$ is greater than $(1 - c^{-d}) \cdot (n/2) \geq n/4$. □

Proposition 2.21 (General case: k-way inner products). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^{m_k}$ be any $1/2$-differentially private algorithm for $\mathcal{I}_k$. Let D_r be a database chosen uniformly at random from $\mathcal{D}$ and $\widetilde{D}_r$ be a database chosen uniformly at random from $\widetilde{\mathcal{D}}$. Then with probability at least $1/n$, at least one of $tr(\Sigma(\mathcal{A}(D_r)))$ or $tr(\Sigma(\mathcal{A}(\widetilde{D}_r)))$ is $\Omega(\min\{m_k^2, nm_k/(\log m_k)\})$.*

Proof. We divide the proof into two cases based on Claim 2.20. Let $o = 1^{m_k}$.
Case 1: $\mathbb{E}_{\mathbf{D_r}}[|\mathbf{S_k(D_r)}|] \geq \mathbf{n/4}$. Let $\Sigma(\mathcal{A}(D_r)) = \mathbb{E}[(\mathcal{A}(D_r) - \mathcal{I}_k(D_r))(\mathcal{A}(D_r) - \mathcal{I}_k(D_r))^\top]$. By definition of $S_k(D_r)$,

$$\sum_{z \in S_k(D_r)} (\pi z)^\top \Sigma(\mathcal{A}(D_r))(\pi z) = \sum_{z \in S_k(D_r)} \mathbb{E}[\langle \mathcal{A}(D_r) - \mathcal{I}_k(D_r), \pi z \rangle^2] = \Omega(m_k^2 |S_k(D_r)|).$$

On the other hand,

$$\begin{aligned}
\sum_{z \in S_k(D_r)} (\pi z)^\top \Sigma(\mathcal{A}(D_r))(\pi z) &= \sum_{z \in S_k(D_r)} tr(\pi^\top \Sigma(\mathcal{A}(D_r)) \pi z z^\top) = tr\left(\pi^\top \Sigma(\mathcal{A}(D_r)) \pi \sum_{z \in S_k(D_r)} z z^\top\right) \\
&= tr(\pi^\top \Sigma(\mathcal{A}(D_r)) \pi m_k M_k(D_r)) \\
&\leq tr(\pi^\top \Sigma(\mathcal{A}(D_r)) \pi) m_k ||M_k(D_r)||_\infty \\
&\leq tr(\Sigma(\mathcal{A}(D_r))) m_k ||M_k(D_r)||_\infty.
\end{aligned}$$

Let $H_k(D_r) = M_k(D_r)/|S_k(D_r)|$. Equating the upper and lower bounds on $\sum_{z \in S_k(D_r)} (\pi z)^\top \Sigma(\mathcal{A}(D_r))(\pi z)$ we get,

$$\Omega(m_k) = tr(\Sigma(\mathcal{A}(D_r))) ||H_k(D_r)||_\infty.$$

The remaining proof of this case is identical to Proposition 2.17. $||H_k(D_r)||_\infty = ||M_k(D_r)||_\infty / |S_k(D_r)|$. Lemma 2.19 can be used to bound $||M_k(D_r)||_\infty$ and by the assumption of this case, $\mathbb{E}_{D_r}[|S_k(D_r)|] \geq n/4$.
Case 2: $\mathbb{E}_{\widetilde{\mathbf{D}}_\mathbf{r}}[|\widetilde{\mathbf{S}}_\mathbf{k}(\widetilde{\mathbf{D}}_\mathbf{r})|] \geq \mathbf{n/4}$. The proof of this case goes as in Case 1. We define, $\widetilde{H}_k(\widetilde{D}_r) = \widetilde{M}_k(\widetilde{D}_r)/|\widetilde{S}_k(\widetilde{D}_r)|$, and use Lemma 2.19 to bound $||\widetilde{M}_k(\widetilde{D}_r)||_\infty$.

Since by Claim 2.20 at least one of the cases hold, we get that with probability at least $1/n$ there exists a database such that trace of its mean squared error matrix is $\Omega(\min\{m_k^2, \frac{nm_k}{\log m_k}\})$. □

2.3 Strengthening the Lower Bounds - Getting ϵ into the Bounds

Let ϵ be the privacy parameter. Let $D_v \in (\{-1,1\}^d)^{2\epsilon n}$. Let $R(D_v) \in (\{-1,1\}^d)^n$ be a database obtained by replicating each row of D_v exactly $1/(2\epsilon)$ times. The first observation is that $\mathcal{I}_k(D_v) = \mathcal{I}_k(R(D_v)) \cdot 2\epsilon$. Let $\mathcal{A}$ be a differentially private algorithm that takes as input databases of size n. Define as follows an algorithm $\mathcal{A}'$ that takes as input databases of size $2\epsilon n$.

ALGORITHM $\mathcal{A}'(D_v)$

1. Construct the database $R(D_v)$.
2. Run algorithm $\mathcal{A}$ with input $R(D_v)$, to get $\mathcal{A}(R(D_v))$.
3. Output $2\epsilon \cdot \mathcal{A}(R(D_v))$.

Claim 2.22. *If $\mathcal{A}$ is ϵ-differentially private then $\mathcal{A}'$ is $1/2$-differentially private.*

Proof. Differential privacy composes well. We state the composition claim for the more general (ϵ, δ)-differential privacy.

Claim 2.23 (Composition and Post-processing [11, 31, 29, 21])**.** *If a randomized algorithm $\mathcal{A}$ runs k algorithms $\mathcal{A}_1, \ldots, \mathcal{A}_k$, where each $\mathcal{A}_i$ is (ϵ_i, δ_i)-differentially private, and outputs a function of the results (that is, $\mathcal{A}(z) = g(\mathcal{A}_1(z), \mathcal{A}_2(z), \ldots, \mathcal{A}_k(z))$ for some probabilistic algorithm g), then $\mathcal{A}$ is $(\sum_{i=1}^k \epsilon_i, e^{\epsilon} \sum_{i=1}^k \delta_i)$-differentially private.*

Consider a database $D_v \in (\{-1,1\}^d)^{2\epsilon n}$. Consider a neighbor $D'_v \in (\{-1,1\}^d)^{2\epsilon n}$ of D_v. By composition property of differential privacy (Claim 2.23), for every output set $\mathcal{S}$

$$\Pr[\mathcal{A}(R(D_v)) \in \mathcal{S}] \leq \exp(1/2) \Pr[\mathcal{A}(R(D'_v)) \in \mathcal{S}] \Rightarrow \Pr[\mathcal{A}'(D_v) \in \mathcal{S}] \leq \exp(1/2) \Pr[\mathcal{A}'(D'_v) \in \mathcal{S}].$$

Since the above inequality holds for all neighboring databases D_v and D'_v, $\mathcal{A}'$ is $1/2$-differentially private. □

Claim 2.24. *There exists a database $D_v \in (\{-1,1\}^d)^{2\epsilon n}$ such that $tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top]) = \Omega(\min\{m_k^2, (\epsilon n m_k)/\log m_k\})$.*

Proof. Since $\mathcal{A}'$ is $1/2$-differentially private (Claim 2.22), means that we can apply Proposition 2.21 to conclude that there exists a database D_v of size $2\epsilon n$ such that $tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top]) = \Omega(\min\{m_k^2, (\epsilon n m_k)/\log m_k\})$. □

Lemma 2.25. *Let $\mathcal{A}$ be any ϵ-differentially private algorithm for $\mathcal{I}_k$. Let D_v be the database such that $tr(\Sigma(\mathcal{A}'(D_v))) = \Omega(\min\{m_k^2, (\epsilon n m_k)/\log m_k\})$. Then,*

$$tr(\Sigma(\mathcal{A}(R(D_v)))) = \Omega(\min\{m_k^2/\epsilon^2, (n m_k)/(\epsilon \log m_k)\}).$$

Proof. We equate $tr(\mathbb{E}[(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))^\top])$ in terms of $tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top])$.

$$\begin{aligned}
& tr(\mathbb{E}[(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))^\top]) \\
&= tr\left(\mathbb{E}\left[\left(\frac{\mathcal{A}'(D_v)}{2\epsilon} - \frac{\mathcal{I}_k(D_v)}{2\epsilon}\right)\left(\frac{\mathcal{A}'(D_v)}{2\epsilon} - \frac{\mathcal{I}_k(D_v)}{2\epsilon}\right)^\top\right]\right) \\
&= \frac{1}{4\epsilon^2} tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top]) \\
&= \Omega\left(\min\left\{\frac{m_k^2}{\epsilon^2}, \frac{n m_k}{\epsilon \log m_k}\right\}\right).
\end{aligned}$$

The last equality follows from Claim 2.24. □

Theorem 2.26 (General Case). *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-way inner products that for every database $D \in (\{-1,1\}^d)^n$ has an average mean squared error of $\min\{o(m_k/\epsilon^2), o(n/(\epsilon \log m_k))\}$ for $\mathcal{A}(D)$ is not ϵ-differentially private. Also, any algorithm $\mathcal{A}$ for releasing all k-way conjunctions that for every database $D \in (\{0,1\}^d)^n$ has an average mean squared error of $\min\{o(m_k/(2^k\epsilon^2)), o(n/(2^k\epsilon \log m_k))\}$ for $\mathcal{A}(D)$ is not ϵ-differentially private.*

3 Lower Bounds on Noise for Minimal Privacy

In this section, we introduce a new reconstruction attack based on analyzing the least singular value of random correlated matrices. We then use the reconstruction attack to prove lower bounds on noise under the notion of strong and attribute non-privacy. Lets first formally define these notions.

Definition 3.1 (Strongly Non-Private). *An algorithm $\mathcal{A}$ is* strongly non-private *if there exists a distribution of databases $\mathbb{D}$ over the domain $(\{0,1\}^d)^n$ under which the rows of the databases are statistically independent and there exists a set $S \subseteq [n]$ with $|S| = \Omega(\min\{n, d\})$ satisfying the following properties:*

a. For any (not-necessarily polynomial time) adversary if $D \sim \mathbb{D}$, the adversary can output any row of D indexed by the elements of S with probability at most $2/3$[5];

b. There exists a polynomial time adversary such that if $D \sim \mathbb{D}$, the adversary on input $\mathcal{A}(D)$ can output $1 - o(1)$ fraction of the rows of D indexed by the elements of S with probability at least $1 - negl(n)$.

Definition 3.2 (Attribute Non-Privacy). *An algorithm $\mathcal{A}$ is attribute non-private if there exists a polynomial time adversary, a database $D \in (\{0,1\}^{d+1})^n$, a set $S \subset [d+1]$ with $|S| = d$, and an integer $t = \Omega(\min\{n, d\})$ such that the adversary on given as input $\mathcal{A}(D)$ and the columns of D indexed by the elements of S, can reconstruct $1 - o(1)$ fraction of the first t entries of the missing column.*

3.1 Upper Bounds for (Not) Strong Non-Privacy

Proposition 3.3 (Strong Non-Privacy upper bound). *There exist an algorithm for releasing all k-way conjunctions ($\mathcal{C}_k$) that is not strongly non-private and that for every database $D \in (\{0,1\}^d)^n$ and for every query in $\mathcal{C}_k(D)$ with constant probability adds $O(\min\{\sqrt{nk \log d}, \sqrt{m_k k \log d}\})$ noise.*

Proof. We call an algorithm not satisfying Definition 3.1 as *not strongly non-private*. Call a set $S \subseteq [n]$ *good* if $|S| = \Omega(\min\{n, d\})$. Call a distribution $\mathbb{D}$ *good* if for every good set S the following is satisfied: if $D \sim \mathbb{D}$ any (not necessarily polynomial time) adversary can output any row of D indexed by the elements of S with probability at most $2/3$. An algorithm $\mathcal{A}$ is not strongly non-private if for every good distribution $\mathbb{D}$ and every set good S, no polynomial time adversary given as input $\mathcal{A}(D)$ (with $D \sim \mathbb{D}$) can reconstruct $1 - o(1)$ fraction of the rows of D indexed by the elements of S with high probability. We construct two different not strongly non-private algorithms for $\mathcal{C}_k$ which when put together will give the claimed noise bound.

Random Sampling. Let $\mathbb{D}$ be a good distribution and let S be a good set. Consider $D \sim \mathbb{D}$. Define an algorithm $\mathcal{A}_{sam}$ that does the following: (1) randomly selects $n/2$ rows from D to construct a new database D_{sam}, (2) evaluates all the k-way conjunction predicates on D_{sam}, and (3) releases the vector $2 \cdot \mathcal{C}_k(D_{sam})$. Firstly, $\mathcal{A}_{sam}$ is not strongly non-private because for all $i \in S$, any adversary can output the ith row only if: (a)

[5]The choice is $2/3$ is arbitrary. Our results also hold for larger constants.

if the ith row is in D_{sam}, or (b) with probability at most $2/3$ if the ith row is not in D_{sam}. Since S is a good set the probability that any adversary can output $1 - o(1)$ rows of S is negligible.

We now invoke Chernoff bound to argue about the noise. Consider some conjunction predicate $c_v \in \mathcal{C}_k$. Now for some constants b, b',

$$\Pr\left[|2 \cdot c_v(D_{sam}) - c_v(D)| \geq \sqrt{bn\log(2^k m_k)}\right] \leq \exp\left(-2 \cdot n \cdot \frac{b\log(2^k m_k)}{n}\right) \leq \frac{1}{b' 2^k m_k}.$$

By applying a union bound it follows that the probability that

$$\forall c_v \in \mathcal{C}_k,\ \Pr\left[|2 \cdot c_v(D_{sam}) - c_v(D)| \geq \sqrt{bn\log(2^k m_k)}\right] \leq \frac{1}{b'}.$$

Adapting Differential Privacy. We use the fact that for some reasonable values of ϵ and δ any (ϵ, δ)-differentially private algorithm is not strongly non-private.

Lemma 3.4. *Any (ϵ, δ)-differentially private algorithm ($\forall(\epsilon, \delta)$ such that $(2/3)e^{\epsilon} + \delta$ is bounded away from 1) is not strongly non-private.*

Proof. Let $\mathcal{A}_d$ be an (ϵ, δ)-differentially private algorithm satisfying the conditions of the lemma, we argue that $\mathcal{A}_d$ is also not strongly non-private. Let $\mathbb{D}$ be a good distribution and S be a good set. Consider $D \sim \mathbb{D}$. Because of the guarantees of (ϵ, δ)-differential privacy, given $\mathcal{A}_d(D)$, no adversary (even with unbounded time) can predict any row of D indexed by the element of S with probability more than $(2/3)e^{\epsilon} + \delta$. Therefore, the probability that adversary can reconstruct $1 - o(1)$ fraction of the rows of D indexed by the elements of S is negligible. □

The SuLQ mechanism of Blum *et al.* [4] adds independent noise drawn according to normal distribution with mean 0 and standard deviation $\sqrt{m_k \log(1/\delta)}/\epsilon$ to each entry in $\mathcal{C}_k(D)$. We set $\epsilon = 0.1$ and $\delta = 0.1$. By Lemma 3.4, the SuLQ mechanism for these values of ϵ and δ is not strongly non-private. A simple analysis of the c.d.f. of the normal distribution, shows that for all predicates in $\mathcal{C}_k$ with constant probability the noise added will be $O(\sqrt{m_k \log(2^k m_k)})$.

Putting Together. Define a new algorithm $\mathcal{A}$ that when $\sqrt{n} \leq \sqrt{m_k}$ outputs $\mathcal{A}_{sam}(D)$, and when $\sqrt{m_k} < \sqrt{n}$ outputs the result of the SuLQ mechanism. It immediately follows that $\mathcal{A}$ is not strongly non-private, and has the claimed noise bounds. □

3.2 1-way Conjunctions - Lower Bounds for Strong Non-Privacy

In the analysis (for simplicity) we restrict our attention to monotone conjunctions. A monotone conjunction predicate $m_v : \{0,1\}^d \to \{0,1\}$ for $v \in \{0,1\}^d$ is defined as $m_v(x) = \prod_i x_i \cdot v_i$, where value of v_i indicates whether the variable x_i is present (if $v_i = 1$) or absent (if $v_i = 0$). Let $\mathcal{M}_k$ be the subset of $\mathcal{C}_k$ restricted to monotone conjunctions. Since $\mathcal{M}_k \subset \mathcal{C}_k$, a lower bound for the monotone case automatically implies a lower bound for the (non-monotone) general case.

Reconstruction Attack. Let $s = (s_1, \ldots, s_n) \in \{0,1\}^n$ be some (secret) vector. We show that there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first $\min\{n, d/2\}$ entries of s if the privacy mechanism allows d inner-product queries and adds $\min\{o(\sqrt{n}), o(\sqrt{d})\}$ noise to every response. The analysis uses some ideas from a recent attack proposed by Dwork and Yekhanin [15].

Let $a = \min\{d/2, n\}$. Let $\Phi \in \{0,1\}^a$ be a vector with independent entries taking values 0 and 1 with probability 1/2. Let $\Phi_1, \ldots, \Phi_d \in \{0,1\}^a$ be d independent copies of Φ. Let $s|_a = (s_1, \ldots, s_a)$ be the first a entries of s. Define a matrix M of dimension $d \times a$ as follows: ith row of M is Φ_i. The attack works as follows: for every row r in M, the adversary asks inner product of r with $s|_a$, and receives noisy responses. Consider any privacy mechanism $\mathcal{A}$. Let $p = \mathcal{A}(Ms|_a)$ be the vector of noisy responses generated by $\mathcal{A}$. Now if e is the noise vector, then $p = Ms|_a + e$. Let $M = P\Gamma Q$ be the singular value decomposition of M. Define a matrix $M' = Q^\top \Gamma^{-1} P^\top$. Given p, the adversary uses M' to constructs $\hat{s} = (\hat{s}_1, \ldots, \hat{s}_a)$ as follows: $\hat{s}_i = 1$ if the ith element in $M'p \geq 1/2$, and 0 otherwise.

Proposition 3.5 (1-way reconstruction attack). *Let M be the $d \times a$ matrix as defined above. If any algorithm adds $\min\{o(\sqrt{d}), o(\sqrt{n})\}$ noise to each entry in Ms, then there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first $\min\{n, d/2\}$ entries of s, with probability at least $1 - \exp(-cd)$.*

Proof. We break the proof into two cases based on the relationship between d and n.
Case 1: $\mathbf{d \geq 2n}$. Let d be greater than $2n$[6]. In this case $s|_a = s$. Now if e is the noise vector, then $p = Ms + e$. Let $M = P\Gamma Q$ be the singular value decomposition of M. Here, Γ is a diagonal matrix of singular values of M and P and Q are orthogonal matrices. P is $d \times d$ matrix, and Q is an $n \times n$ matrix. Now Γ is a $d \times n$ diagonal matrix, let $\Gamma = \binom{G}{0}$. Here G is an $n \times n$ diagonal matrix of singular values, and $\mathbf{0}$ is $d - n \times n$ zero matrix. We use the following theorem of Rudelson and Vershynin to lower bound the least singular value of M.

Theorem 3.6 (Rudelson and Vershynin [34]). *Let R be a $d \times n$ random matrix with $d \geq n$, whose elements are independent copies of mean zero subgaussian random variable[7] with unit variance. Let $\sigma_1(R), \ldots, \sigma_n(R)$ be the singular values of R in the non-increasing order. Then, for every $\gamma > 0$, we have*

$$\Pr[\sigma_n(R) \leq \gamma(\sqrt{d} - \sqrt{n-1})] \leq (\kappa\gamma)^{d-n+1} + \exp(-\tau d),$$

where $\kappa, \tau > 0$ depend (polynomially) only on the subgaussian moment of the random variable.

Corollary 3.7. *The least singular value of matrix M is $\Omega(\sqrt{d})$ with probability at least $1 - \exp(-cd)$, where c is an absolute constant.*

Proof. Theorem 3.6 doesn't directly apply to matrix M as the entries of M are not centered. However, M is just a rank one perturbation of a random matrix whose entries are centered ($M = R + (1/2)J$, where J is the all 1's matrix and R is a random centered matrix satisfying the conditions of Theorem 3.6). The proof of Rudelson and Vershynin (Theorem 3.6) can be extended to handle matrices that are small perturbations of random centered matrices (proof of Theorem 3.10 shows how this can be done). □

Define $G^{-1} = diag(1/\sigma_1(M), \ldots, 1/\sigma_n(M))$. Define $n \times d$ matrix Γ^{-1} as $(G^{-1}|\mathbf{0}^\top)$. Now, $\Gamma^{-1}\Gamma = \mathbb{I}_n$ (identity matrix of dimension $n \times n$). Define a matrix $M' = Q^\top \Gamma^{-1} P^\top$. Given p, the adversary uses M' to reconstruct s. Since, $M'p = s + M'e$. Define $\hat{s} = (\hat{s}_1, \ldots, \hat{s}_n)$ as follows: $\hat{s}_i = 1$ if the ith element in $M'p \geq 1/2$, and 0 otherwise. We now ague that if an algorithm adds $o(\sqrt{n})$ noise to each entry in Ms, then with high probability $d_H(s, \hat{s}) = o(n)$.

Now, $M'e = Q^\top \Gamma^{-1} P^\top e$ and $||M'e|| = ||Q^\top \Gamma^{-1} P^\top e|| = ||\Gamma^{-1} P^\top e||$ (Q is an orthogonal matrix, therefore multiplication by it preserves the norm). Now since, $||P^\top e|| = ||e||$ ($P^\top$ is an orthogonal matrix) implies

$$||M'e|| \leq ||\Gamma^{-1}||_\infty ||P^\top e|| = ||\Gamma^{-1}||_\infty ||e|| = ||G^{-1}||_\infty ||e||.$$

Corollary 3.7 implies that with probability at least $1 - \exp(-cd)$, $||G^{-1}||_\infty = O(1/\sqrt{d})$.

[6] For the proof it is only necessary that d be greater than $(1+\gamma)n$ for any constant $\gamma > 0$.

[7] A random variable Z is subgaussian if there exists $b > 0$ such that $\Pr[|Z| > a] \leq 2\exp(-a^2/b^2)$ for all $a > 0$.

Let us condition on the event that the smallest singular value of M is $\Omega(\sqrt{d})$. Now,

$$||M'e|| \leq ||\Gamma^{-1}||_\infty ||e|| = O(||e||/\sqrt{d}).$$

Now if an algorithm adds $o(\sqrt{n})$ noise to each monotone conjunction query then $||e|| = o(\sqrt{nd})$, which implies that $||M'e|| = o(\sqrt{n})$. In particular it implies that $M'e$ cannot have $\Omega(n)$ coordinates with absolute value above $1/2$, therefore $d_H(s, \hat{s}) = o(n)$. Since, by Corollary 3.7 with probability at least $1 - \exp(-cd)$ the smallest singular value of M is $\Omega(\sqrt{d})$, therefore, if an algorithm adds $o(\sqrt{n})$ noise to each entry in Ms, then with probability at least $1 - \exp(-cd)$, $d_H(s, \hat{s}) = o(n)$.

Case 2: $\mathbf{d < 2n}$. Let $d \leq 2n$. We can carry out the same analysis as in the previous case (n gets substituted by $d/2$ in the analysis). So if an algorithm adds $o(\sqrt{d})$ noise to each entry in Ms, then by using the previous mentioned attack the adversary can construct $\hat{t}$ such that $d_H(t, \hat{t}) = o(|t|) = o(d)$. □

Strong Non-Privacy. We construct a database $D_s \in (\{0,1\}^d)^n$ from s and $\Phi_1, \ldots, \Phi_d$ and show that there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first a rows of D_s if given too accurate vector $\mathcal{M}_1(D_s)$. We assume that the adversary knows $\Phi_1, \ldots, \Phi_d$. The database $D_s \in (\{0,1\}^d)^n$ is constructed as follows: (i,j)th entry of D_s is s_i if the ith entry in $\Phi_j = 1$, and 0 otherwise. Using the following we define a distribution over databases to prove our strong non-privacy result (Theorem 3.9).

Lemma 3.8. *Let D_s be the database as constructed above. If any algorithm adds $\min\{o(\sqrt{d}), o(\sqrt{n})\}$ noise to each entry in $\mathcal{M}_1(D_s)$ (or $\mathcal{C}_1(D_s)$), then there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first $\min\{n, d/2\}$ rows of D_s with probability at least $1 - \exp(-cd)$.*

Proof. From Case 1 of Proposition 3.5, if any algorithm adds $o(\sqrt{n})$ noise to each entry in $\mathcal{M}_1(D_s) = Ms$, then the adversary can reconstruct $1 - o(1)$ fraction of s (and hence, $1 - o(1)$ fraction of the rows of D_s) with probability at least $1 - \exp(-cd)$. Similarly from Case 2 of Proposition 3.5, if any algorithm adds $o(\sqrt{d})$ noise to each entry in $\mathcal{M}_1(D_s) = Mt$, then the adversary can reconstruct $1 - o(1)$ fraction of the first $d/2$ entries of s (and hence, $1 - o(1)$ fraction of the first $d/2$ rows of D_s) with probability at least $1 - \exp(-cd)$. Putting these two statements together concludes the proof of Lemma 3.8. □

Theorem 3.9. *Any algorithm for releasing all 1-way conjunctions ($\mathcal{C}_1$) that for every database $D \in (\{0,1\}^d)^n$ adds $\min\{o(\sqrt{n}), o(\sqrt{d})\}$ noise to each entry in $\mathcal{C}_1(D)$ is strongly non-private.*

Proof. Let $a = \min\{n, d/2\}$. Let $S = \{1, 2, \ldots, a\}$, be a set of row positions. Let Φ be a random vector from $\{0,1\}^a$. Let $\Phi_1, \ldots, \Phi_d$ be d independent copies of Φ. We fix these vectors Φ_j's for rest of the construction. We also assume that these Φ_j's are known to the adversary. For a vector $s \in \{0,1\}^n$, let D_s be a database constructed as follows: (i,j)th entry in D_s is s_i if the ith entry in $\Phi_j = 1$, and 0 otherwise. Define a random variable (probability distribution) $\mathbb{D}$ over the set of databases as follows: draw a vector s_r uniformly at random from $\{0,1\}^n$ and output D_{s_r}.

Consider $D \sim \mathbb{D}$. Let consider some ith row where $i \in S$. Let E be the event that there exists a Φ_j such that ith entry in Φ_j is 1. Conditioned on event E, an adversary can only predict the ith row of D by guessing the ith entry in s_r. Since s_r is picked uniformly at random, this implies that conditioned on E no adversary can guess the ith row of D with probability more than $1/2$. Finally, since $\Pr[\overline{E}] = 1/2^d$, therefore, no adversary can guess the ith row of D with probability more than $1/2 + 1/2^d \leq 2/3$. Thus, $\mathbb{D}$ satisfies the first condition of Definition 3.1.

The attack described in Section 3.2 (Lemma 3.8) shows that there exists a polynomial time adversary that can reconstruct $1 - o(1)$ fraction of the rows of S when given $\mathcal{C}_1(D)$ with $\min\{o(\sqrt{n}), o(\sqrt{d})\}$ noise in each entry. Therefore, for $\mathbb{D}$, both the conditions of Definition 3.1 are satisfied. □

3.3 Constant k-way Conjunctions - Lower Bounds for Strong and Attribute Non-Privacy

The idea is similar to the 1-way case. We will assume that k is a constant throughout this section. Let $u_1 = (u_1(1), \ldots, u_1(n)), u_2 = (u_2(1), \ldots, u_2(n)), \ldots, u_k = (u_k(1), \ldots, u_k(n)) \in \mathbb{R}^n$ be k vectors. The entry-wise product of $u_1, \ldots, u_k$ is, $u_1 \odot u_2 \odot \cdots \odot u_k = (u_1(1) \cdot u_2(1) \cdot \cdots \cdot u_k(1), u_1(2) \cdot u_2(2) \cdot \cdots \cdot u_k(2), \ldots, u_1(n) \cdot u_2(n) \cdot \cdots \cdot u_k(n))$.

Reconstruction Attack. Let $m'_k = \binom{d+1}{k}$. Let $a = \min\{n, c'd^k/\log d\}$ (where c' is the constant from Theorem 3.10). Let $\Phi_1, \ldots, \Phi_d \in \{0,1\}^a$ be d independent random vectors. Let $\Phi_{d+1} = 1^a$ (which is just added for ease of analysis). Define a matrix $M^{(k)}$ of dimension $m'_k \times a$ as follows: rows of $M^{(k)}$ are the entry-wise product of every set of k vectors from $\Phi_1, \Phi_2, \ldots, \Phi_{d+1}$. The attack works in the same manner as before: for every row r in $M^{(k)}$ the adversary asks the inner product of r with $s|_a$. The crucial difference comes in the analysis of the least singular value of $M^{(k)}$, which we bound using the following theorem.

Theorem 3.10 (Least Singular Value). *Let k be a constant. Let $d^{k-1} \leq n \leq c'd^k/\log d$, where c' is a constant. Let $M^{(k)}$ be the $m'_k \times n$ matrix as defined above ($a = \min\{n, c'd^k/\log d\} = n$). Let $\sigma_1(M^{(k)}), \ldots, \sigma_n(M^{(k)})$ be the least singular values of $M^{(k)}$ in non-decreasing order. Then, there exists a constant $c_k < 1$ such that $\Pr\left[\sigma_n(M^{(k)}) \leq c_k d^{k/2}\right] \leq e^{-cd}$.*

Proof Outline for k = 2. The complete proof for $k = 2$ is presented in Section 4. The proof is a development of techniques introduced in [25, 26, 33, 34]. The extension to larger constant k follows easily. For $k = 2$ the idea is as follows. Instead of analyzing $\sigma_n(M^{(k)})$, we analyze the $\sigma_n(B)$, where the rows of the matrix B are all the entry-wise products Φ_i and Φ_j, where $\Phi_i \in \{\Phi_1, \ldots, \Phi_{d/2}\}$ and $\Phi_j \in \{\Phi_{d/2+1}, \ldots \Phi_d\}$. Note that, $\sigma_n(M^{(k)}) \geq \sigma_n(B)$. In Lemma 4.4, we show that the vectors $\Phi_{d/2+1}, \ldots \Phi_d$ are in a certain regular position with probability close to 1. Then we condition on these rows and obtain a matrix consisting of $d/2$ independent groups of rows. Analysis of the behavior of the least singular value of such matrix is the core of the argument.

One important tool is bounding the small ball probability, which is the probability that the matrix B maps a *fixed* vector in the unit sphere into a small ball in the space. Instead of obtaining a uniform lower bound for $\|Bx\|$, we decompose the sphere in numerous regions, and estimate the probability that $\|Bx\|$ is small for each part separately. The regions are defined by *compressibility* of the vectors. A vector is compressible, if its norm is concentrated on a small number of coordinates. For each part we apply the *epsilon-net argument* especially tailored for a certain degree of compressibility. Namely, the region is discretized, by using an epsilon-net for a certain epsilon. Then we obtain a uniform lower estimate on the net, using the small ball probability and the union bound. This estimate is extended to the whole region by approximation. This method requires a careful balance between the small ball probability, and the size of the net. The better the small ball probability is, the bigger epsilon-net we can consider, and so the bigger region we can cover. This balance dictates the aforementioned decomposition of the sphere. The highly compressible vectors admit a small epsilon-net, and can be treated as in [25, 26]. This is done in Lemma 4.12. The rest of the sphere is decomposed into regions defined in Lemma 4.13, where we use a careful epsilon-net argument to obtain a uniform lower bound for each region. Lemmata 4.14 and 4.15 show that the rest of the sphere can be assembled from these regions. This allows to finish the proof by using the union bound. □

Corollary 3.11. *When $n = c'd^k/\log d$, the least singular value of $M^{(k)}$ is $\Omega(d^{k/2})$ with probability at least $1 - \exp(-cd)$.*

Proposition 3.12 (k-way reconstruction attack). *If any algorithm adds $\min\{o(\sqrt{n/\log d}), o(\sqrt{d^k/\log d})\}$ noise to each entry in $M^{(k)}s$, then there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first $\min\{n, d^k/\log d\}$ entries of s, with probability at least $1 - \exp(-cd)$.*

Proof. We divide the proof into three cases. Let $k \geq 2$.
Case 1: $\mathbf{n \leq d^{k-1}}$. For $n < d^{k-1}$, the proof follows by analyzing the Cases 2 and 3 of this proof with k replaced by $k-1$. Consider all the entry-wise product of every set of $k-1$ vectors from $\Phi_1, \ldots, \Phi_d$. Since $\Phi_{d+1} = 1^a = 1^n$, $M^{(k)}s$ also contains inner-product of s with all these $(k-1)$-way entry-wise products. We get that if a private algorithm adds $o(\sqrt{n/\log d})$ noise to each entry in $M^{(k)}s$, then with probability at least $1 - \exp(-cd)$, $d_H(s, \hat{s}) = o(n)$.
Case 2: $\mathbf{d^{k-1} \leq n \leq c'd^k/\log d}$. The analysis is similar to Case 1 of Proposition 3.5. We use Theorem 3.10 to bound the least singular value of $M^{(k)}$.

Analogous to Proposition 3.5, we take the inverse M' of $M^{(k)}$ (defined using the singular value decomposition of $M^{(k)}$), and show that $||M'e|| \leq ||\Gamma^{-1}||_\infty ||e||$. Let us condition on the event that the least singular value of $M^{(k)}$ is $\Omega(d^{k/2})$. Then,

$$||M'e|| = O(||e||/d^{k/2}).$$

Now if an algorithm adds $o(\sqrt{n})$ noise to each monotone conjunction query then $||e|| = o(d^{k/2}\sqrt{n})$, therefore $||M'e|| = o(\sqrt{n})$. So again, $M'e$ cannot have $\Omega(n)$ coordinates with absolute value above $1/2$, therefore $d_H(s, \hat{s}) = o(n)$ ($\hat{s}$ is constructed in the same manner as in Proposition 3.5). Since the least singular value of $M^{(k)}$ is $\Omega(d)$ with probability at least $1 - \exp(-cd)$, therefore if a private algorithm adds $o(\sqrt{n})$ noise to each entry in $M^{(k)}s$, then with probability at least $1 - \exp(-cd)$, $d_H(s, \hat{s}) = o(n)$.
Case 3: $\mathbf{c'd^k/\log d < n}$. Let $b = c'd^k/\log d$. We can carry out the same analysis as in the previous case for this $M^{(k)}$ (n gets substituted by b in the analysis). Corollary 3.11 shows that the least singular value of $M^{(k)}$ is $\Omega(d^{k/2})$ with probability at least $1 - \exp(-cd)$. If an algorithm adds $o(\sqrt{d^k/\log d})$ noise to each entry in $M^{(k)}s$, then with probability at least $1 - \exp(-cd)$, the adversary can construct $\hat{t}$ such that $d_H(t, \hat{t}) = o(|t|) = o(|b|)$. □

Remark: Substituting $d = 2n$ in Proposition 3.5 or $d = (n \log d)^{1/k}$ in Proposition 3.12 gives new attacks for achieving blatant non-privacy. Our attack requires only $\tilde{O}(n)$ queries. The main difference is that the Fourier attack of Dwork and Yekhanin is deterministic (there is no failure probability), whereas our attack has an exponentially small failure probability.

Strong Non-Privacy. We construct a database $D_s \in (\{0,1\}^d)^n$ as in Section 3.2: (i,j)th entry of D_s is s_i if the ith entry in $\Phi_j = 1$, and 0 otherwise. The proof of the following lemma follows along the lines of Lemma 3.8.

Lemma 3.13. *If any algorithm adds* $\min\{o(\sqrt{n/\log d}), o(\sqrt{d^k/\log d})\}$ *noise to each entry in* $\mathcal{M}_k(D_s)$ *(or* $\mathcal{C}_k(D_s)$*), then there exists an adversary that can reconstruct* $1 - o(1)$ *fraction of the first* $\min\{n, d^k/\log d\}$ *rows of* D_s *with probability at least* $1 - \exp(-cd)$.

Proof. We split the analysis into three cases.
Case 1: $\mathbf{n \leq d^{k-1}}$. $\mathcal{M}_k(D_s) = M^{(k)}s$. From Case 1 of Proposition 3.12, if any algorithm adds $o(\sqrt{n/\log d})$ noise to each entry in $\mathcal{M}_k(D_s)$, then the adversary can reconstruct $1 - o(1)$ fraction of the rows of s (and hence, $1 - o(1)$ fraction of the rows of D_s) with probability at least $1 - \exp(-cd)$.
Case 2: $\mathbf{d^{k-1} \leq n \leq c'd^k/\log d}$. $\mathcal{M}_k(D_s) = M^{(k)}s$. From Case 2 of Proposition 3.12, if any algorithm adds $o(\sqrt{n})$ noise to each entry in $\mathcal{M}_k(D_s)$, then the adversary can reconstruct $1 - o(1)$ fraction of the rows of s (and hence, $1 - o(1)$ fraction of the rows of D_s) with probability at least $1 - \exp(-cd)$.
Case 3: $\mathbf{c'd^k/\log d < n}$. Let $b = c'd^k/\log d$. Since Φ_j's are now random vectors from $\{0,1\}^b$, the last $n - b$ rows of D_s are 0^d. Let $t \subset s$, be the vector containing the first b positions in s. Then, $\mathcal{M}_k(D_s) = M^{(k)}t$. From Case 3 of Proposition 3.12, if any algorithm adds $o(\sqrt{d^k/\log d})$ noise to each entry in $\mathcal{M}_k(D_s)$, then the

adversary can reconstruct $1 - o(1)$ fraction of the first b entries of s (and hence, $1 - o(1)$ fraction of the first b rows of D_s) with probability at least $1 - \exp(-cd)$.

Putting these three cases together concludes the proof of Lemma 3.13. □

The proof of the following theorem is identical to Theorem 3.9 with Lemma 3.13 playing the role of Lemma 3.8.

Theorem 3.14. *Let k be a constant. Any algorithm for releasing all k-way conjunctions $(\mathcal{C}_k)$ that for every database $D \in (\{0,1\}^d)^n$ adds $\min\{o(\sqrt{n/\log d}), o(\sqrt{d^k/\log d})\}$ noise to each entry in $\mathcal{C}_k(D)$ is strongly non-private.*

Attribute Non-Privacy. We construct a database $D_a \in (\{0,1\}^{d+1})^n$ from s and $\Phi_1, \ldots, \Phi_d$ and show that there exists an adversary that when given the first d columns of D_a can reconstruct a major fraction of the last column of D_a. We assume that the adversary knows $\Phi_1, \ldots, \Phi_d$. The database $D_a \in (\{0,1\}^{d+1})^n$ is constructed as follows: the first d columns are $\Phi_1, \ldots, \Phi_d$, and the last column is s. The proof of the following theorem relies on the Proposition 3.12.

Theorem 3.15. *Let k be a constant. Any algorithm for releasing all constant k-way conjunctions $(\mathcal{C}_k)$ that for every database $D \in (\{0,1\}^{d+1})^n$ adds $\min\{o(\sqrt{n/\log d}), o(\sqrt{d^{k-1}/\log d})\}$ noise to each entry in $\mathcal{C}_k(D)$ is attribute non-private.*

Proof. Let D_a be a database with $\Phi_1, \ldots, \Phi_d$ in its first d columns and s in the last column. Now, $\mathcal{C}_k(D_a)$ contains all the entries of $M^{(k-1)}s$. Now consider an algorithm $\mathcal{A}$ that releases $\mathcal{C}_k(D_a)$ with

$$\min\{o(\sqrt{n/\log d}), o(\sqrt{d^{k-1}/\log d})\}$$

noise to each entry. Set $t = \min\{n, d^{k-1}/\log d\}$. By Proposition 3.12, there exists an adversary that, can reconstruct $1 - o(1)$ fraction of first t entries of s (first t entries of the last column of D_a). □

4 Bounding the Least Singular Value

In this section, we provide the proof of Theorem 3.10. Throughout this section C, c, c', etc. denote absolute constants, whose value may change from line to line. The constants in the proof are not optimized. For $\gamma \in \mathbb{N}$ define the function $\log^{(\gamma)}$ by induction. For $N > 0$ set $\log^{(1)} N = \max(\log N, 1)$. If $\log^{(\gamma)}$ is defined, then

$$\log^{(\gamma+1)} N = \max(\log \log^{(\gamma)} N, 1).$$

We start off by restating Theorem 3.10 in a more general form. Fixing $\gamma = 1$, gives the previous statement of this theorem. We ignore Φ_{d+1} (as having that only leads to increase in the least singular value).

Theorem 4.1 (Theorem 3.10 Restated). *Let k be a constant. Let d, n, γ be natural numbers such that $d^{k-1} \leq n \leq \frac{c' d^k}{\log^{(\gamma)} d}$. Let $m_k = \binom{d}{k}$. Let $\Phi \in \{0,1\}^n$ be a vector with independent entries taking values 0 and 1 with probability $1/2$. Let $\Phi_1, \ldots, \Phi_d$ be d independent copies of Φ. Let $M^{(k)}$ be a matrix of dimension $m_k \times n$ whose rows are then entry-wise product of every set of k vectors from $\Phi_1, \ldots, \Phi_d$. Then there exists a constant $c_k < 1$ (where c_k depends only on k) such that,*

$$\Pr\left[\sigma_n(M^{(k)}) \leq c_k^{\gamma} d^{k/2}\right] \leq e^{-cd},$$

provided d is big enough ($d \geq d(k, \gamma)$).

We now present the full proof for the $k = 2$ case. The extension to higher k's is similar. We start off by restating Theorem 4.1 for the $k = 2$ case.

Theorem 4.2 (Theorem 4.1 for $k = 2$). *Let d, n, γ be natural numbers such that $d \leq n \leq \frac{c'd^2}{\log^{(\gamma)} d}$. Let $m_2 = \binom{d}{2}$. Let $\Phi \in \{0,1\}^n$ be a vector with independent entries taking values 0 and 1 with probability $1/2$. Let $\Phi_1, \ldots, \Phi_d$ be d independent copies of Φ. Let $M^{(2)}$ be a matrix of dimension $m_2 \times n$ whose rows are then entry-wise product of every pair of vectors from $\Phi_1, \ldots, \Phi_d$ (i.e., $\Phi_1 \odot \Phi_2, \Phi_1 \odot \Phi_3, \ldots, \Phi_{d-1} \odot \Phi_d$). Then there exists a constant $c < 1$ such that,*

$$\Pr\left[\sigma_n(M^{(2)}) \leq c^\gamma d\right] \leq e^{-cd},$$

provided d is big enough ($d \geq d(\gamma)$).

Over the next few subsections we prove statements needed for the proof of Theorem 4.2. In Section 4.4, we put together all these statements to prove the theorem. The idea is to show that

$$\Pr\left[\exists x \in S^{n-1} \text{ s.t. } ||M^{(2)}x|| \leq c^\gamma d\right] \leq \exp(-cd).$$

Notations. The Euclidean sphere centered at origin is denoted by S^{n-1}. For a vector $x \in \mathbb{R}^n$, $x(i)$ represents the ith entry of the vector. The Euclidean distance from a point p to a subset T is denoted by $dist(a, T)$. For vectors $x, y \in \mathbb{R}^n$, $y \geq x$ if each entry in y is greater than the corresponding entry in x. We will let $[x, y]$ denote the set of all vectors z such that $x \leq z \leq y$.

Consider a subset T of $\mathbb{R}^n$, and let $\alpha > 0$. A α-net of T is a subset $\mathcal{N} \subseteq T$ such that for every $x \in T$ one has $dist(x, \mathcal{N}) \leq \alpha$. Throughout this section, we would use the following well-known result about α-nets.

Proposition 4.3 (Bounding the size of a α-Net [30]). *Let T be a subset of S^{n-1} and let $\alpha > 0$. Then there exists a α-net of T of cardinality at most $(1 + 2/\alpha)^n$.*

4.1 Norm estimates

Let N be a natural number. Denote by $\mathcal{W}$ the set of all $N \times n$ matrices V satisfying

$$\|V|_J\| \leq C\left(\sqrt{N} + \sqrt{|J|} \cdot \sqrt{\log \frac{en}{|J|}}\right) \tag{3}$$

for all subsets $J \subset \{1, \ldots, n\}$. Here $V|_J$ denotes the submatrix of V with columns belonging to J.

Lemma 4.4. *Let V be an $N \times n$ random ± 1 matrix (each entry in V is 1 or -1 independently with probability $1/2$). Then,*

$$\Pr[V \notin \mathcal{W}] \leq e^{-cN}.$$

Proof. Let $x \in S^{N-1}$, and let $y \in S^{n-1} \cap \mathbb{R}^J$. Then, $\langle x, V|_J y\rangle$ is a subgaussian random variable[8] of variance 1. Hence,

$$\Pr[|\langle x, V|_J y\rangle| > t] \leq e^{-ct^2}$$

for any $t \geq 1$. Let $J \subset \{1, \ldots, n\}$, $|J| = m$. Let $\mathcal{N}$ be a $(1/2)$-net in S^{N-1}, and let $\mathcal{P}$ be a $(1/2)$-net in $S^{n-1} \cap \mathbb{R}^J$. Then

$$\|V|_J\| \leq 4 \sup_{x \in \mathcal{N}} \sup_{y \in \mathcal{P}} \langle x, V|_J y\rangle.$$

[8] A random variable Z is subgaussian if there exists $b > 0$ such that $\Pr[|Z| > a] \leq 2\exp(-a^2/b^2)$ for all $a > 0$.

The nets $\mathcal{N}$ and $\mathcal{P}$ can be chosen so that $|\mathcal{N}| \le 6^N$ and $|\mathcal{P}| \le 6^m$ (by substituting $\alpha = 1/2$ in Proposition 4.3). Let c' be another constant. Combining this with the union bound, we get

$$\Pr[\|V|_J\| \ge 4t] \le |\mathcal{N}| \cdot |\mathcal{P}| \cdot e^{-16ct^2} \le \exp\left(-16ct^2 + (m+N)\log 6\right) \le e^{-c't^2}$$

provided that $t \ge C(\sqrt{N} + \sqrt{m})$. Applying the previous inequality with $t = t_m = \sqrt{N} + \sqrt{m}\sqrt{\log(en/m)}$, and taking the union bound, we get

$$\begin{aligned}\Pr[V \notin \mathcal{W}] &\le \sum_{m=1}^{n} \sum_{J:|J|=m} \Pr[\|V|_J\| > 4t_m] \le \sum_{m=1}^{n} \left(\frac{en}{m}\right)^m e^{-ct_m^2} \\ &\le \sum_{m=1}^{n} \exp\left(-C\left(\sqrt{N} + \sqrt{m}\sqrt{\log\frac{en}{m}}\right)^2 + m\log\frac{en}{m}\right) \le e^{-cN}.\end{aligned}$$

For the second inequality we used the fact $\binom{n}{m} \le (en/m)^m$. □

Lemma 4.5. *Let $\Lambda = (\lambda_{i,j})$ and $\Lambda' = (\lambda'_{i,j})$ be two $N \times n$ matrices, whose entries are independent random variables, taking values 0 and 1 with probability $1/2$. Let $B = (b_{k,j})$ be the $N^2 \times n$ matrix with entries $b_{k,j} = \lambda_{i,j} \cdot \lambda'_{i',j}$, where $k = (i-1)N + i'$ for $i, i' \in [N]$. Let R be the $N^2 \times n$ matrix, all whose entries equal $1/4$. Then, for some constants C, c,*

$$\Pr[\|B - R\| \ge CN] \le \exp(-cN).$$

Proof. Set $\theta_{i,j} = 2\lambda_{i,j} - 1$, and $\theta'_{i,j} = 2\lambda'_{i,j} - 1$. Then, $\theta_{i,j}$ and $\theta'_{i,j}$, $i = 1, \dots, N$, $j = 1, \dots, n$ are independent ± 1 random variables. Let $i' = 1, \dots, N$. Since, $\lambda_{i,j} = (\theta_{i,j} + 1)/2$ and $\lambda'_{i,j} = (\theta'_{i,j} + 1)/2$, the matrix B can be decomposed as

$$B = \frac{1}{4}B' + \frac{1}{4}B'' + \frac{1}{4}B''' + R,$$

where the matrices B', B'', B''' have entries $b'_{k,j} = \theta_{i,j} \cdot \theta'_{i',j}$, $b''_{k,j} = \theta_{i,j}$, $b'''_{k,j} = \theta'_{i',j}$ for $k = (i-1)N + i'$, respectively. The matrix B'' consists of N copies of the $N \times n$ matrix U with entries $\theta_{i,j}$. Hence, $\|B''\| = \sqrt{N} \cdot \|U\|$. Since U is a ± 1 random matrix, a $(1/2)$-net argument similar to the one used in Lemma 4.4 yields,

$$\Pr[\|B''\| > CN] = \Pr[\|U\| > C\sqrt{N}] \le e^{-cN}.$$

A similar inequality holds for the norm of B'''.

The matrix B' can be written in a similar fashion: $B' = (U_1, \dots, U_N)^\top$, where U_i is the $N \times n$ matrix with entries $u_{i',j} = \theta_{i,j} \cdot \theta'_{i',j}$. Hence,

$$\Pr[\|B'\| > CN] \le \Pr[\exists i \text{ s.t. } \|U_i\| > C\sqrt{N}] \le \sum_{i=1}^{N} \Pr[\|U_i\| > C\sqrt{N}].$$

U_i conditioned on $\theta = (\theta_{i,j} \mid i = 1, \dots, N,\ j = 1, \dots, n)$ is a ± 1 random matrix. We obtain

$$\Pr[\|U_i\| > C\sqrt{N}] = \mathop{\mathbb{E}}_{\theta}[\Pr[\|U_i\| > C\sqrt{N} \mid \theta]] \le e^{-cN}.$$

Together with the previous inequality this implies (for some constants c, c')

$$\Pr[\|B'\| > CN] \le Ne^{-cN} \le e^{-c'N}.$$

The result follows by combining the bounds for the norms of B', B'', and B'''. □

4.2 Small Ball Probability - Bounds for the Lévy concentration function

Starting from the works of Lévy [23], Kolmogorov [22], and Esséen [16] a number of results in probability theory have been concerned with the question of how spread the sums of independent random variables are. Lévy concentration is a convenient way to quantify the spread of a random variable.

Definition 4.6. *Let $\rho > 0$. Define the Lévy concentration function if a random vector $X \in \mathbb{R}^N$ by*

$$\mathcal{L}(X, \rho) = \sup_{x \in \mathbb{R}^N} \Pr[\|X - x\| \leq \rho].$$

The Lévy concentration function measures small ball probabilities which is the likelihood that the random vector X enters a small ball in the space. We will use the following standard lemma.

Lemma 4.7. *Let $X \in \mathbb{R}^N$ be a random vector, and let X' be an independent copy of X. Then, for any $\rho > 0$*

$$\mathcal{L}(X, \rho) \leq \sqrt{\Pr[\|X - X'\| \leq 2\rho]}.$$

Proof. Let $x \in \mathbb{R}^n$. Then

$$(\Pr[\|X - x\| \leq \rho])^2 = \Pr[\|X - x\| \leq \rho \text{ and } \|X' - x\| \leq \rho] \leq \Pr[\|X - X'\| \leq 2\rho].$$

Taking the supremum over $x \in \mathbb{R}^n$ completes the proof. □

For $t \in (0, 1)$ and $x = (x(1), \ldots, x(n)) \in S^{n-1}$ define the vector $x|_t \in \mathbb{R}^n$ by

$$x|_t(j) = x(j) \cdot \chi_{[-t,t]}(x(j)),$$

where $\chi_{[-t,t]}(x(j))$ is the indicator function that is 1 if $-t \leq x(j) \leq t$, and 0 otherwise. Denote

$$I_t(x) = \operatorname{supp}(x|_t - x|_{t/2}) = \{j \in \{1, \ldots, n\} \mid t/2 < |x(j)| \leq t\}.$$

Lemma 4.8. *Let $CN \leq n \leq N^2$. Let $t > 0$, and let $x \in S^{n-1}$ be a vector satisfying $m := |I_t(x)| \geq C'$. Let $U = (u_{i,j})$ be any $N \times n$ matrix with ± 1 entries. For $j = 1, \ldots, n$ let $v_j \in \mathbb{R}^N$ be the vector with coordinates $v_j(i) = x(j) \cdot u_{i,j}$. Let $\eta_1, \ldots, \eta_n$ be independent ± 1 random variables. Assume that $U \in \mathcal{W}$. Then,*

$$\mathcal{L}\left(\sum_{j=1}^{n} \varepsilon_j v_j, c\sqrt{N} \cdot \|x|_t\|\right) \leq \exp\left(-\frac{cN}{\frac{N}{m} + \log\frac{en}{m}}\right).$$

Proof. Denote $J = \operatorname{supp}(x|_{2t})$ and $I = I_t(x)$. Let $\eta'_1, \ldots, \eta'_n$ be independent copies of $\eta_1, \ldots, \eta_n$. Let $\eta = (\eta_1, \ldots, \eta_n)$, and $\eta' = (\eta'_1, \ldots, \eta'_n)$. Conditioning on $\{\eta_j\}_{j \notin J}$ and applying Lemma 4.7, we obtain for all $\rho > 0$

$$\mathcal{L}\left(\sum_{j=1}^{n} \eta_j v_j, \rho\right) \leq \mathcal{L}\left(\sum_{j \in J} \eta_j v_j, \rho\right) \leq \sqrt{\Pr\left[\left\|\sum_{j \in J} (\eta_j - \eta'_j) v_j\right\| \leq \rho\right]}.$$

Consider a function $F : \mathbb{R}^J \to \mathbb{R}$, defined by

$$F(y) = \left\|\sum_{j \in J} y_j v_j\right\| = \|Vy\|,$$

where V is the matrix with columns $v_j, j \in J$. Then F is a convex function with the Lipschitz constant[9] $L = \|V\|$. Note that $V = U|_J \cdot H$, where H is the diagonal matrix: $H = \mathrm{diag}(x_j)_{j\in J}$. For $l \in \mathbb{N}$ denote $I_l = \{j \in J \mid 2^{-l}t < |x(j)| \le 2^{1-l}t\}$. Let $Q = \{l \in \mathbb{N} \mid I_l \neq \emptyset\}$. Then

$$L \le \left(\sum_{l\in Q} \left(\|U|_{I_l}\| \cdot \max_{j\in I_l} |x(j)| \right)^2 \right)^{1/2}.$$

Since $U \in \mathcal{W}$,

$$\begin{aligned} L &\le C \left(\sum_{l\in Q} \left(N + |I_l| \cdot \log \frac{en}{|I_l|} \right) \cdot 2^{-2l}t^2 \right)^{1/2} \\ &\le C \left(Nt^2 + \sum_{l\in Q} |I_l| \cdot \log \frac{en}{|I_l|} \cdot 2^{-2l}t^2 \right)^{1/2}. \end{aligned}$$

To bound the last expression denote $Q_1 = \{j \in Q \mid |I_l| \le |I_1|\}$. Since the function $f(x) = x\log(en/x)$ increases on the interval $(0, e^2n)$,

$$\sum_{l\in Q_1} |I_l| \cdot \log \frac{en}{|I_l|} \cdot 2^{-2l}t^2 \le |I_1| \cdot \log \frac{en}{|I_1|} \cdot \sum_{l\in Q_1} 2^{-2l}t^2 \le \|x|_t\|^2 \cdot \log \frac{en}{|I_1|}.$$

Also, for any $l \notin Q_1$, $\log \frac{en}{|I_l|} \le \log \frac{en}{|I_1|}$, so

$$\sum_{l\in Q\setminus Q_1} |I_l| \cdot \log \frac{en}{|I_l|} \cdot 2^{-2l}t^2 \le \left(\sum_{l\in Q\setminus Q_1} |I_l| \cdot 2^{-2l}t^2 \right) \cdot \log \frac{en}{|I_1|} \le \|x|_t\|^2 \cdot \log \frac{en}{|I_1|}.$$

Combining the previous inequalities, and using $t^2 \cdot |I_1| \le \|x|_t\|^2$, we obtain

$$L \le C \left(Nt^2 + \|x|_t\|^2 \cdot \log \frac{en}{|I_1|} \right)^{1/2} \le C\,\|x|_t\| \cdot \left(\frac{N}{|I_1|} + \log \frac{en}{|I_1|} \right)^{1/2}.$$

By Talagrand's measure concentration theorem for convex functions [36],

$$\Pr[|F(\eta|_J - \eta'|_J) - \mathbb{M}(F)| > s] \le 2\exp\left(-\frac{cs^2}{L^2}\right),$$

where $\mathbb{M}(F)$ is a median of F. This tail estimate implies

$$|\mathbb{M}(F) - \left(\mathbb{E}[F^2]\right)^{1/2}| \le cL.$$

Since $|I_1| = |I_t(x)| \ge C'$,

$$\sqrt{\mathbb{E}[F^2]} = \sqrt{2} \left(\sum_{j\in J} \|v_j\|^2 \right)^{1/2} \ge \sqrt{2N} \cdot \|x|_t\| \ge 4L,$$

[9] The Lipschitz constant is the smallest value K such that $|F(a) - F(b)| \le K||a - b||$ for all a, b in the domain of F.

if the constant C' is large enough. We conclude that $(3/4)\cdot\sqrt{\mathbb{E}[F^2]} \le \mathbb{M}(F) \le \sqrt{\mathbb{E}[F^2]}$. Hence,

$$\Pr\left[\left\|\sum_{j\in J}(\eta_j-\eta'_j)v_j\right\| \le \frac{1}{8}\sqrt{N}\cdot\|x|_t\|\right] \le \Pr\left[|F(\eta|_J-\eta'|_J)-\mathbb{M}(F)| \ge \frac{1}{4}\sqrt{\mathbb{E}[F^2]}\right]$$
$$\le 2\exp\left(-\frac{c\,\mathbb{E}[F^2]}{L^2}\right) \le \exp\left(-\frac{c'N}{\frac{N}{|I_1|}+\log\frac{en}{|I_1|}}\right).$$

This inequality and Equation 4 finish the proof. □

Lemma 4.8 implies the following

Corollary 4.9. *Let $t>0$. Let $x\in S^{n-1}$ be a vector satisfying $\|x|_t\| \ge n^{-2}$, and $m=|I_t(x)| \ge c\log N$. Let $\Lambda'=(\lambda'_{i,j})$ be a $\{0,1\}$ matrix of size $N\times n$. For $j=1,\dots,n$ let $w_j\in\mathbb{R}^N$ be the vector with coordinates $w_j(i)=x(j)\cdot\lambda'_{i,j}$. Set $v_{i,j}=2\lambda'_{i,j}-1$, and assume that $V=(v_{i,j})\in\mathcal{W}$. Let $\eta_1,\dots,\eta_n$ be independent ±1 random variables. Then, with the notation of the previous lemma,*

$$\mathcal{L}\left(\sum_{i=1}^n \eta_i w_j, c\sqrt{N}\cdot\|x|_t\|\right) \le \exp\left(-\frac{c'N}{\frac{N}{m}+\log\frac{en}{m}}\right).$$

Proof. Note that $w_j=\frac{1}{2}v_j+\frac{1}{2}o$, where $o=1^N$. Hence,

$$\sum_{j=1}^n \eta_j w_j = \frac{1}{2}\sum_{i=1}^n \eta_i w_j + \frac{1}{2}\left(\sum_{j=1}^n \eta_j\right)\cdot o.$$

Since $\left|\sum_{j=1}^n \eta_j\right| \le n$, and $\|o\|=\sqrt{N}$, the vector $\frac{1}{2}\left(\sum_{j=1}^n \eta_j\right)\cdot d$ belongs to an interval in $\mathbb{R}^N$ of length $n\sqrt{N}$. Covering this interval by balls of radius $c\sqrt{N}\cdot\|x|_t\|$, we obtain

$$\mathcal{L}\left(\sum_{j=1}^n \eta_j w_j, c\sqrt{N}\cdot\|x|_t\|\right) \le \frac{n\sqrt{N}}{c\sqrt{N}\cdot\|x|_t\|}\cdot\mathcal{L}\left(\sum_{j=1}^n \eta_j v_j, c\sqrt{N}\cdot\|x|_t\|\right).$$

The result follows from Lemma 4.8, since $\|x|_t\| \ge n^{-2}$, and $m=|I_t(x)| \ge c\log N$. □

For the next result we need the following simple lemma.

Lemma 4.10. *Let $a_1,\dots,a_N$ be independent non-negative random variables such that $\Pr[a_i\le K]\le p$ for all $i\in[N]$. Then,*

$$\Pr\left[\sum_{i=1}^N a_i^2 \le \frac{1}{2}K^2N\right] \le (4p)^{N/2}.$$

Proof. If $\sum_{i=1}^N a_i^2 \le \frac{1}{2}K^2N$, then $a_i\le K$ for at least $N/2$ numbers i. □

Proposition 4.11 (Small Ball Probability). *Let $\Lambda=(\lambda_{i,j})$ and $\Lambda'=(\lambda'_{i,j})$ be two $N\times n$ matrices, whose entries are independent random variables, taking values 0 and 1 with probability $1/2$. Let $B=(b_{k,j})$ be the $N^2\times n$ matrix with entries $b_{k,j}=\lambda_{i,j}\cdot\lambda'_{i',j}$, where $k=i(N-1)+i'$ (obtained from Lemma 4.5). Let V be the $N\times n$ matrix with entries $v_{i,j}=2\lambda'_{i,j}-1$.*

Let $t > 0$, and let $x \in S^{n-1}$ be a vector satisfying $\||x|_t\| \geq n^{-2}$, and $m = |I_t(x)| \geq c \log N$. Then

$$\mathcal{L}\left(Bx, cN \cdot \||x|_t\| \text{ and } V \in \mathcal{W}\right) \triangleq \sup_{y \in \mathbb{R}^{N^2}} \Pr[(\|Bx - y\| \leq cN \cdot \||x|_t\|) \text{ and } (V \in \mathcal{W})] \leq \exp\left(-\frac{cN^2}{\frac{N}{m} + \log\frac{en}{m}}\right).$$

Proof. Let X and Y be random variables. Then, for any measurable sets $\mathcal{E}, \mathcal{F}$

$$\begin{aligned}\Pr[X \in \mathcal{E} \text{ and } Y \in \mathcal{F}] &= \mathop{\mathbb{E}}_{Y}\Big(\Pr[X \in \mathcal{E} \mid Y \in \mathcal{F}] \cdot \chi_{\mathcal{F}}(Y)\Big) \\ &\leq \sup_{Y \in \mathcal{F}} \Pr[X \in \mathcal{E} \mid Y \in \mathcal{F}],\end{aligned}$$

where $\chi_{\mathcal{F}}(Y)$ is the indicator random variable which is 1 if $Y \in \mathcal{F}$, and 0 otherwise. Hence,

$$\mathcal{L}\left(Bx, cN \cdot \||x|_t\| \text{ and } V \in \mathcal{W}\right) \leq \sup_{V \in \mathcal{W}} \mathcal{L}_\Lambda\left(Bx, cN \cdot \||x|_t\| \mid V\right).$$

Here $\mathcal{L}_\Lambda$ is the Lévy concentration function with respect to the random entries of Λ, while the matrix Λ' (and so $V \in \mathcal{W}$) is fixed. For $j = 1, \ldots, n$ denote by w_j the vector in $\mathbb{R}^N$ with coordinates $w_j(i) = x(j) \cdot \lambda'_{i,j}$. Decompose $\mathbb{R}^{N^2} = \bigoplus_{i=1}^N E_i$, where $E_i = \mathrm{span}(e_{(i-1)N+1}, \ldots, e_{iN})$ (where $\bigoplus$ represents direct sum and e_a is the vector with a 1 in position a and 0's everywhere else). Then, for any $i = 1, \ldots, N$, the projection of the vector $Bx \in \mathbb{R}^{N^2}$ on the subspace E_i is distributed like

$$X_i := \sum_{j=1}^n \lambda_{i,j} w_j.$$

Fix i, and denote $\eta_j = 2\lambda_{i,j} - 1$. Then,

$$X_i = \frac{1}{2}\sum_{j=1}^n \eta_j w_j + \frac{1}{2}\sum_{j=1}^n w_j,$$

where the last term doesn't depend on the random variables $\eta_1, \ldots, \eta_n$. Hence, by Corollary 4.9,

$$\mathcal{L}(X_i, 2c\sqrt{N} \cdot \||x|_t\|) = \mathcal{L}(\sum_{j=1}^n \varepsilon_j w_j, c\sqrt{N} \cdot \||x|_t\|) \leq \exp\left(-\frac{c'N}{\frac{N}{m} + \log\frac{en}{m}}\right).$$

By definition,

$$\mathcal{L}_\Lambda\left(Bx, cN \cdot \||x|_t\| \mid V\right) \leq \sup_{y_1, \ldots, y_N \in \mathbb{R}^N} \Pr\left[\sum_{i=1}^N \|X_i - y_i\|^2 \leq c^2 N \cdot \||x|_t\|^2 \cdot N \mid V\right]$$

Since the random variables $\|X_i - y_i\|$ are independent, the proposition follows from Lemma 4.10. □

4.3 Decomposition of the sphere

To prove the main result, we will decompose the sphere into several regions, and treat them by applying different modifications of the epsilon-net argument. The regions will be defined by the degree of *compressibility* of the vectors. We say that a vector is compressible, if it can be approximated by another vector having a relatively small support. The idea of classifying vectors according to their compressibility comes from [33, 34]. But unlike [33, 34], where the vectors were divided simply into compressible and incompressible, we use here a more elaborate scheme. In the following subsection we investigate vectors with different levels of compressibility.

4.3.1 Highly compressible vectors

Lemma 4.12. *Let S_0 be the set of all points $x \in S^{n-1}$, which can be decomposed as $x = y + z$, where*

$$|\mathrm{supp}(y)| \le c_0 \frac{N}{\log n/N}, \qquad \|z\| \le \rho_0$$

for some appropriately chosen constants c_0, ρ_0. Let Λ, Λ', B be as in Lemma 4.5. Then

$$\Pr[\exists x \in S_0 \text{ s.t. } \|Bx\| \le cN] \le \exp(-c'N).$$

Proof. Denote $U = \Lambda - 2R$. Then $2U$ is an $N \times n$ random ± 1 matrix with independent entries. Then, for any $x \in S^{n-1}$, $2Ux \in \mathbb{R}^N$ is a vector with independent coordinates of variance 1. Hence,

$$\mathcal{L}(Ux, c_1\sqrt{N}) \le e^{-cN} \tag{4}$$

Also,

$$\Pr[\|U\| \ge C_1\sqrt{N}] \le e^{-cn}.$$

Let $P : \mathbb{R}^N \to \mathbb{R}^N$ be the orthogonal projection, whose kernel is spanned by the vector $o = 1^N$. Then the previous inequality implies

$$\Pr[\|P\Lambda\| \ge C_1\sqrt{N}] = \Pr[\|PU\| \ge C_1\sqrt{N}] \le e^{-cn}.$$

Let $Q : \mathbb{R}^{N^2} \to \mathbb{R}^{N^2}$ be the block-diagonal matrix $Q = \mathrm{diag}(P, \dots, P)$. Then $\|QB\| \le \sqrt{N} \cdot \|P\Lambda\|$, so

$$\Pr[\|QB\| \ge C_1 N] \le e^{-cn}. \tag{5}$$

Let $H = \sqrt{n} \cdot [-o, o]$, and let $y_1, \dots, y_\ell \in H$ be a $(c_1/2)\sqrt{N}$-net in H. Here

$$\ell \le \frac{1}{\sqrt{n}(c_1/2)\sqrt{N}}.$$

By (4), for any $x \in S^{n-1}$

$$\begin{aligned}\Pr[\|P\Lambda x\| \le (c_1/2)\sqrt{N}] &\le \Pr[\mathrm{dist}(Ux, H) \le (c_1/2)\sqrt{N}] \\ &\le \Pr[\exists i \le \ell \text{ s.t. } \mathrm{dist}(Ux, y_i) \le c_1\sqrt{N}] \\ &\le \ell \cdot \mathcal{L}(Ux, c_1\sqrt{N}) \le e^{-c'N}.\end{aligned}$$

For any $i \in \{1, \dots, N\}$ let x_i be the vector with coordinates $x_i(j) = \lambda'_{i,j} \cdot x(j)$. Denote by $\Pr_\Lambda$ and $\Pr_{\Lambda'}$ the probability with respect to the entries of Λ and Λ', respectively. By the Paley–Zygmund inequality[10] $\Pr_{\Lambda'}[\|x_i\| \le \kappa] \le \mu$ for some absolute constants $\kappa, \mu < 1$. Let $1/2 < \tau < 1$ and let Ω be the event that $\|x_i\| \le \kappa$ for at least τN indices i. Then

$$\begin{aligned}\Pr[\Omega] &\le \sum_{k=\lceil \tau N \rceil}^{N} \binom{N}{k} \cdot \mu^{\tau N} \\ &\le N \exp\left(N \cdot \left((1-\tau) \cdot \log \frac{e}{1-\tau} - \tau \cdot \log \frac{1}{\mu}\right)\right) \le e^{-cN},\end{aligned} \tag{6}$$

[10] Paley–Zygmund inequality states for a random variable Z that $\Pr[|Z| \ge a] \ge (\mathbb{E}[Z^2] - a)^2 / \mathbb{E}[Z^4]$.

if the constant τ is appropriately chosen.

Let C_0 be a constant to be chosen later. Let $\mathcal{S}_{\overline{\Omega}}$ be the set of all matrices Λ' that don't satisfy Ω. We get

$$\Pr[\|QBx\| \le C_0 N] \le \sup_{\Lambda' \in \mathcal{S}_{\overline{\Omega}}} \Pr_{\Lambda}[\|QBx\| \le C_0 N \mid \Lambda'] + \Pr[\Omega]. \tag{7}$$

The vector $Bx \in \mathbb{R}^{N^2}$ consists of N blocks of the form Λx_i. Hence,

$$\|QBx\|^2 = \sum_{i=1}^{N} \|P\Lambda x_i\|^2 .$$

For $\Lambda' \in \mathcal{S}_{\overline{\Omega}}$, there exists a set $I \subset \{1, \dots, N\}$ such that $|I| \ge (1-\tau)N$ and $\|x_i\| \ge \kappa$ for all $i \in I$. Assuming that $\|P\Lambda x_i\| \ge (c_1/2)\sqrt{N} \cdot \|x_i\|$ for all $i \in I$, we get

$$\|QBx\| \ge (c_1/2)\sqrt{N} \cdot \kappa \cdot \sqrt{(1-\tau)N} =: C_0 N.$$

Therefore, for $\Lambda' \in \mathcal{S}_{\overline{\Omega}}$,

$$\begin{aligned}\Pr_{\Lambda}[\|QBx\| \le C_0 N \mid \Lambda'] &\le \Pr_{\Lambda}[\exists i \in |I| \text{ s.t. } \|P\Lambda x_i\| \le (c_1/2)\sqrt{N} \cdot \|x_i\| \mid \Lambda'] \\ &\le |I| \cdot e^{-cN} \le e^{-c'N}.\end{aligned}$$

Combining this with (6) and (7), we get that for any $x \in S^{n-1}$

$$\Pr[\|QBx\| \le C_0 N] \le e^{-cN}.$$

The proof now finishes by another application of the epsilon-net argument. By the volumetric estimate (cf. Proposition 4.3), there exists a ρ_0-net $\mathcal{N}$ in the set $\{y \in S^{n-1} \mid |\mathrm{supp}(y)| \le m\}$ of cardinality

$$|\mathcal{N}| \le \binom{n}{m} \cdot \left(\frac{3}{\rho_0}\right)^m \le \exp\left(\log\left(\frac{3en}{\rho_0 m}\right) \cdot m\right).$$

The last inequality follows by using $\binom{n}{m} \le (en/m)^m$. Then,

$$\Pr[\exists x \in \mathcal{N} \text{ s.t. } \|QBx\| \le C_0 N] \le |\mathcal{N}| \cdot e^{-c'N} \le e^{-c''N}$$

if $m \le c_0 N/\log(n/N)$ for some constant c_0. Assume that for any $x \in \mathcal{N}$ $\|QBx\| \ge C_0 N$. Let $x' \in S_0$, and choose $x \in \mathcal{N}$ such that $\|x' - x\| < 2\rho_0$. Then the inequality (5) implies

$$\begin{aligned}\|Bx'\| \ge \|QBx'\| &\ge \|QBx\| - \|QB\| \cdot \|x' - x\| \\ &\ge C_0 N - C_1 N \cdot 2\rho_0 \ge C_0/2 \cdot N\end{aligned}$$

for an appropriately chosen ρ_0. □

4.3.2 Remaining Part of the Unit Sphere

We start with deriving a uniform lower estimate of $\|Bx\|$ over a certain part of the sphere. To this end we combine the bound of Proposition 4.11 with the epsilon-net argument.

Lemma 4.13. *Let $l, m \in \mathbb{N}$ be such that $l, m \le n$, and let $b, r, t \in (0,1)$. Consider the set $S(l,b,m,t)$ of all points $x \in S^{n-1}$, which satisfy $\|x|_t\| \ge b$, $|I_t(x)| \ge m$, and can be decomposed as $x = u + v$, where*

$$|\mathrm{supp}(u)| \le l, \qquad \|v\| \le c_1 b.$$

Assume that

$$m \ge C \log N; \tag{8}$$

$$l \log \frac{cn}{lb} \le \frac{c_0 N^2}{\frac{N}{m} + \log \frac{en}{m}}. \tag{9}$$

Let B and R be the matrices defined in Lemma 4.5. Then,

$$\Pr[\exists \hat{x} \in S(l,b,m,t) \text{ s.t. } \|B\hat{x}\| \le cNb, \ \|B - R\| \le CN, \text{ and } V \in \mathcal{W}] \le \exp\left(-\frac{cN^2}{\frac{N}{m} + \log \frac{en}{m}}\right).$$

Proof. By the standard volumetric estimate (cf. Proposition 4.3), there exists a $c_1 b$-net for the set $\{u \in S^{n-1} \mid |\mathrm{supp}(u)| \le l\}$ of cardinality less than

$$\binom{n}{l} \cdot \left(\frac{6}{c_1 b}\right)^l \le \exp\left(l \log \frac{cn}{lb}\right).$$

Any such net is a $(2c_1 b)$-net for $S(l,b,m,t)$. Hence, there exists a $(4c_1 b)$-net $\mathcal{N} \subset S(l,b,m,t)$ satisfying

$$|\mathcal{N}| \le \exp\left(l \log \frac{cn}{lb}\right).$$

Denote $A = [-h, h]$, where $h = (\sqrt{n}, \ldots, \sqrt{n}) \in \mathbb{R}^n$. Denote

$$\tau = (c/2)Nb, \tag{10}$$

and let $y_1, \ldots, y_T \in A$ be a τ-net in A. Here,

$$T \le \frac{2n}{\tau}.$$

Let $x \in \mathcal{N}$. By Proposition 4.11,

$$\begin{aligned}\Pr[\mathrm{dist}(Bx, A) \le \tau \text{ and } V \in \mathcal{W}] \le \Pr[\exists j \le T \text{ s.t. } \|Bx - y_j\| \le 2\tau] &\le T \cdot \exp\left(-\frac{cN^2}{\frac{N}{m} + \log \frac{en}{m}}\right) \\ &\le \exp\left(-\frac{c'N^2}{\frac{N}{m} + \log \frac{en}{m}}\right).\end{aligned}$$

Thus (9) implies that

$$\begin{aligned}\Pr[\exists x \in \mathcal{N} \text{ s.t. } \mathrm{dist}(Bx, A) \le \tau \text{ and } V \in \mathcal{W}] \le |\mathcal{N}| \cdot \exp\left(-\frac{c'N^2}{\frac{N}{m} + \log \frac{en}{m}}\right) &\le \exp\left(l \log \frac{cn}{lr} - \frac{c'N^2}{\frac{N}{m} + \log \frac{en}{m}}\right) \\ &\le \exp\left(-\frac{(c'/2)N^2}{\frac{N}{m} + \log \frac{en}{m}}\right), \end{aligned} \tag{11}$$

if we choose $c_0 = c'/2$. Assume now that the following events hold.

- $\exists \hat{x} \in S(l,b,m,t)$ such that $\|B\hat{x}\| < \tau/2$;
- $V \in \mathcal{W}$;
- $\|B - R\| \le CN$.

Let $x \in \mathcal{N}$ be such that $\|x - \hat{x}\| \le 4c_1 b$. Then

$$\begin{aligned}\operatorname{dist}(Bx, A) &= \operatorname{dist}((B-R)x, A) \\ &\le \operatorname{dist}((B-R)\hat{x}, A) + \|(B-R)(\hat{x}-x)\| \\ &\le \operatorname{dist}(B\hat{x}, A) + \|B-R\| \cdot \|\hat{x}-x\| \\ &< \tau/2 + CN \cdot 4c_1 b \le \tau,\end{aligned}$$

where the last inequality follows from (10) and an appropriate choice of the constant c_1. Therefore,

$$\begin{aligned}&\Pr[\exists \hat{x} \in S(l,b,m,t) \text{ s.t. } \|B\hat{x}\| \le cNb, \ \ \|B-R\| \le CN, \text{ and } V \in \mathcal{W}] \\ &\le \Pr[\exists x \in \mathcal{N} \text{ s.t. } \operatorname{dist}(Bx, A) \le \tau \text{ and } V \in \mathcal{W}] \\ &\le \exp\left(-\frac{(c'/2)N^2}{\frac{N}{m} + \log\frac{en}{m}}\right).\end{aligned}$$

The last step follows from Equation 11. This completes the proof of the lemma. □

Lemma 4.13 allows to prove the uniform bound for bigger sets of vectors.

Lemma 4.14. *Let $b, t < 1$ be real numbers satisfying*

$$C \log N \le \frac{b}{t} \le n^{1/2}.$$

Let $V(l,b,t)$ be the set of all points $x \in S^{n-1}$, such that $\|x|_t\| \ge b$, and which can be decomposed as $x = y + z$, where

$$|\operatorname{supp}(y)| \le l, \quad \|z\| \le (c_1/2)b$$

and l satisfies

$$l \log\frac{cn}{lb} \le \frac{c_0 N^2}{(\frac{N}{n}\vartheta + 2)\log\vartheta}, \quad \text{where } \vartheta = \frac{nt^2}{b^2}. \tag{12}$$

Let Λ, Λ', B be as in Lemma 4.5. Then

$$\Pr\left[\exists x \in V(l,b,t) \text{ s.t. } \|Bx\| \le cNb, \ \ \|B-R\| \le CN, \text{ and } V \in \mathcal{W}\right] \le \exp\left(-cN\right).$$

Proof. Let $x = y + z$ be a decomposition as above. Set $\tau = (c_1/2)bn^{-1/2}$ (where c_1 is defined in 4.13). Decompose a vector y according to the sizes of its coordinates: $y = u + v + w$, where $w = y|_\tau$, $v = y|_t - w$, and $u = y - y|_t$. Then $\|w\| \le \tau\sqrt{n} \le c_1/2b$. Thus, $x = (u+v) + (w+z)$, where $|\operatorname{supp}(u+v)| \le l$, and $\|w+z\| \le c_1 b$. Furthermore, decompose the coordinates of v in dyadic blocks:

$$v = \sum_{r=0}^{r_1} v_r, \quad \text{where } v_r = y|_{2^{-r}t} - y|_{2^{-r-1}t}.$$

Here r_1 is the smallest number such that $2^{-r_1}t \le \tau$, so

$$r_1 \le c \log\left(\frac{nt^2}{b^2}\right).$$

Since

$$b^2/2 \le \|y|_t - y|_\tau\|^2 = \sum_{r=0}^{r_1} \|v_r\|^2,$$

it can be easily shown that there exists $r \le r_1$ such that $\|y|_{2^{-r}t}\| \ge b/2$, and $\|v_r\|^2 \ge b^2/4r_1$. Indeed, let $r_0 < r_1$ be the biggest number such that $\|y|_{2^{-r_0}t}\| \ge b/2$. Assume that $\|v_r\|^2 < b^2/4r_1$ for all $r \le r_0$. Then $\sum_{r=0}^{r_0} \|v_r\|^2 < b^2/4$, and so

$$\|y|_{2^{-r_0-1}t}\|^2 \ge \sum_{r=r_0+1}^{r_1} \|v_r\|^2 \ge b^2/4,$$

which contradicts the maximality of r_0.

Therefore, this r satisfies

$$|\mathrm{supp}(v_r)| \ge \frac{\|v_r\|^2}{(2^{-r}t)^2} \ge \frac{b^2}{4r_1t^2} \ge \frac{Cb^2}{t^2 \cdot \log\left(\frac{nt^2}{b^2}\right)} =: m.$$

For this definition of m,

$$\log\frac{en}{m} \le 2\log\vartheta,$$

and so the inequality (12) implies condition (9). Since $b/t \ge C\log N$, the condition (8) is also satisfied. In the notation of Lemma 4.13 this means that $x \in S(l, b/2, m, 2^{-r}t)$. Thus, we have shown that

$$V(l,b,t) \subset \bigcup_{r=0}^{r_1} S(l, b/2, m, 2^{-r}t).$$

Thus by Lemma 4.13,

$$\Pr\left[\exists x \in V(l,b,t) \text{ s.t. } \|Bx\| \le cNR \ \ \|B-R\| \le CN, \text{ and } V \in \mathcal{W}\right] \le r_1 \cdot \exp\left(-\frac{cN^2}{\frac{N}{m} + \log\frac{en}{m}}\right)$$
$$\le \exp\left(-\frac{c_0N^2}{\left(\frac{N\vartheta}{n}+2\right)\log\vartheta}\right).$$

Since by the assumptions of the Lemma, $\vartheta \le cn/\log^2 N$ the last quantity does not exceed e^{-cN}. □

We now provide a uniform estimate for the $\|Bx\|$ over $x \in S^{n-1} \setminus S_0$.

Lemma 4.15. *Assume that*

$$n \le \frac{cN^2}{(\log^{(\gamma)} N)^2}.$$

Let Λ, Λ', B be as in Lemma 4.5. If $N \ge N(\gamma)$, then

$$\Pr\left[\exists x \in S^{n-1} \setminus S_0 \text{ s.t. } \|Bx\| \le c^\gamma N, \ \ \|B-R\| \le CN, \text{ and } V \in \mathcal{W}\right] \le \exp\left(-cN\right).$$

Proof. We define the numbers $l_0 < l_1 < \ldots < l_\gamma$ by induction. Set

$$l_0 = \frac{c_0 N}{\log n/N}.$$

For $1 \le j \le \gamma$ set $b_j = \rho_0 \cdot (c_1/2)^{j-1}$, and $t_j = \sqrt{l_{j-1}}$. Remember, that ρ_0 is a constant from Lemma 4.12, and c_1 is a constant from Lemma 4.13.

Also, define sets S_j by

$$S_j := (S^{n-1} \setminus S_0) \setminus \bigcup_{i=0}^{j-1} V(l_i, b_i, t_i).$$

We claim that any $x \in S_j$ satisfies $\|x|_{t_j}\| \ge b_j$. Indeed, assume that $\|x|_{t_j}\| < b_j$. Since $x \notin S_0$, $\|x|_{t_1}\| \ge b_1 = \rho_0$. Hence, there exists $i \le j$ such that $\|x|_{t_{i-1}}\| \ge b_{i-1}$, while $\|x|_{t_i}\| \le b_i = (c_1/2)b_{i-1}$. Any $x \in S^{n-1}$ satisfies $|\mathrm{supp}(x - x|_{t_i})| \le t_i^{-2} = l_{i-1}$, so $x \in V(l_{i-1}, b_{i-1}, t_{i-1})$.

The numbers l_j will be chosen so that

$$\Pr\left[\exists x \in S_j \cap V(l_j, b_j, t_j) \text{ s.t. } \|Bx\| \le cNb_j \ \ \|B - R\| \le CN, \text{ and } V \in \mathcal{W}\right] \le \exp\left(-cN\right)$$

for any $j \in \{1, \ldots, \gamma\}$. Set

$$l_1 = \frac{cN^2}{\log^3 N}.$$

then $l = l_1$ satisfies condition (12) with $t = t_1$ and $b = b_1$. By Lemma 4.14, the inequality (13) is satisfied.

Let $j > 1$, and assume that l_{j-1} is already constructed. Set $\vartheta_j = \frac{nt_j^2}{b_j^2}$. Then for any $1 \le j \le \gamma$

$$\frac{N\vartheta_j}{n} \le \frac{Nt_1^2}{b_j^2} \le \frac{C\log^3 N}{N\rho_0(c_1/2)^{j-1}} \le 1,$$

provided that $N > N(\gamma)$. If

$$l_j = \frac{cN^2}{3\log^2 \vartheta_j}$$

then

$$\log \frac{cn}{l_j b_j} \le \log \frac{c\log^2 \vartheta_j}{\rho_0(c_1/2)^{j-1}} \le \log \vartheta_j,$$

whenever $N > N(\gamma)$ for some appropriately chosen $N(\gamma)$. Thus, condition (12) is satisfied. Again, in this case Lemma 4.14 implies (13).

Note that by construction $l_j \ge \frac{cN^2}{\log^{(j-1)} N}$ for any $j > 1$. Thus the claim of the lemma follows by summing up inequalities (13) over $j = 1, \ldots, \gamma$. □

4.4 Proof of the Theorem 4.1

Assume that

$$n \le \frac{cN^2}{(\log^{(\gamma)} N)^2}.$$

Then

$$n \le \frac{cN^2}{\log^{(\gamma+1)} N}.$$

Set $N = \lfloor d/2 \rfloor$, and let Λ and Λ' be matrices whose rows are the vectors $\Phi_1, \ldots, \Phi_N$ and $\Phi_{N+1}, \ldots, \Phi_d$, respectively. Form the matrix B, as in Lemma 4.5. Then $\sigma_n(M^{(2)}) \geq \sigma_n(B)$ (as B is constructed out of a subset of rows of $M^{(2)}$). We have,

$$\Pr\left[\sigma_n(M^{(2)}) \leq c^{\gamma+1}N\right] \leq \Pr\left[\exists x \in S^{n-1} \text{ s.t } ||Bx|| \leq c^{\gamma+1}N\right]$$
$$\leq \Pr\left[\exists x \in S^{n-1} \ \|Bx\| \leq c^{\gamma+1}N, \ \|B - R\| \leq CN, \text{ and } V \in \mathcal{W}\right] + \Pr\left[\|B - R\| > CN\right] + \Pr\left[V \notin \mathcal{W}\right].$$

The last inequality uses the fact that for any events E_1, E_2, E_3, $\Pr[E_1] \leq \Pr[E_1 \text{ and } E_2 \text{ and } E_3] + \Pr[\overline{E_2}] + \Pr[\overline{E_3}]$. The proof finishes by combining Lemmas 4.4, 4.5, 4.12, and 4.15. These lemmas bound each of the probability term on the right hand side to $\exp(-cd)$ (for some absolute constant c). □

Acknowledgments

The authors would like to thank Jon Ullman for many discussions during initial stages of this work.

References

[1] AHLSWEDE, R., AND WINTER, A. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory 48*, 3 (2002), 569–579.

[2] BARAK, B., CHAUDHURI, K., DWORK, C., KALE, S., MCSHERRY, F., AND TALWAR, K. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *PODS* (2007), ACM, pp. 273–282.

[3] BISHOP, Y., FIENBERG, S., AND HOLLAND, P. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge MA, 1975.

[4] BLUM, A., DWORK, C., MCSHERRY, F., AND NISSIM, K. Practical privacy: The SuLQ framework. In *PODS* (2005), ACM, pp. 128–138.

[5] BLUM, A., LIGETT, K., AND ROTH, A. A learning theory approach to non-interactive database privacy. In *STOC* (2008), ACM, pp. 609–618.

[6] CANDES, E., RUDELSON, M., TAO, T., AND VERSHYNIN., R. Error correction via linear programming. In *ANNUAL SYMPOSIUM ON FOUNDATIONS OF COMPUTER SCIENCE* (2005), vol. 46, IEEE COMPUTER SOCIETY PRESS, p. 295.

[7] CANDES, E., AND TAO., T. Decoding by linear programming. *IEEETransactions on Information Theory 51*, 12 (2005), 4203–4215.

[8] CHAUDHURI, K., AND MONTELEONI, C. Privacy-preserving logistic regression. In *NIPS* (2008), D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds., MIT Press.

[9] DINUR, I., DWORK, C., AND NISSIM, K. Revealing information while preserving privacy, full version of [10], in preparation, 2009.

[10] DINUR, I., AND NISSIM, K. Revealing information while preserving privacy. In *PODS* (2003), ACM, pp. 202–210.

[11] DWORK, C., KENTHAPADI, K., MCSHERRY, F., MIRONOV, I., AND NAOR, M. Our data, ourselves: Privacy via distributed noise generation. In *EUROCRYPT* (2006), LNCS, Springer, pp. 486–503.

[12] DWORK, C., AND LEI, J. Differential privacy and robust statistics. In *Symposium on the Theory of Computing (STOC)* (2009).

[13] DWORK, C., MCSHERRY, F., NISSIM, K., AND SMITH, A. Calibrating noise to sensitivity in private data analysis. In *TCC* (2006), LNCS, Springer, pp. 265–284.

[14] DWORK, C., MCSHERRY, F., AND TALWAR, K. The price of privacy and the limits of lp decoding. In *STOC* (2007), ACM, pp. 85–94.

[15] DWORK, C., AND YEKHANIN, S. New efficient attacks on statistical disclosure control mechanisms. In *CRYPTO* (2008), Springer, pp. 469–480.

[16] ESSÉEN, C. On the Kolmogorov-Rogozin inequality for the concentration function. *Probability Theory and Related Fields 5*, 3 (1966), 210–216.

[17] FELDMAN, D., FIAT, A., KAPLAN, H., AND NISSIM, K. Private coresets. In *To appear in STOC 2009* (2009), ACM.

[18] GHOSH, A., ROUGHGARDEN, T., AND SUNDARARAJAN, M. Universally utility-maximizing privacy mechanisms. In *STOC* (2009).

[19] GUPTA, A., LIGETT, K., MCSHERRY, F., ROTH, A., AND TALWAR, K. Title: Differentially private approximation algorithms. *CoRR arXiv:0903.4510v1[cs.DS]* (2009).

[20] KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S., AND SMITH, A. What can we learn privately? In *FOCS* (2008), IEEE Computer Society, pp. 531–540.

[21] KASIVISWANATHAN, S. P., AND SMITH, A. A note on differential privacy: Defining resistance to arbitrary side information. *CoRR arXiv:0803.39461 [cs.CR]* (2008).

[22] KOLMOGOROV, A. Sur les propriétés des fonctions de concentrations de MP Lévy. *Ann. Inst. H. Poincaré 16* (1958), 27–34.

[23] LEVY, P. Théorie de l'addition des variables aléatoires. *Gauthier-Villars* (1937).

[24] LI, N., LI, T., AND VENKATASUBRAMANIAN, S. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE* (2007), IEEE Computer Society, pp. 106–115.

[25] LITVAK, A., PAJOR, A., RUDELSON, M., AND TOMCZAK-JAEGERMANN, N. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics 195*, 2 (2005), 491–523.

[26] LITVAK, A., PAJOR, A., RUDELSON, M., TOMCZAK-JAEGERMANN, N., AND VERSHYNIN, R. Euclidean embeddings in spaces of finite volume ratio via random matrices. *Journal für die reine und angewandte Mathematik 2005*, 589 (2005), 1–19.

[27] MACHANAVAJJHALA, A., GEHRKE, J., KIFER, D., AND VENKITASUBRAMANIAM, M. l-diversity: Privacy beyond k-anonymity. In *ICDE* (2006), p. 24.

[28] MACHANAVAJJHALA, A., KIFER, D., ABOWD, J. M., GEHRKE, J., AND VILHUBER, L. Privacy: Theory meets practice on the map. In *24th International Conference on Data Engineering (ICDE)* (2008), IEEE, pp. 277–286.

[29] MCSHERRY, F., AND TALWAR, K. Mechanism design via differential privacy. In *FOCS* (2007), IEEE, pp. 94–103.

[30] MILMAN., V., AND SCHECHTMAN, G. *Asymptotic theory of finite dimensional normed spaces*. Springer, 1986.

[31] NISSIM, K., RASKHODNIKOVA, S., AND SMITH, A. Smooth sensitivity and sampling in private data analysis. In *STOC* (2007), ACM, pp. 75–84.

[32] RUDELSON, M., AND VERSHYNIN, R. The least singular value of a random square matrix is $O(n^{-1/2})$. *Comptes rendus-Mathématique* (2008).

[33] RUDELSON, M., AND VERSHYNIN, R. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics 218*, 2 (2008), 600–633.

[34] RUDELSON, M., AND VERSHYNIN, R. The smallest singular value of a random rectangular matrix. *ArXiv e-prints* (2008).

[35] SWEENEY, L. k-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10*, 5 (2002), 557–570.

[36] TALAGRAND, M. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathematiques de l'IHES 81*, 1 (1995), 73–205.

[37] TAO, T., AND VU, V. Random matrices: The distribution of the smallest singular values. *CoRR arXiv:0903.0614v1 [math.PR]* (2009).

[38] VADHAN, S. P. *A study of statistical zero-knowledge proofs*. PhD thesis, Massachusetts Institute of Technology, 1999. Supervisor-Shafi Goldwasser.

[39] WASSERMAN, L., AND ZHOU, S. A statistical framework for differential privacy. *ArXiv.org*, arXiv:0811.2501v1 [math.ST] (2008).

[40] WIGDERSON, A., AND XIAO, D. Derandomizing the Ahlswede-Winter matrix-valued Chernoff bound using pessimistic estimators, and applications. *Theory of Computing 4*, 1 (2008), 53–76.

[41] ZHOU, S., LIGETT, K., AND WASSERMAN, L. Differential privacy with compression. *ArXiv.org*, arXiv:0901.1365v1 [stat.ML] (2009).

A Lower Bounds for Instance-Independent Additive Case

1-way inner products. Let $\mathcal{A}$ be an algorithm that adds instance-independent additive noise. Let $Z \in \mathbb{R}^d$ be the additive noise distribution. The covariance matrix[11], $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])^\top]$, is independent of D (i.e., $\forall D$, $\Sigma(\mathcal{A}(D)) = \Sigma = \mathbb{E}[(Z - E[Z])(Z - E[Z])^\top]$). The proof of the following lemma is identical Lemma 2.1.

Lemma A.1. *Let $\mathcal{A}$ be any ϵ-differentially private algorithm for $\mathcal{I}_1$ that adds instance-independent additive noise. Let $\mathcal{A}(D) = I_{1,d}(D) + Z$. Let $\mathcal{A}(D) \approx_\epsilon \mathcal{A}(D') = \mathcal{A}(D) + \Delta$ for some vector $\Delta \in \{-2, 2\}^d$. Then,*

$$\mathbb{E}[\langle \mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)], \Delta\rangle^2] = \mathbb{E}[\langle \mathcal{A}(D') - \mathbb{E}[\mathcal{A}(D')], \Delta\rangle^2] = \mathbb{E}[(Z - E[Z])(Z - E[Z])^\top] = \Omega(d^2/\epsilon^2).$$

The rest of the proof follows as in Section 2.1.1. We get the following result.

Proposition A.2 (Instance-independent additive case: 1-way inner products). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^d$ be any ϵ-differentially private algorithm for $\mathcal{I}_1$ that adds instance-independent additive noise. Let D be any database which has at least a row of both -1^d and 1^d. Then, $tr(\Sigma(\mathcal{A}(D))) = \Omega(d^2/\epsilon^2)$.*

k-way inner products. The analysis is same as for unbiased case which is explained in Section 2.1.2. We get the following results.

Lemma A.3. *Let $\mathcal{A}$ be any ϵ-differentially private algorithm for $\mathcal{I}_k$ that adds instance-independent additive noise. Let $D_c, D'_c, \widehat{D}$ be the databases in Section 2.1.2. Let $z = \mathcal{I}_k(D'_c) - \mathcal{I}_k(\widehat{D})$ and $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o\rangle$. Then, $\mathbb{E}[\langle \mathcal{A}(D_c) - \mathbb{E}[\mathcal{A}(D_c)], \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2/\epsilon^2)$.*

Proposition A.4 (Instance-independent additive case: k-way inner products over $\{-1, 1\}^d$). *Let $\mathcal{A} : \{-1, 1\}^d \to \mathbb{R}^{m_k}$ be any ϵ-differentially private algorithm for $\mathcal{I}_k$ that adds instance-independent additive noise. Let $D_c = 1^d$. Then, $tr(\Sigma(\mathcal{A}(D_c))) = \Omega(m_k^2/\epsilon^2)$.*

Theorem A.5 (Instance-independent additive case). *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-way inner products that adds instance-independent additive noise and that for every database $D \in (\{-1, 1\}^d)^n$ has an average variance of $o(m_k/\epsilon^2)$ for $\mathcal{A}(D)$ is not ϵ-differentially private. Also, any algorithm $\mathcal{A}$ for releasing all k-way conjunctions that adds instance-independent additive noise and that for every database $D \in (\{0, 1\}^d)^n$ has an average variance of $o(m_k/(2^k\epsilon^2))$ for $\mathcal{A}(D)$ is not ϵ-differentially private.*

[11] All our results for instance-independent case also hold if we replace the covariance matrix, by mean squared error matrix $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{I}_1(D))(\mathcal{A}(D) - \mathcal{I}_1(D))^\top] = \mathbb{E}[ZZ^\top]$.

B Going From Inner Products to Conjunctions

Let $D_{\pm}$ be a database from $(\{-1,1\}^d)^n$. Let the Boolean variables $y_1,\ldots,y_d$ represent the d columns of $D_{\pm}$ (i.e., column i in $D_{\pm}$ contains assignments to variable y_i). Define variables $x_1,\ldots,x_d$ as $x_i = (y_i+1)/2$. Construct $D_0 \in (\{0,1\}^d)^n$ from $D_{\pm}$ by replacing all the -1's by 0's. The variables $x_1,\ldots,x_d$ represent the d columns of D_0.

Let us set $k=2$, and look at all the 4 possible conjunctions on two variables x_i and x_2. The conjunction predicates on x_1, x_2 and the inner product predicates on y_i, y_j can be related using a Hadamard matrix as,

$$\underbrace{\begin{pmatrix} n \\ i_{y_j}(D_{\pm}) \\ i_{y_i y_j}(D_{\pm}) \\ i_{y_i}(D_{\pm}) \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}}_{H} \underbrace{\begin{pmatrix} c_{x_1x_2}(D_0) \\ c_{\bar{x}_i\bar{x}_j}(D_0) \\ c_{\bar{x}_i x_2}(D_0) \\ c_{x_1\bar{x}_j}(D_0) \end{pmatrix}}_{V} \tag{13}$$

Therefore, $||U|| = ||HV||$, or $||U|| = 2||V||$.

Now consider the vectors, $\mathcal{I}_2(D_{\pm})$ (= 2-way inner product predicates evaluated on $D_{\pm}$) , $\mathcal{I}_1(D_{\pm})$ (= 1-way inner product predicates evaluated on $D_{\pm}$), and $\mathcal{I}_0(D_{\pm}) = (n)$, and let $\mathcal{I}_{\leq 2}(D_{\pm})$ be a concatenation of these three vectors. By an extension of Equation 13, it can be shown that

$$\mathcal{I}_{\leq 2}(D_{\pm}) = \Pi \cdot diag(H,\ldots,H) \cdot \mathcal{C}_2(D_0),$$

where $diag(H,\ldots,H)$ is a block diagonal matrix and Π is a suitable projection matrix. Therefore,

$$||\mathcal{I}_2(D_{\pm})|| \leq ||\mathcal{I}_{\leq 2}(D_{\pm})|| = ||\Pi \cdot diag(H,\ldots,H) \cdot \mathcal{C}_2(D_0)|| \leq 2||\mathcal{C}_2(D_0)||.$$

In general, for higher k's, $||\mathcal{I}_k(D_{\pm})|| \leq 2^{k/2}||\mathcal{C}_k(D_0)||$. The following proposition and corollary follow immediately.

Proposition B.1. *If there exists an algorithm $\mathcal{A}$ for $\mathcal{C}_k$ that has $tr(\Sigma(\mathcal{A}(D_0))) \leq T$, then there exists an algorithm $\mathcal{B}$ for $\mathcal{I}_k$ that has $tr(\Sigma(\mathcal{B}(D_{\pm}))) \leq 2^{k/2}T$.*

Corollary B.2. *If there exists a database $D \in (\{-1,1\}^d)^n$ such that no algorithm $\mathcal{B}$ for $\mathcal{I}_k$ has $tr(\Sigma(\mathcal{B}(D))) \leq T$, then there exists a database $D^* \in (\{0,1\}^d)^n$ such that no algorithm $\mathcal{A}$ for $\mathcal{C}_k$ has $tr(\Sigma(\mathcal{A}(D^*))) \leq T/2^{k/2}$.*

C Extension to (ϵ,δ)-differential Privacy

Our results extend to (ϵ,δ)-differential privacy. Letš start by formally defining (ϵ,δ)-differential privacy.

Definition C.1 ((ϵ,δ)-differential privacy [13]). *Let $\delta = \delta(n)$ be a negligible function of n. A randomized algorithm $\mathcal{A}$ is (ϵ,δ)-differentially private if for all neighboring databases D, D', and for all sets $\mathcal{S}$ of possible outputs $\Pr[\mathcal{A}(D) \in \mathcal{S}] \leq \exp(\epsilon) \cdot \Pr[\mathcal{A}(D') \in \mathcal{S}] + \delta$. The probability is taken over the random coins of the algorithm $\mathcal{A}$.*

Let X and Y be random variables taking values in a set $\mathcal{O}$. We use $X \approx_{(\epsilon,\delta)} Y$ to indicate that random variables X and Y are (ϵ,δ)-indistinguishable, i.e.,

$$\forall \mathcal{S} \subseteq \mathcal{O}, \exp(-\epsilon) \cdot \Pr[Y \in \mathcal{S}] - \delta \leq \Pr[X \in \mathcal{S}] \leq \exp(\epsilon) \cdot \Pr[Y \in \mathcal{S}] + \delta.$$

C.1 Lower Bounds for Instance-Independent Additive (ϵ, δ)-differential Privacy

If $\mathcal{A}$ is an (ϵ, δ)-differentially private algorithm that adds instance-independent additive noise then there is an easy extension of Theorem A.5.

Theorem C.2 (Instance-Independent Additive Noise: Extension of Theorem A.5). *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-way inner products that adds instance-independent additive noise and that for every database $D \in (\{-1,1\}^d)^n$ has an average variance of $o(m_k(1-\delta)^2/\epsilon^2)$ for $\mathcal{A}(D)$ is not (ϵ, δ)-differentially private. Also, any algorithm $\mathcal{A}$ for releasing all k-way conjunctions that adds instance-independent additive noise and that for every database $D \in (\{0,1\}^d)^n$ has an average variance of $o(m_k(1-\delta)^2/(2^k\epsilon^2))$ for $\mathcal{A}(D)$ is not (ϵ, δ)-differentially private.*

C.2 Lower Bounds for Unbiased (ϵ, δ)-differential Privacy

Unlike the instance-independent additive noise case, there is no extension of the lower bound in Theorem 2.9 to (ϵ, δ)-differential privacy. In fact, we now show that any (ϵ, δ)-differential privacy mechanism can be converted into an unbiased (ϵ, δ)-differential privacy mechanisms with a little more noise. We analyze lower bounds for general (ϵ, δ)-differential privacy mechanisms in the next subsection.

Lemma C.3. *Let $\mathcal{F}$ be any function class. Let $\mathcal{A}$ be any (ϵ, δ)-differential privacy mechanism for $\mathcal{F}$, there exists an unbiased $(\epsilon, 2\delta)$-differential privacy mechanism $\mathcal{B}$ for $\mathcal{F}$ such that for all databases D, $tr(\Sigma(\mathcal{B}(D))) \leq tr(\Sigma(\mathcal{A}(D))) + (|\mathcal{F}|n^2)/\delta$.*

Proof. Define $\mathcal{B}$ as follows:

$$\mathcal{B}(D) = \begin{cases} \mathcal{A}(D) & \text{with probability } 1-\delta, \\ \frac{\mathcal{F}(D) - (1-\delta)\,\mathbb{E}[\mathcal{A}(D)]}{\delta} & \text{with probability } \delta. \end{cases}$$

$\mathcal{B}$ is unbiased because for all D, $\mathbb{E}[\mathcal{B}(D)] = \mathcal{F}(D)$. Since, $\mathcal{A}$ is (ϵ, δ)-differentially private, $\mathcal{B}$ is $(\epsilon, 2\delta)$-differentially private.

Let $m = |\mathcal{F}|$. Let $\mathcal{F}(D) = (f_1(D), \ldots, f_m(D))$. Similarly, let $\mathcal{A}(D) = (\mathcal{A}_1(D), \ldots, \mathcal{A}_m(D))$ and $\mathcal{B}(D) = (\mathcal{B}_1(D), \ldots, \mathcal{B}_m(D))$. Then, for every $j \in [m]$, $\mathbb{E}[(\mathcal{B}_j(D) - f_j(D))^2]$ can be bounded as,

$$\begin{aligned} \mathbb{E}[(\mathcal{B}_j(D) - f_j(D))^2] &= (1-\delta)\,\mathbb{E}[(\mathcal{A}_j(D) - f_j(D))^2] + \delta\left(\frac{f_j(D) - (1-\delta)\,\mathbb{E}[\mathcal{A}_j(D)]}{\delta} - f_j(D)\right)^2 \\ &= (1-\delta)\,\mathbb{E}[(\mathcal{A}_j(D) - f_j(D))^2] + \frac{(1-\delta)^2(f_j(D) - \mathbb{E}[\mathcal{A}_j(D)])^2}{\delta} \\ &\leq \mathbb{E}[(\mathcal{A}_j(D) - f_j(D))^2] + n^2/\delta. \end{aligned}$$

The last inequality follows because $0 \leq f_j(D) \leq n$ and $0 \leq \mathbb{E}[\mathcal{A}_j(D)] \leq n$. Therefore,

$$tr(\Sigma(\mathcal{B}(D))) = \sum_{j=1}^{m} \mathbb{E}[(\mathcal{B}_j(D) - f_j(D))^2] \leq \sum_{j=1}^{m} (\mathbb{E}[(\mathcal{A}_j(D) - f_j(D))^2] + n^2/\delta) = tr(\Sigma(\mathcal{A}(D))) + mn^2/\delta.$$

□

C.3 Lower Bounds for General (ϵ, δ)-differential Privacy

We state the extensions of the statements in Section 2.2 to (ϵ, δ)-differential privacy. Let us first consider 1-way inner products ($\mathcal{I}_1$).

Lemma C.4 (Extension of Lemma 2.10). *Let $\mathcal{A}$ be any $(1/2, \delta)$-differentially private algorithm for $\mathcal{I}_1$. Let $\mathcal{A}(D) \approx_{(1/2,\delta)} \mathcal{A}(D')$. Let $\mathcal{I}_1(D') = \mathcal{I}_1(D) + \Delta$ for some $\Delta \in \{-2, 2\}^d$. Then at least one of $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta\rangle^2]$ or $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_1(D'), \Delta\rangle^2]$ is $\Omega(d^2(1-\delta)^2)$.*

Proof. As in Lemma 2.10 we set $X = \langle \mathcal{A}(D) - \mathcal{I}_1(D), \Delta\rangle$, $Y = \langle \mathcal{A}(D') - \mathcal{I}_1(D'), \Delta\rangle$, and $a = \Delta^\top \Delta = d$ in Lemma 2.11. □

Using Lemma C.4 instead of Lemma 2.10 in the proof of Proposition 2.17 we get the following result.

Proposition C.5 (Extension of Proposition 2.17). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^d$ be any $(1/2, \delta)$-differentially private algorithm for $\mathcal{I}_1$. Then with probability at least $1/n$ over D chosen uniformly at random from $\mathcal{D}$, $tr(\Sigma(\mathcal{A}(D))) = \Omega(\min\{d^2(1-\delta)^2, nd(1-\delta)^2/\log d\})$.*

Even for higher way inner products, the extension is easy. Let $\Sigma(\mathcal{A}(D)) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{I}_k(D))(\mathcal{A}(D) - \mathcal{I}_k(D))^\top]$. Using the same framework as in Section 2.2.2 we get.

Lemma C.6 (Extension of Lemma 2.18). *Let $\mathcal{A}$ be any $(1/2, \delta)$-differentially private algorithm for $\mathcal{I}_k$. Let $\mathcal{A}(D) \approx_{(1/2,\delta)} \mathcal{A}(D')$. Let $\mathcal{I}_k(D') = \mathcal{I}_k(D) + \tilde{z}$ for some $\tilde{z} \in \text{supp}(\tilde{z}_r)$ and $z = o - \tilde{z}$. Then, at least one of $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z\rangle^2]$ or $\mathbb{E}[\langle \pi z, \mathcal{A}(D') - \mathcal{I}_k(D')\rangle^2]$ is $\Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)$.*

Using this lemma, Proposition 2.21 can be extended as follows.

Proposition C.7 (Extension of Proposition 2.21). *Let $\mathcal{A} : (\{-1, 1\}^d)^n \to \mathbb{R}^{m_k}$ be any $(1/2, \delta)$-differentially private algorithm for $\mathcal{I}_k$. Let D_r be a database chosen uniformly at random from $\mathcal{D}$ and $\widetilde{D}_r$ be a database chosen uniformly at random from $\widetilde{\mathcal{D}}$. Then with probability at least $1/n$, at least one of $tr(\Sigma(\mathcal{A}(D_r)))$ or $tr(\Sigma(\mathcal{A}(\widetilde{D}_r)))$ is $\Omega(\min\{m_k^2(1-\delta)^2, nm_k(1-\delta)^2/\log m_k\})$.*

By using the trick described in Section 2.3 we can introduce ϵ into the lower bound. The algorithm $\mathcal{A}'$ will be $(1/2, (e^\epsilon\delta)/(2\epsilon))$-differentially private (using Claim 2.23). For small ϵ, $e^\epsilon\delta/(2\epsilon) \approx \delta$.

Theorem C.8 (Extension of Theorem 2.26). *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-way inner products that for every database $D \in (\{-1, 1\}^d)^n$ has an average mean squared error of*

$$\min\left\{o\left(m_k(1-\delta)^2/\epsilon^2\right), o\left(n(1-\delta)^2/\epsilon \log m_k\right)\right\}$$

for $\mathcal{A}(D)$ is not (ϵ, δ)-differentially private. Also, any algorithm $\mathcal{A}$ for releasing all k-way conjunctions that for every database $D \in (\{0, 1\}^d)^n$ has an average mean squared error of

$$\min\left\{o\left(m_k(1-\delta)^2/(2^k\epsilon^2)\right), o\left(n(1-\delta)^2/(2^k\epsilon \log m_k)\right)\right\}$$

for $\mathcal{A}(D)$ is not (ϵ, δ)-differentially private.