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

Shiva Kasiviswanathan[*] Mark Rudelson[†] Adam Smith[‡] Jonathan Ullman[§]

Abstract

Marginal (contingency) tables are the method of choice for government agencies 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 [9, 14, 15, 16] 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. We provide lower bounds for releasing k-attribute marginal tables under (i) *minimal privacy*, a general privacy notion which captures a large class of privacy definitions, and (ii) *differential privacy*, a rigorous notion of privacy that has received extensive recent study. Our main contributions are:

- We give efficient polynomial time attacks which allow an adversary to reconstruct sensitive information given insufficiently perturbed marginal table releases. Using these reconstruction attacks, we show that for releasing all k-attribute marginal tables with constant k, $\widetilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^{k-1}}\})$ average distortion per entry is necessary for any privacy notion satisfying at least a minimalistic privacy guarantee. Under this privacy guarantee this bound is tight.
- Our above reconstruction-based attacks require a new lower bound on the least singular value of a random matrix with correlated rows. For a constant k, consider a matrix $M^{(k)}$ with d^k rows which are formed by taking all possible k-way entry-wise products of an underlying set of d random vectors from $\{0,1\}^n$. We show that even if $M^{(k)}$ is nearly square its least singular value is $\widetilde{\Omega}(\sqrt{d^k})$ with high probability— asymptotically, the same bound as one gets for a matrix with *independent* rows. The proof introduces several new tools for dealing with random matrices with correlated entries and could be of independent interest.
- We obtain stronger lower bounds for differential privacy. For releasing all k-attribute marginal tables with constant k, previous work showed that $\widetilde{O}(\min\{n, (n^2 d)^{1/3}, \sqrt{d^k}\})$ average distortion per entry is *sufficient* for satisfying differential privacy (ignoring the dependence on privacy parameters). We give a lower bound of $\widetilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$, which is tight for $n = \widetilde{\Omega}(d^k)$. Moreover, for a natural and popular class of mechanisms based on adding *instance-independent* noise, our lower bound can be strengthened to $\Omega(\sqrt{d^k})$, which is tight for all n. Our lower bounds for differential privacy extend even to non-constant k, losing roughly a factor of $\sqrt{2^k}$ compared to best-known upper bounds for large n.

[*] 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.

[§] SEAS, Harvard University, jullman@seas.harvard.edu.

Contents

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 [9] initiated a rigorous study of the tradeoff between privacy and utility. The notion of *differential privacy* [14] 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 (see [10, 11] for two recent surveys). 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 (noise) is necessary for particular functions f. Some of these bounds apply only to differential privacy (e.g., [14, 20, 21]); other bounds rule out *any* reasonable notion of privacy by showing how to reconstruct almost all of the data D given sufficiently accurate approximations to $f(D)$ [9, 15, 16]. 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 *marginal contingency tables* (marginal tables, in short), under both minimal and differential 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 k attributes in T. We call a marginal table for a set of k attributes a k-attribute marginal table. The d-attribute marginal table is the "full" contingency table for the data set.

Marginal tables are the workhorses of categorical data analysis and, in particular, of data analysis in the medical, social and behavioral sciences (e.g., clinical trials, public health studies, and education statistics). In addition to being easy to interpret, they are sufficient statistics for popular classes of probabilistic models [4]. (As a simple example: for binary data, the mean vector and covariance matrix, which capture linear dependencies among attributes, are equivalent to the set of all 2-attribute marginal tables.) Because of this, they are the format of choice for data release by government statistical bureaus [3]. However, many of the fields in which categorical data are used generate highly sensitive data. Researchers and government agencies have ethical and legal responsibilities to protect the confidentiality of the individuals whose data they collect. Consequently, the confidentiality of contingency table releases has been an active topic of research in statistics for over thirty years (see, for example, [19, 36]). Understanding the extent to which marginal tables can be released while guaranteeing a rigorous, meaningful notion of privacy is an important problem.

1.1 Our Contributions

Let $\mathcal{C}_k(D)$ be the set of all k-attribute marginal tables (equivalently, the frequencies of all possible k-attribute conjunctions) for a database $D \in (\{0,1\}^d)^n$. There are $\binom{d}{k}$ such tables; however, it is convenient to think of $\mathcal{C}_k(D)$ as a single real vector of length $2^k\binom{d}{k}$. The $\widetilde{O}(\cdot)$ notation below hides polylogarithmic factors in n, d, k, and $\widetilde{\Omega}(\cdot)$ hides inverse polylogarithmic factors in n, d, k.

We give lower bounds for simultaneously estimating all the entries of $\mathcal{C}_k(D)$ privately. As a point of reference, for constant k, the best-known differentially private algorithms [5, 6] have an $\widetilde{O}(\min\{n, (n^2 d)^{1/3}, \sqrt{d^k}\})$ average distortion per entry. Our lower bounds match this upper bound in different respects.

(1) Lower Bounds for Minimal Privacy: We show that algorithms that do not sufficiently distort the marginal tables fail to satisfy a large class of "privacy" definitions. We define two violations of privacy: *attribute non-privacy* and *row non-privacy*.[1]

Each of these rules out a large class of popular definitions of privacy. Row non-privacy rules out definitions that protect an entire row of the database even given leakage of other rows; such definitions include differential privacy as well as several definitions popular in the *randomized response* literature [40, 1, 18]. Attribute non-privacy rules out any definition that guarantees the secrecy of a particular "sensitive" attribute even when all other attributes are known to an attacker; such definitions include K-anonymity [38] and its variants [27, 26, 7, 28, 43], as well as the notions ruled out by row non-privacy.

Using a "reconstruction" attack outlined below **(2)**, we show that for any constant k, releasing $\mathcal{C}_k(D)$ with distortion $o(\min\{\sqrt{n}, \sqrt{d^{k-1}}\})$ per entry allows an adversary to efficiently reconstruct large fraction of the sensitive attribute entries given the nonsensitive values, thus *violating* attribute privacy (the bound holds even for releasing only all those k-attribute tables that involve the sensitive attribute and $k-1$ other attributes). Moreover, releasing $\mathcal{C}_k(D)$ with distortion $o(\min\{\sqrt{n}, \sqrt{d^k}\})$ per entry allows an adversary to efficiently reconstruct large fraction of the rows of D, even though this would not be possible without the release, thus *violating* row privacy. Both these bounds are (almost) tight, as there is an algorithm which is neither attribute non-private nor row non-private and which for every database D adds $\widetilde{O}(\min\{\sqrt{n}, \sqrt{d^k}\})$ distortion per entry of $\mathcal{C}_k(D)$. The formal bounds for these privacy notions are stated in Table 1 and discussed in Section 2. We discuss the significance of these bounds below (Section 1.2).

(2) Reconstruction Attack & the Least Singular Value of Random Matrices with Correlated Rows: The bounds on minimal privacy **(1)** above require significantly different techniques from previous work. Previous lower bounds [9, 15, 16] 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 [9, 15, 16] or from an algebraic code [16]). When $k = 1$ a similar flexibility is available in our lower bounds; the matrix $M^{(1)}$ that arise in our lower bounds is a $\{0,1\}^{d\times n}$ matrix with independent random entries. However, for $k > 1$ the rows of the matrix $M^{(k)}$ that arise in our lower bounds are highly correlated: the matrix $M^{(k)}$ has d^k rows which are formed by taking all possible k-way *entry-wise products*[2] of the rows of the random matrix $M^{(1)}$. 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 $(0,1)$-matrix

[1] Alternatively, we might call these "attribute leakage" and "row leakage". We use "non-privacy" for consistency with [9, 15, 16].

[2] The entry-wise product of k vectors $p_1, \ldots, p_k \in \mathbb{R}^n$ is the vector in $q \in \mathbb{R}^n$ with entries $q_i = \prod_{j=1}^k p_{j_i}$.

with *correlated* rows has approximately the same least singular value as a random $(0,1)$-matrix with *independent* rows.

Tight bounds are known on the least singular values of various types of matrices (e.g., square, rectangular) with independent random entries (see, e.g., [34, 35, 33] and references therein). The least singular value of an $N \times n$ matrix with $(0,1)$ independent random entries and $N \geq n$ is $\Theta(\sqrt{N})$ with exponentially high probability (in fact, even non-asymptotic bounds are known, see Rudelson and Vershynin [35]). To deal with the dependencies, we develop several new tools, which may be of independent interest. We show that for any constant k if the random matrix $M^{(1)}$ has less than $d^k/\log^{k-2} n$ columns, then the least singular value of $M^{(k)}$ is $\widetilde{\Omega}(\sqrt{d^k})$ with exponentially high probability. Therefore, the least singular value of $M^{(k)}$ is approximately the same as that of a $d^k \times n$ random matrix with independent entries, but $M^{(k)}$ (constructed out of $M^{(1)}$) uses far lower randomness.

The proof is challenging because correlations make powerful measure concentration tools hard to apply. We first reduce the problem to bounding the least singular value of a (related) random centered matrix $\tilde{\Pi}$. The smallest singular value of $\tilde{\Pi}$ is the minimum of $\|\tilde{\Pi}x\|$, over x from the unit sphere. An important tool in the proof is bounding the *small ball probability*, which is the probability that $\|\tilde{\Pi}x\|$ is small for a *fixed* vector x. To obtain a uniform lower bound for $\|\tilde{\Pi}x\|$, we decompose the unit sphere into many pieces, and for each piece use epsilon-net arguments tailored according to the small ball probability. Then, we obtain a uniform lower estimate on the net, which is then extended to the whole unit sphere by approximation.

In the privacy context, our spectral bound allows for a reconstruction algorithm of the form $\hat{s} = \text{round}(M^{(k)}_{inv} \cdot (M^{(k)} s + e))$, where s is a $(0,1)$-vector and $M^{(k)}_{inv}$ is an appropriate pseudoinverse of $M^{(k)}$. We show that releasing $M^{(k)}s$ with distortion $o(\sqrt{n})$ per entry allows the adversary to reconstruct $n - o(n)$ bits of s (that is, to find $\hat{s}$ that agrees in almost all entries of s), as long as $n = o(d^k)$. One can extend the result to get a lower bound of $\widetilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$ for all n.

(3) Lower Bounds for Differential Privacy: Using a disjoint set of techniques, we show a stronger lower bound for releasing k-attribute marginal tables under the notion of (ϵ, δ)-differential privacy. The precise bounds are stated in Table 1 and discussed in Section 4. Here, we treat ϵ and δ as constants.

For constant k, the best-known (ϵ, δ)-differentially private algorithms [5, 6] yield an average distortion per entry of $\widetilde{O}(\min\{n, (n^2 d)^{1/3}, \sqrt{d^k}\})$, while our lower bound is $\widetilde{\Omega}(\min\{\sqrt{n}, \sqrt{d^k}\})$. Our bounds imply that the technique of Blum *et al.* [5], which adds carefully calibrated Gaussian noise to each entry in $\mathcal{C}_k$ is tight for large databases (when $n = \widetilde{\Omega}(d^k)$). Moreover, for a natural and popular class of algorithms based on adding *instance-independent* noise [5, 14, 3], we strengthen this bound to $\Omega(\sqrt{d^k})$, which is tight for all n.

Our lower bounds for differential privacy extend even to non-constant k. Here, we get a lower bound of $\widetilde{\Omega}(\min\{\sqrt{n}/2^k, \sqrt{m_k}/2^k\})$ on the average distortion per entry, where $m_k = \binom{d}{k}$. For $n = \widetilde{\Omega}(m_k)$, this is loose by a factor of $\sqrt{2^k}$ when compared to the best-known upper bound. This lower bound can again be strengthened for the instance-independent case.

Consider any differentially private algorithm $\mathcal{A}$ for $\mathcal{C}_k$. The rough idea behind these lower bounds is to start with a particular database D and bound the projection of the *mean squared error (MSE)* matrix of $\mathcal{A}(D)$ along a large set of directions. If the algorithm adds instance-independent noise then we show that this set of directions contains an (almost) orthonormal basis, allowing us to lower bound the trace of the MSE matrix, and hence the average distortion per entry. In the general case (when the distortion is instance-dependent), we use concentration inequalities for matrix-valued random variables to show that for appropriately chosen *random* databases, the trace of the MSE matrix is large with high probability.

We expect that the linear algebraic techniques developed for this bound should be useful for bound-

Privacy Guarantee	Upper Bound on Noise	Lower Bound on Noise
Not attribute non-privacy	$\widetilde{O}(\min\{\sqrt{n}, \sqrt{d^k}\})$	$\widetilde{\Omega}\left(\min\left\{\sqrt{n}, \sqrt{d^{k-1}}\right\}\right)$
Not row non-privacy	$\widetilde{O}(\min\{\sqrt{n}, \sqrt{d^k}\})$	$\widetilde{\Omega}\left(\min\left\{\sqrt{n}, \sqrt{d^{k}}\right\}\right)$
(ϵ, δ)-diff. privacy (Inst. Ind.)	$O\left(\frac{\sqrt{\binom{d}{k}\log(1/\delta)}}{\sqrt{2^k}\epsilon}\right)$ [5, 3]	$\Omega\left(\frac{\sqrt{\binom{d}{k}}(1-\delta/\epsilon)}{2^k\epsilon}\right)$
(ϵ, δ)-diff. privacy	$\widetilde{O}\left(\min\left\{n, \left(\frac{n^2dk}{\epsilon}\right)^{\frac{1}{3}}, \frac{\sqrt{\binom{d}{k}\log(1/\delta)}}{\sqrt{2^k}\epsilon}\right\}\right)$ [5, 3, 6]	$\widetilde{\Omega}\left(\min\left\{\frac{\sqrt{n}(1-\delta/\epsilon)}{2^k\sqrt{\epsilon}}, \frac{\sqrt{\binom{d}{k}}(1-\delta/\epsilon)}{2^k\epsilon}\right\}\right)$

Table 1: *Upper and lower bounds on the average noise per cell entry for releasing all k-attribute marginal tables (or all k-way conjunction predicates) under various privacy guarantees. The results on attribute non-privacy and row non-privacy are for k being a constant. The n term in the upper bound for (ϵ, δ)-differential privacy (last row) comes from an algorithm that releases a vector of $n/2$'s for all D's. The $\widetilde{O}(\cdot)$ notation hides polylogarithmic factors d, n, k, and $\widetilde{\Omega}(\cdot)$ hides inverse polylogarithmic factors in d, n, k. All the uncited results appear in this paper.*

ing the required distortion of a wide range of differentially private releases.

1.2 Significance of the Privacy Lower Bounds

We summarize previous and new bounds in Table 1. Dinur and Nissim [9] showed that if a mechanism answers (or allows the user to compute) $O(n \log n)$ arbitrary inner product queries on a database (vector) $s \in \{0,1\}^n$ with noise $o(\sqrt{n})$ per response, then an adversary can reconstruct $n - o(n)$ entries of s. Their attack was subsequently extended to use a linear number of queries [15], allow a small fraction of answers to be arbitrarily distorted [15], and run significantly more quickly [16]. These reconstruction attacks provide lower bounds for various minimal notions of privacy; our results extend the scope of these bounds significantly.

There were also several known lower bounds specific to differential privacy, though they are not directly relevant to marginal tables [14, 32, 20]. *Subsequently to our work,* Hardt and Talwar [21] gave upper and lower bounds for releasing a variety of linear functions (including marginal tables) for the special case of "pure" ϵ-differential privacy (with $\delta = 0$). For the case of 1-attribute marginal tables, their bound of $\Omega(d/\epsilon)$ improves on ours; we conjecture that their techniques lead to a bound of $\widetilde{\Omega}(d^k/\epsilon)$ for releasing constant k-attribute marginal tables under ϵ-differential privacy. However, their techniques break down for even slightly relaxed privacy notions such as (ϵ, δ)-differential privacy.

We see our new lower bounds as interesting for several reasons.

Natural symmetric functions. In their simplest form, the inner product queries considered by [9, 15, 16] require the adversary to be able to "name rows", that is, specify a coefficient for each entry of the vector s.

Thus, the lower bound does not apply directly to any functionality that is symmetric in the rows of the data set such as marginal tables. It was pointed out in [8] 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 of 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). The lower bounds we give for marginal table releases are the first for symmetric functions regularly released by official statistics agencies; one can think of our reconstruction attacks as using *conjunctions* as weak hash functions to implement the idea of [8].

When is distortion acceptably low? It is natural to ask at what point the distortion required for privacy interferes with statistical analysis. There is no simple answer, but for the "predicate queries" considered here, where each entry counts the number of occurrences of a predicate in the underlying data set, there is a large class of statistical models which inherently have "sampling error", that is standard deviation of the observed statistics, of $\Omega(\sqrt{n})$. A crude rule of thumb, then, is that the distortion interferes seriously when it is not $o(\sqrt{n})$ [14, 3, 37]. Our lower bounds of $\widetilde{\Omega}(\min(\sqrt{n}, \sqrt{d^k}))$ show that for even modest values of d and k, the data set n must be very large to get distortion $o(\sqrt{n})$.

The "dimension" of marginal tables. The reconstruction attacks [9, 15, 16] above show a lower bound of roughly $\min\{\sqrt{n}, \sqrt{m}\}$ on the distortion required to answer a set of m random, independent queries about a data set of size n. However, the bounds heavily rely on independence of the queries. This raises the question of whether certain interesting classes of queries could be answered with much less noise. For example, if a set of queries is linearly dependent, then one can compute noisy answers to only a few queries (a spanning set), and deduce the rest using the linear relationships. Both of our bounds can be interpreted as showing that the marginal statistics of a data set are, in a sense that depends on the notion of privacy, *far* from any low dimensional subspace. In particular, we show that the $\binom{d}{k}2^k$ different entries of the k-attribute marginal tables hide a set of $\Omega(d^k)$ "nearly independent" underlying features–as far as privacy is concerned, they have dimension close to $\binom{d}{k}2^k$. It is natural to ask: what properties of a set of queries lead to this type of behavior, in general? Our techniques suggest that the right notion is related to inapproximability by low-dimensional linear spaces; it is still unclear how to formulate this notion precisely and how it relates to concepts such as VC-dimension that play a role in other works on privacy [6, 22].

1.3 Known Upper Bounds for Differential Privacy

Let $m_k = \binom{d}{k}$. In [5, 14] it was shown that addition of carefully calibrated noise to functions satisfying a Lipschitz condition is enough to ensure differential privacy. Applied to conjunctions, they show that 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 [5], while adding 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 (with $\delta = 0$) [14]. Barak *et al.* [3] improve the dependency on k in these results, saving a factor of approximately $\sqrt{2^k}$ in the required distortion.

In a different vein, Blum *et al.* [6] adapt the exponential sampling technique of [30] to release a synthetic data set. One can use their techniques to release $\mathcal{C}_k(D)$ with distortion $\widetilde{O}((n^2dk/\epsilon)^{2/3})$ in each entry (see Appendix A). The dependency on d and k in [6] is much better than in the additive noise mechanisms, but the dependency on n is significantly worse; in particular, our results show that one cannot reduce the dependency on n without incurring a dependency on d^k.

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 $\mathrm{supp}(\cdot)$, 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 used in the paper are by default column vectors. We use $(a)^n$ to denote a vector of length n of all a's. 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 to denote 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. The length of projection of v_1 onto v_2 is then $\langle v_1, v_2 \rangle / \|v_2\|$. For a matrix M, $tr(M)$ denotes the trace and $\|M\|_\infty$ denotes the operator norm. We use $diag(a_1, \dots, a_n)$ to denote an $n \times n$ diagonal matrix with entries $a_1, \dots, a_n$ along the main diagonal. 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 $\sigma_n(M) = \inf_{z:\|z\|=1} \|Mz\|$.

In our analysis, we assume that a (private) algorithm $\mathcal{A}$ for a function class $\mathcal{F}$ on input D releases a vector

$$\mathcal{A}(D) = (\mathcal{A}_1(D), \dots, \mathcal{A}_{|\mathcal{F}|}(D)),$$

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 $\mathcal{F}(\widehat{D}) = (f_1(\widehat{D}), \dots, f_{|\mathcal{F}|}(\widehat{D}))$. The perturbation introduced doesn't change by this second step, therefore, we can think of $\mathcal{A}$ as directly releasing the sanitized vector.

Boolean Conjunctions. It is convenient to describe the results in terms of releasing conjunction predicates over the domain $\{0,1\}^d$. Each $x \in \{0,1\}^d$ is interpreted as an assignment to d Boolean variables $x_1, \dots, x_d$. 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 for all $i \in [d], x_i = 1$ if $v_i = 1$ and $x_i = 0$ if $v_i = -1$. 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 $x_1, \dots, x_d$. 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 of D 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 predicates in $\mathcal{C}_k$ evaluated on D.

1.5 Organization of the Paper

In Section 2, we define row non-privacy and attribute non-privacy. We present a new reconstruction attack and analyze the lower bound on noise needed to prevent this attack. We then show how this lower bound implies that releasing all k-way marginal tables with not enough noise leads to row non-privacy and attribute non-privacy. A crucial tool in these lower bound analyses is the bound on the least singular value of a random matrix with correlated rows. In Section 3, we present this least singular value bound. In Section 4, we present our lower bounds on noise for releasing all k-way marginal tables under the popular notion of differential privacy. As mentioned earlier, our results are tighter if restricted to differentially private algorithms that add instance-independent noise. We present our lower bounds for the instance-independent case in Section 4.1, and in Section 4.2, we present our lower bounds for the general case.

2 Lower Bounds on Noise for Minimal Privacy

In this section, we introduce a new reconstruction attack based on analyzing the least singular value of a random matrix with correlated entries. We then use the reconstruction attack to establish lower bounds on noise needed for releasing k-attribute marginal tables under the notions of attribute non-privacy and row non-privacy. We treat k as a constant in this section. The upper bounds on noise are discussed in Appendix B.

The lower bounds for our minimal privacy definitions proceed by "reducing" an instance of the reconstruction problem for a matrix with correlated rows into a marginal table release problem. We now formally define attribute non-privacy and row non-privacy, and explain the reductions from the reconstruction problem.

Definition 2.1 (Attribute Non-Privacy). *An algorithm $\mathcal{A}$ for releasing all k-way conjunction predicates is* attribute non-private *if there exists a polynomial time adversary such that for every $s \in \{0,1\}^n$ there exists a database $D_{at}(s) \in (\{0,1\}^d)^n$ whose last column is s, such that the adversary on given as input $\mathcal{A}(D_{at}(s))$ and the first $d-1$ columns of $D_{at}(s)$, can reconstruct at least $\widetilde{\Omega}(\min\{n, d^{k-1}\})$ entries of s with probability $1 - negl(d)$.*

This definition captures a common model in the data privacy literature (e.g., [38, 42, 27, 26, 7, 28, 43]) where one assumes that a database consists of $d-1$ *nonsensitive* attributes (e.g., demographic information), which can be learned from other sources, and one *sensitive* attribute (e.g., disease). The attribute non-privacy lower bound applies to any notion of privacy that purports to protect individual values of the sensitive attribute (the lower bound applies in particular, to differential privacy but also, e.g., to the notion of privacy implicit in the popular "K-anonymization" scheme [38] and its recent variants [27, 26, 7, 28, 43]).

To define the reduction from the reconstruction problem, we need the following definition of entry-wise product of vectors and matrices.

Definition 2.2 (Entry-wise Product). *The entry-wise product of vectors $p, q \in \mathbb{R}^n$ is the vector in $p \odot q \in \mathbb{R}^n$ with entries $(p \odot q)_i = p_i \cdot q_i$. If A is an $N_1 \times n$ matrix, and B is an $N_2 \times n$ matrix, denote by $A \odot B$ an $N_1 N_2 \times n$ matrix, whose rows are entry-wise products of the rows of A and B: $(A \odot B)_{j,k} = A_j \odot B_k$, where $(A \odot B)_{j,k}, A_j, B_k$ denote rows of the corresponding matrices.*

Reduction from the reconstruction problem to attribute non-privacy. Consider a matrix $M \in \{0,1\}^{d \times n}$. Let $M^{(k)} = M \odot M \odot \ldots \odot M$ be the dimension $d^k \times n$ matrix obtained by applying $\odot$ operator $k-1$ times. Let $s \in \{0,1\}^n$. Consider the database $D = (M^\top | s) \in (\{0,1\}^{d+1})^n$. That is, the first d columns of D are given by the rows of M, and the last column is s. Then $\mathcal{C}_k(D)$ contains the vector $M^{(k-1)}s$. This reduction holds for every $M \in \{0,1\}^{d \times n}$, but in the analysis, we use a random matrix M.

Definition 2.3 (Row Non-Privacy). *An algorithm $\mathcal{A}$ for releasing all k-way conjunction predicates is* row 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 $S \subseteq [n]$ whose size is at least $\widetilde{\Omega}(\min\{n, d^k\})$ 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 a constant (say $2/3$);

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 $1 - negl(d)$.

The row non-privacy lower bound 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 as well as its relaxations to metrics on probability distributions such as total variation distance or KL divergence [14, 40, 1, 18].

Reduction from the reconstruction problem to row non-privacy. Consider a matrix $M \in \{0,1\}^{d \times n}$. Consider the database $D = diag(s) \cdot M^\top \in (\{0,1\}^d)^n$, 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 $M_i^\top$ or 0^d, according to the ith bit of s (where M_i is the ith row of M). Then $\mathcal{C}_k(D)$ contains the vector $M^{(k)}s$. Again, this reduction holds for every $M \in \{0,1\}^{d \times n}$, but in the analysis, we use a random matrix M.

2.1 Lower Bounds for the Reconstruction Problem

Let $s = (s_1, \dots, s_n) \in \{0,1\}^n$ be some (secret) vector. Let c_k be a constant (we will define it later in Theorem 2.5). Let $a = \min\{n, c_k d^k / \log^{2k-2} n\}$. Let $s|_a = (s_1, \dots, s_a)$ be the first a entries of s. Let $\Phi \in \{0,1\}^a$ be a vector with independent entries taking values 0 and 1 with probability 1/2 (to simplify the exposition, we shall ignore rounding issues). Let $\Phi_1, \dots, \Phi_d \in \{0,1\}^a$ be d independent copies of Φ. Let M be a $d \times a$ matrix whose rows are $\Phi_1, \dots, \Phi_d$. Again, we consider the matrix $M^{(k)}$ constructed out of M.

The attack works as follows: for every row R in $M^{(k)}$, the adversary asks inner product of R with $s|_a$, and receives noisy responses. Consider any privacy mechanism $\mathcal{A}$. Let $p = \mathcal{A}(M^{(k)} s|_a)$ be the vector of noisy answers generated by $\mathcal{A}$. Define the error (noise) vector as $e = p - M^{(k)} s|_a$. Let $M^{(k)} = P\Gamma Q$ be the singular value decomposition of $M^{(k)}$. Here, P is a $d^k \times d^k$ orthogonal matrix, Γ is a $d^k \times a$ diagonal matrix, and Q is an $a \times a$ orthogonal matrix. Let $\mathbf{0}$ be a $(d^k - a) \times a$ matrix with all entries zero. Define $\Gamma^{-1} = (diag(\sigma_1(M^{(k)})^{-1}, \dots, \sigma_a(M^{(k)})^{-1}) | \mathbf{0}^\top)$. The dimension of Γ^{-1} is $a \times d^k$. Define $M_{inv}^{(k)} = Q^\top \Gamma^{-1} P^\top$.

Given p, the adversary uses $M_{inv}^{(k)}$ to construct $\hat{s} = (\hat{s}_1, \dots, \hat{s}_a)$ as follows: $\hat{s}_i = 1$ if the ith entry in $M_{inv}^{(k)} p \geq 1/2$, and 0 otherwise. Now, the claim is that $\hat{s}$ is a good reconstruction of $s|_a$. The idea behind the analysis is that $M_{inv}^{(k)} p = s|_a + M_{inv}^{(k)} e$, and therefore (as P and Q are orthogonal matrices),

$$\|M_{inv}^{(k)} e\| = \|Q^\top \Gamma^{-1} P^\top e\| = \|\Gamma^{-1} P^\top e\| \leq \|\Gamma^{-1}\|_\infty \|P^\top e\| = \|e\| / \sigma_a(M^{(k)}).$$

Corollary 2.6 shows that with high probability $\sigma_a(M^{(k)}) = \widetilde{\Omega}(\sqrt{d^k})$. If an algorithm only adds $o(\sqrt{n})$ noise to each query (i.e., all the entries in e are $o(\sqrt{n})$) then $\|e\| = o(\sqrt{d^k n})$, and therefore, $\|M_{inv}^{(k)} e\| \approx o(\sqrt{n})$ with high probability. In particular, if $a = n$, then this implies that with high probability $M_{inv}^{(k)} e$ cannot have $\Omega(n)$ entries with absolute value above $1/2$, and therefore, the Hamming distance between $\hat{s}$ and $s|_a$ is $o(a) = o(n)$ (as the adversary only fails to recover those entries of $s|_a$ whose corresponding $M_{inv}^{(k)} e$ entries are greater than $1/2$). The following proposition formalizes this observation. The proof uses some ideas from a recent reconstruction attack proposed by Dwork and Yekhanin [16].

Proposition 2.4 (k-way reconstruction attack)**.** *Let k be a constant. If an algorithm adds*

$$o(\min\{\sqrt{n}/\log^{(k^2+k+1)} n, \sqrt{d^k}/\log^{(k^2+3k-1)} n\})$$

noise to each entry in $M^{(k)} s|_a$, then there exists an adversary that can reconstruct $1 - o(1)$ fraction of $s|_a$ with probability at least $1 - negl(d)$.

Proof. Let $P\Gamma Q$ be the singular value decomposition of $M^{(k)}$. Here, Γ is a diagonal matrix containing the singular values of $M^{(k)}$, and P, Q are orthogonal matrices. P is a $d^k \times d^k$ matrix, Γ is a $d^k \times a$ diagonal matrix, and Q is an $a \times a$ matrix. Let $\Gamma = \binom{G}{\mathbf{0}}$. Here, G is an $a \times a$ diagonal matrix of singular values, and $\mathbf{0}$ is $(d^k - a) \times a$ matrix with all entries zero.

Define a diagonal matrix

$$G^{-1} = diag\left(\frac{1}{\sigma_1(M^{(k)})}, \dots, \frac{1}{\sigma_a(M^{(k)})}\right).$$

Define a matrix $\Gamma^{-1} = (G^{-1}|\mathbf{0}^\top)$. The dimension of Γ^{-1} is $a \times d^k$. Now, $\Gamma^{-1}\Gamma = \mathbb{I}_a$ (identity matrix of dimension $a \times a$). Define a matrix $M_{inv}^{(k)} = Q^\top \Gamma^{-1} P^\top$.

Let $p = M^{(k)} s|_a + e$. Given p, the adversary uses $M_{inv}^{(k)}$ to construct $\hat{s} = (\hat{s}_1, \dots, \hat{s}_a)$ as follows: $\hat{s}_i = 1$ if the ith entry in $M_{inv}^{(k)} p \geq 1/2$, and 0 otherwise. We have $M_{inv}^{(k)} p = s + M_{inv}^{(k)} e$.

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

$$\|M_{inv}^{(k)} e\| \leq \|\Gamma^{-1}\|_\infty \|P^\top e\| = \|\Gamma^{-1}\|_\infty \|e\| = \|G^{-1}\|_\infty \|e\|.$$

We break the reminder of the proof into two cases based on the relationship between d and n.

Case 1: $n \leq c_k d^k / \log^{2k-2} n$**.** In this case $a = n$, and $s|_a = s$. Using the bound from Corollary 2.6 for the least singular value of $M^{(k)}$, implies with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$,

$$\|M_{inv}^{(k)} e\| \leq \|G^{-1}\|_\infty \|e\| = \frac{\|e\| \log^{(k^2+k+1)} n}{c'_k \sqrt{d^k}}.$$

Now if an algorithm adds $o(\sqrt{n} / \log^{(k^2+k+1)} n)$ noise to each entry in $M^{(k)} s$ then

$$\|e\| = o(d^{k/2} \sqrt{n} / \log^{(k^2+k+1)} n).$$

Therefore, $\|M_{inv}^{(k)} e\| = o(\sqrt{n})$ with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$. So, with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$, $M_{inv}^{(k)} e$ cannot have $\Omega(n)$ coordinates with absolute value above $1/2$, and therefore, $d_H(s, \hat{s}) = o(n)$.

Case 2: $n > c_k d^k / \log^{2k-2} n$**.** In this case $a = c_k d^k / \log^{2k-2} n$. As in the previous case,

$$\|M_{inv}^{(k)} e\| \leq \frac{\|e\| \log^{(k^2+k+1)} n}{c'_k \sqrt{d^k}}.$$

Now if an algorithm adds $o(\sqrt{d^k} / \log^{(k^2+3k-1)} n)$ noise to each entry in $M^{(k)} s|_a$ then

$$\|e\| = o(d^k / \log^{(k^2+3k-1)} n).$$

Therefore, $\|M_{inv}^{(k)} e\| = o(\sqrt{a})$ with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$. So, with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$, $M_{inv}^{(k)} e$ cannot have $\Omega(a)$ coordinates with absolute value above $1/2$, and therefore, $d_H(s|_a, \hat{s}) = o(a)$. □

Remark: For $k = 1$, we can improve the bounds in the above proposition. In this case, if an algorithm adds $o(\min\{\sqrt{d}, \sqrt{n}\})$ noise to each entry in $M^{(1)} s|_a (= M s|_a)$ where $a = \min\{n, d/2\}$, then there exists an adversary that can reconstruct $1 - o(1)$ fraction of $s|_a$ with exponentially high probability. The reason being that $M^{(1)} = M$ is a random matrix with independent entries, so in Proposition 2.4, as opposed to, Theorem 2.5, we can use the least singular value bound of a random matrix with independent entries from [35].

Theorem 2.5 (Least Singular Value)**.** *Let k, m, d be natural numbers such that $m \leq c_k d^k / \log^{2k-2} m$ where c_k depends only on k, and let A be a $d \times m$ matrix with independent entries taking values 0 and 1 with probability $1/2$. Then there exists numbers C_k, c'_k which depend only on k such that the k-times entry-wise product $\tilde{A} = A \odot A \odot \ldots \odot A$ is a $d^k \times m$ matrix satisfying*

$$\Pr\left[\sigma_m(\tilde{A}) \leq c'_k \sqrt{d^k} / \log^{(k^2+k+1)} m\right] \leq 2\exp\left(-C_k d / \log^{2k-2} m\right).$$

The complete proof of this theorem is presented in Section 3, and an outline of the proof is given in the Section 3. Now, as M is a $d \times a$ matrix with independent entries taking values 0 and 1 with probability $1/2$, and $a \leq c_k d^k / \log^{2k-2} n$, we can apply the above theorem to conclude the following.

Corollary 2.6. $\Pr[\sigma_a(M^{(k)}) \geq \sqrt{d^k} / \log^{(k^2+k+1)} n] \geq 1 - negl(d)$.

2.2 Lower Bounds on Noise to Avoid Attribute Non-Privacy

We use the reduction from the reconstruction attack described earlier. For a vector $s \in \{0,1\}^n$, define a database $D_{at}(s) \in (\{0,1\}^d)^n$ as follows: the first d columns are $\Phi_1, \ldots, \Phi_{d-1}$, the last column is s, and if $n > a$ then the last $n - a$ entries in each Φ_j ($j \in [d-1]$) are all 0's. The following theorem uses Proposition 2.4 to show that there exists an adversary that can reconstruct $1 - o(1)$ fraction of the first $\widetilde{\Omega}(\min\{n, d^{k-1}\})$ entries of s if given too accurate vector $\mathcal{C}_k(D_{at}(s))$.

Theorem 2.7 (Attribute Non-Privacy)**.** *Let k be a constant. Any algorithm $\mathcal{A}$ for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that for every database $D \in (\{0,1\}^d)^n$ adds*

$$o\left(\min\left\{\sqrt{n}/\log^{(k^2-k+1)} n, \sqrt{d^{k-1}}/\log^{(k^2+k-3)} n\right\}\right)$$

noise to each entry in $\mathcal{A}(D)$ is attribute non-private.

Proof. Let $D_{at}(s)$ be a database with $\Phi_1, \ldots, \Phi_{d-1}$ in its first d columns and s in the last column, and if $n > a$ then the last $n-a$ entries in each Φ_j ($j \in [d-1]$) are all 0's. Let $b = \min\{n, c_{k-1}(d-1)^{k-1}/\log(d-1)\}$ (where c_{k-1} is the constant from Theorem 2.5). Now, $\mathcal{C}_k(D_{at}(s))$ contains all the entries of $M^{(k-1)} s|_b$. Now consider an algorithm $\mathcal{A}$ that releases $\mathcal{C}_k(D_{at}(s))$ with

$$o\left(\min\left\{\sqrt{n}/\log^{(k^2-k+1)} n, \sqrt{d^{k-1}}/\log^{(k^2+k-3)} n\right\}\right)$$

noise to each entry. By Proposition 2.4, there exists an adversary that can reconstruct at least $\widetilde{\Omega}(\min\{n, d^{k-1}\})$ entries of s with probability $1 - negl(d)$. □

2.3 Lower Bounds on Noise to Avoid Row Non-Privacy

Again, we use the reduction from the reconstruction attack described earlier. For a vector $s \in \{0,1\}^n$, define a database $D_{st}(s) \in (\{0,1\}^d)^n$ as follows: (i,j)th entry of $D_{st}(s)$ is s_i if the ith entry in $\Phi_j = 1$ and 0 otherwise, and if $n > a$ then the last $n - a$ entries in each Φ_j ($j \in [d]$) are all 0's. The following lemma is based on the observation that $\mathcal{C}_k(D_{st}(s))$ contains all the entries of $M^{(k)} s|_a$.

Lemma 2.8. *If an algorithm adds $o(\min\{\sqrt{n}/\log^{(k^2+k+1)} n, \sqrt{d^k}/\log^{(k^2+3k-1)} n\})$ noise to each entry in $\mathcal{C}_k(D_{st}(s))$, then there exists an adversary who if given $\Phi_1, \ldots, \Phi_d$ can reconstruct $1 - o(1)$ fraction of the first a rows of $D_{st}(s)$ with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$.*

Proof. We split the proof into two cases.

Case 1: $n \leq c_k d^k / \log^{2k-2} n$**.** $\mathcal{C}_k(D_{st}(s))$ contains all the entries of $M^{(k)} s$. From Case 1 of Proposition 2.4, if an algorithm adds $o(\sqrt{n}/\log^{(k^2+k+1)} n)$ noise to each entry in $\mathcal{C}_k(D_{st}(s))$, then an adversary can reconstruct $1 - o(1)$ fraction of s (and hence, $1 - o(1)$ fraction of the rows of $D_{st}(s)$) with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$.

Case 2: $n > c_k d^k / \log^{2k-2} n$**.** $\mathcal{C}_k(D_{st}(s))$ contains all the entries of $M^{(k)} s|_a$. From Case 2 of Proposition 2.4, if an algorithm adds $o(\sqrt{d^k}/\log^{(k^2+3k-1)} n)$ noise to each entry in $\mathcal{C}_k(D_{st}(s))$, then an adversary can reconstruct $1 - o(1)$ fraction of $s|_a$ (and hence, $1 - o(1)$ fraction of the first a rows of $D_{st}(s)$) with probability at least $1 - 2\exp(-C_k d / \log^{2k-2} n)$. □

Define a distribution $\mathbb{D}$ over the set of databases as follows: draw a vector s_r uniformly at random from $\{0,1\}^n$ and output $D_{st}(s_r)$. Let consider some ith row where $i \in [a]$. 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 $D_{st}(s_r)$ 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 $D_{st}(s_r)$ with probability more than $1/2$. Finally, since $\Pr[\overline{E}] = 1/2^d$, therefore, no adversary (even with access to $\Phi_1, \dots, \Phi_d$) can guess the ith row of $D_{st}(s_r)$ with probability more than $1/2 + 1/2^d \leq 2/3$. Thus, $\mathbb{D}$ satisfies the first condition of Definition 2.3 for every set S. The following theorem uses this distribution $\mathbb{D}$ to obtain a lower bound on noise needed for *not* row non-privacy.

Theorem 2.9 (Row Non-Privacy)**.** *Let k be a constant. Any algorithm for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that for every database $D \in (\{0,1\}^d)^n$ adds*

$$o\left(\min\left\{\sqrt{n}/\log^{(k^2+k+1)} n, \sqrt{d^k}/\log^{(k^2+3k-1)} n\right\}\right)$$

noise to each entry in $\mathcal{A}(D)$ is row non-private.

Proof. Define a distribution $\mathbb{D}$ over the set of databases as follows: draw a vector s_r uniformly at random from $\{0,1\}^n$ and output $D_{st}(s_r)$. As discussed above, $\mathbb{D}$ satisfies the first condition of Definition 2.3. The set S in the Definition 2.3 is $[a]$.

Consider $D_{st}(s_r) \sim \mathbb{D}$. Lemma 2.8 shows that if an algorithm adds

$$o\left(\min\left\{\sqrt{n}/\log^{(k^2+k+1)} n, \sqrt{d^k}/\log^{(k^2+3k-1)} n\right\}\right)$$

noise to each entry of $\mathcal{C}_k(D_{st}(s_r))$ then there exists an adversary that if given access to $\Phi_1, \dots, \Phi_d$ can reconstruct $1 - o(1)$ fraction of the first a rows of $D_{st}(s_r)$ with probability $1 - negl(d)$. □

3 Lower Bounding the Least Singular Value: Proof of Theorem 2.5

In this section, we present the complete proof of Theorem 2.5. Estimating the smallest singular value of the matrix $\tilde{A}$ presents two challenges. The entries of this matrix are interdependent, which makes powerful measure concentration tools hard to apply. Also, the entries are non-centered, and hence its (operator) norm is of order $\sqrt{d^k n}$ with high probability. The norm of the matrix enters many probabilistic bounds involved in the proof, and such big norm would render most of these bounds meaningless. To remove these obstacles, we apply in Section 3.1 a simple decoupling and symmetrization argument to reduce the problem to bounding the smallest singular value of a matrix $\tilde{\Pi}$, which is an entry-wise product of k independent random matrices with centered $\{-1, 0, 1\}$ entries. Analysis of the behavior of the least singular value of such matrix is the core of the argument.

The first step in this analysis is obtaining a probabilistic bound for the norm of this matrix. This bound is proved in Section 3.2 by induction on k, with Talagrand's convex concentration inequality (see [24], Corollary 4.10) applied at each step.

The smallest singular value of $\tilde{\Pi}$ is the minimum of $\|\tilde{\Pi}x\|$, over x from the unit sphere. Before we analyze this quantity in full generality, we consider in Section 3.3 a simpler question of estimating *the small ball probability*. This is the probability that $\|\tilde{\Pi}x\|$ is small for a fixed vector x. Measure concentration plays a prominent role in this estimate as well.

We finish the proof in Section 3.4. Instead of obtaining a uniform lower bound for $\|\tilde{\Pi}x\|$ in one step, we decompose the sphere in numerous regions, and estimate the probability that $\|\tilde{\Pi}x\|$ 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.

We start with obtaining a uniform estimate of $\|\tilde{\Pi}x\|$ over a set of all vectors x having a given level of sparsity. This is done in Lemma 3.13, and the argument essentially depends on how sparse the vectors x are. In Lemma 3.16, we extend the bound from the set of sparse vectors to the set of compressible vectors with a certain level of compressibility. Finally, in Lemma 3.17 we show that the whole sphere can be assembled from these sets. This allows to finish the proof by using the union bound.

Remember that the entry-wise product of vectors $p, q \in \mathbb{R}^n$ is the vector in $p \odot q \in \mathbb{R}^n$ with entries $(p \odot q)_i = p_i \cdot q_i$. If A is an $N_1 \times n$ matrix, and B is an $N_2 \times n$ matrix, $A \odot B$ is an $N_1 N_2 \times n$ matrix, whose rows are entry-wise products of the rows of A and B: $(A \odot B)_{j,k} = A_j \odot B_k$, where $(A \odot B)_{j,k}, A_j, B_k$ denote rows of the corresponding matrices.

We start off by restating Theorem 2.5.

Theorem 3.1 (Theorem 2.5 Restated)**.** *Let K, n, d be natural numbers such that*

$$n \leq \frac{c_K d^K}{\log^{2K-2} n},$$

where c_K depends only on K and let A be an $d \times n$ matrix with independent entries taking values 0 and 1 with probability $1/2$. Then there exists numbers c'_K, C_K that depend only on K such that the K-times entry-wise product $\tilde{A} = A \odot A \odot \ldots \odot A$ is a $d^K \times n$ matrix satisfying

$$\Pr\left[\sigma_n(\tilde{A}) \leq \frac{c'_K \sqrt{d^K}}{\log^{(K^2+K+1)} n}\right] \leq 2 \exp\left(-\frac{C_K d}{\log^{2K-2} n}\right).$$

Note that (for convenience) we changed the notation a bit from Theorem 2.5 (n replaces m and K replaces k). The proof of Theorem 3.1 follows from Theorem 3.4 and is described in Section 3.1. We note some remarks about this proof.

Remark: The powers of $\log n$ can be significantly reduced. Such reduction, however, would make the proof more complicated.

Remark: The same proof with minor modifications works for more general random matrices. namely, it is applicable to a matrix A, whose entries are independent $\{0,1\}$ random variables $a_{i,j}$, taking the value 1 with probabilities $p_{i,j} \in (q_1, q_2)$, where $0 < q_1 < q_2 < 1$. In this case the parameters C_K, c_K, c'_K will depend on q_1 and q_2, as well as on K.

Notations. The Euclidean sphere centered at origin is denoted by S^{n-1}, and by B_2^n we denote the unit Euclidean ball in $\mathbb{R}^n$. We use $e_1, \dots, e_n$ to denote the standard basis in $\mathbb{R}^n$. For a vector $x \in \mathbb{R}^n$, $x(i)$ represents the ith entry of the vector (we use this notation in this proof instead of our usual x_i for convenience). For vectors $x, y \in \mathbb{R}^n$, $y \geq x$ if each entry in y is greater than the corresponding entry in x. Throughout the proof C, c, c', etc. denote absolute constants, whose value may change from line to line. Denote by $\|A\|$ the operator norm of the matrix A (we use this notation in this proof instead of our usual $\|A\|_\infty$ for convenience). Denote by $\|A\|_{HS}$ the Hilbert–Schmidt norm:

$$\|A\|_{HS} = \left(\sum_{j,k} |a_{j,k}|^2 \right)^{1/2}.$$

Consider a subset T of $\mathbb{R}^n$, and let $\alpha > 0$. An 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 3.2 (Bounding the size of an α-Net [31]). *Let T be a subset of S^{n-1} and let $\alpha > 0$. Then there exists an α-net of T of cardinality at most $(1 + 2/\alpha)^n$.*

3.1 Reduction of Theorem 3.1 to a Variant

Definition 3.3 (Γ-random variable). *Let Γ be a random variable taking values 1 and -1 with probability $1/4$, and value 0 with probability $1/2$. We will call a copy of this variable a* Γ-random variable.

Theorem 3.1 follows from the following decoupled version.

Theorem 3.4. *Let K, n, d be natural numbers such that $n \leq \frac{c_K d^K}{\log^{2K-2} n}$ where c_K depends only on K and let $\Pi_1, \dots, \Pi_K$ be $d \times n$ matrices with independent Γ-random entries. Then there exists numbers c'_K, C_K that depend only on K such that the K-times entry-wise product $\tilde{\Pi} = \Pi_1 \odot \Pi_2 \odot \dots \odot \Pi_K$ is a $d^K \times n$ matrix satisfying*

$$\Pr\left[\sigma_n(\tilde{\Pi}) \leq \frac{c'_K \sqrt{d^K}}{\log^{(K^2+K+1)} n} \right] \leq 2 \exp\left(-\frac{C_K d}{\log^{2K-2} n} \right).$$

Over the next few subsections we prove this theorem. Now we use it to derive Theorem 3.1.

Proof of Theorem 3.1. Let d and K be as in Theorem 3.1. Let $d = 2Kd' + m$, where $0 \leq m < 2K$. For $j = 1, \dots, K$ denote by Π_j^1 the submatrix of A consisting of rows $(2d'(j-1)+1), \dots, (2d'(j-1)+d')$, and by Π_J^0 the submatrix consisting or rows $(2d'(j-1)+d'+1), \dots, 2d'j$. Set $\Pi_j = \Pi_j^1 - \Pi_j^0$. Then $\Pi_1, \dots, \Pi_K$ are $d' \times n$ matrices with independent Γ-random entries.

For any $x \in S^{n-1}$

$$\begin{aligned}\|(\Pi_1 \odot \ldots \odot \Pi_K)x\| &\leq \sum_{\alpha=(\alpha_1,\ldots,\alpha_K)\in\{0,1\}^K} \left\|(\Pi_1^{\alpha_j} \odot \ldots \odot \Pi_K^{\alpha_j})x\right\| \\ &\leq 2^K \left\|(A \odot \ldots \odot A)x\right\|,\end{aligned}$$

because the coordinates of $(\Pi_1^{\alpha_j} \odot \ldots \odot \Pi_K^{\alpha_j})x$ form a part of coordinates of $(A \odot \ldots \odot A)x$. Therefore, for any $t > 0$

$$\Pr[\sigma_n(A \odot \ldots \odot A) < t] \leq \Pr[\sigma_n(\Pi_1 \odot \ldots \odot \Pi_K) < 2^K t].$$

To complete the proof we use Theorem 3.4 with d' in place of d, and note that $d \leq 3Kd'$. □

3.2 Norm Estimates

Lemma 3.5. *Let W be an $m \times n$ matrix. Let $\theta \in \mathbb{R}^n$ be a vector with independent Γ-random coordinates. For $l \in \mathbb{N}$, let $Y_1, \ldots, Y_l$ be independent copies of the random variable $Y = \|W\theta\|$. Then for any $s > 0$*

$$\Pr\left[\sum_{j=1}^{l} Y_j^2 \geq 2l \|W\|_{HS}^2 + s\right] \leq 2^l \cdot \exp\left(-\frac{cs}{\|W\|^2}\right).$$

Proof. Note that $F : \mathbb{R}^n \to \mathbb{R},\ F(x) = \|Wx\|$ is a Lipschitz convex function with the Lipschitz constant $\|W\|$. By Talagrand's convex concentration inequality,

$$\Pr[|Y - M| \geq t] \leq 2\exp\left(-\frac{ct^2}{\|W\|^2}\right),$$

where $M = \mathbb{M}(Y)$ is the median of Y. For $j = 1, \ldots, l$ set $Z_j = |Y_j - M|$. Then the previous inequality means that Z_j is a ψ_2 random variable, i.e.

$$\mathbb{E}\left[\exp\left(\frac{c' Z_j^2}{\|W\|^2}\right)\right] \leq 2$$

for some constant $c' < c$. By the Chebychev inequality and independence of $Z_1, \ldots, Z_l$,

$$\Pr\left[\sum_{j=1}^{l} Z_j^2 > t\right] = \Pr\left[\frac{c'}{\|W\|^2}\sum_{j=1}^{l} Z_j^2 > \frac{c't}{\|W\|^2}\right] \leq 2^l \cdot \exp\left(-\frac{c't}{\|W\|^2}\right).$$

Using the elementary inequality $x^2 \leq 2(x-a)^2 + 2a^2$, valid for all $x, a \in \mathbb{R}$, we derive that

$$\Pr\left[\sum_{j=1}^{l} Y_j^2 > 2lM^2 + 2t\right] \leq \Pr\left[\sum_{j=1}^{l} Z_j^2 > t\right] \leq 2^l \cdot \exp\left(-\frac{c't}{\|W\|^2}\right).$$

To finish the proof, notice that $M^2 = \mathbb{M}(Y^2) \leq \mathbb{E}[Y^2] = \|W\|_{HS}^2$. □

To bound the norms of the matrices we construct a subset of the sphere with a special structure.

Lemma 3.6. *Let $\mathcal{N}_0 \subset S^{n-1}$ be the set of all points*

$$x(I, \nu) = \frac{1}{\sqrt{|I|}} \sum_{j \in I} \nu_j e_j,$$

where $e_1, \dots, e_n$ is the standard basis in $\mathbb{R}^n$, I is a non-empty subset of $\{1, \dots, n\}$, and $\nu \in \{-1, 1\}^I$. Then for any linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$

$$\|T\| \le 2\sqrt{\log n} \cdot \max_{x \in \mathcal{N}_0} \|Tx\| .$$

Proof. Assume that $\max_{x \in \mathcal{N}_0} \|Tx\| = M$. Let $x \in S^{n-1}$. Decompose each coordinate of x as follows:

$$x(k) = \sum_{j=1}^{\log_2 n} \frac{x_j(k)}{2^j} + y(k),$$

where $x_1(k), \dots, x_{\log_2 n}(k) \in \{-1, 1, 0\}$ are the first $\log_2 n$ digits of the number $x(k)$ in the dyadic notation, and $|y(k)| \le 1/n$. Set

$$x_j = \sum_{k=1}^{n} \frac{x_j(k)}{2^j} e_k, \quad y = \sum_{k=1}^{n} y(k) e_k.$$

Then $\|y\| < 1/2$, so the set of all vectors $\sum_{j=1}^{\log_2 n} x_j$, where $x \in S^{n-1}$ is a $(1/2)$-net. Hence,

$$\|T\| \le 2 \sup \left(\left\| \sum_{j=1}^{\log_2 n} T x_j \right\| \mid x \in S^{n-1} \right).$$

Since for any j $\|Tx_j\| \le M \|x_j\|$,

$$\left\| \sum_{j=1}^{\log_2 n} T x_j \right\| \le M \sum_{j=1}^{\log_2 n} \|x_j\| \le M \sqrt{\log_2 n} \left(\sum_{j=1}^{\log_2 n} \|x_j\|^2 \right)^{1/2} \le M \sqrt{\log_2 n}.$$

The last inequality follows from

$$\sum_{j=1}^{\log_2 n} \|x_j\|^2 = \sum_{j=1}^{\log_2 n} \sum_{k=1}^{n} \frac{x_j(k)^2}{4^j} \le \|x\|^2 = 1.$$

□

For $k \in \mathbb{N}$ denote by $\mathcal{W}_k$ the set of all $d^k \times n$ matrices V satisfying

$$\|V|_J\| \le C_k \left(d^{k/2} + \sqrt{|J|} \cdot \log^{k/2} n \right) \cdot \log^{(k-1)/2} n. \tag{1}$$

for all non-empty subsets $J \subset \{1, \dots, n\}$. Here $V|_J$ denotes the submatrix of V with columns belonging to J, and C_k is a constant depending on k only.

Lemma 3.7. *Let $d, n, k \in \mathbb{N}$ be numbers satisfying $d \geq \log n$. Let $V_1, \ldots, V_k$ matrices with independent Γ-random entries. Define a $d^k \times n$ matrix W, whose rows are entry-wise products of the rows of $V_1, \ldots, V_k$: $W = V_1 \odot V_2 \odot \ldots \odot V_k$. Then*

$$\Pr[W \notin \mathcal{W}_k] \leq ke^{-cd}.$$

Proof. We use the induction on k.

Step 1. Let $k = 1$. Then $W = V$ is a matrix with independent Γ-random entries. Let $x \in S^{d-1}$, and let $y \in S^{n-1} \cap \mathbb{R}^J$. Then $\langle x, V|_J y\rangle$ is a subgaussian[3] random variable of variance $1/2$. 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^{d-1}, and let $\mathcal{M}$ be a $(1/2)$-net in $S^{n-1} \cap \mathbb{R}^J$. Then

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

The nets $\mathcal{N}$ and $\mathcal{M}$ can be chosen so that $|\mathcal{N}| \leq 6^d$ and $|\mathcal{M}| \leq 6^m$. Combining this with the union bound, we get

$$\Pr[\|V|_J\| \geq 4t] \leq |\mathcal{N}| \cdot |\mathcal{M}| \cdot e^{-ct^2} \leq \exp\left(-ct^2 + (m+d)\log 6\right) \leq e^{-c't^2}$$

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

$$\begin{aligned}\Pr[V \notin \mathcal{W}_1] &\leq \sum_{m=1}^{n} \sum_{|J|=m} \Pr[\|V|_J\| > 4t_m] \leq \sum_{m=1}^{n} n^m e^{-ct_m^2} \\ &\leq \sum_{m=1}^{n} \exp\left[-C\left(\sqrt{d} + \sqrt{m}\sqrt{\log n}\right)^2 + m\log n\right] \leq e^{-cd}.\end{aligned}$$

Step 2. Let $k > 1$, and let $U = V_1 \odot \ldots \odot V_{k-1}$. Assume that $U \in \mathcal{W}_{k-1}$ and condition on U. It is enough to prove that

$$\Pr[U \odot V_k \notin \mathcal{W}_k \mid U] \leq e^{-cd}. \tag{2}$$

Indeed, in this case the induction hypothesis yields

$$\Pr[W \notin \mathcal{W}_k] \leq \Pr[U \odot V_k \notin \mathcal{W}_k \mid U \in \mathcal{W}_{k-1}] + \Pr[U \notin \mathcal{W}_{k-1}] \leq ke^{-cd}.$$

Fix $I \subset \{1, \ldots, n\}$ and ν, and consider the random variable $\|Wx(I,\nu)\|$, where $x(I,\nu)$ was defined in Lemma 3.6. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a row of the matrix V_k. Then the coordinates of the vector $Wx(I,\nu)$ corresponding to this row form the vector

$$(U \odot \alpha)x(I,\nu) = (U \odot x(I,\nu))\alpha.$$

Let U' be the $d^{k-1} \times |I|$ matrix defined as

$$U' = (U \odot x(I,\nu))|_I.$$

[3] 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 $a > 0$.

Since all coordinates of $x(I, \nu)$ have absolute value $1/\sqrt{|I|}$, the assumption $U \in \mathcal{W}_{k-1}$ implies

$$\|U'\| \leq \frac{1}{\sqrt{|I|}} \|U|_I\| \leq \frac{C_{k-1}}{\sqrt{|I|}} \left(d^{(k-1)/2} + \sqrt{|I|} \cdot \log^{(k-1)/2} n \right) \cdot \log^{(k-2)/2} n.$$

Also, since all entries of U have absolute value at most 1,

$$\|U'\|_{HS}^2 \leq d^{k-1}.$$

The sequence of coordinates of the vector $Wx(I, \nu)$ consists of d independent copies of $U'\alpha_I$. Therefore, applying Lemma 3.5 with $l = d$, we get

$$\begin{aligned} p := \Pr[\|Wx(I, \nu)\|^2 \geq 2d \cdot d^{k-1} + s] &\leq 2^d \exp\left(-\frac{cs}{\|U'\|^2} \right) \\ &\leq 2^d \exp\left(-\frac{c_{k-1}s}{\left((d^{k-1}/|I|) + \log^{k-1} n \right) \cdot \log^{k-2} n} \right) \end{aligned} \tag{3}$$

with $c_{k-1} = c \cdot C_{k-1}^{-2}$ depending only on k. Choosing

$$s = (Cd + C|I| \log n) \frac{\left((d^{k-1}/|I|) + \log^{k-1} n \right) \cdot \log^{k-2} n}{c_{k-1}},$$

we get

$$p \leq e^{-cd} \cdot \exp\left(-C|I| \log n \right).$$

Since $d \geq \log n$, we have an estimate

$$2d^k + s \leq C_k''(d^k + |I| \log^k n) \cdot \log^{k-2} n =: s(|I|).$$

The union bound implies

$$\begin{aligned} \Pr[\exists I \subset \{1, \ldots, n\} \ \exists \nu \in \{-1, 1\}^I \text{ such that } \|Wx(I, \nu)\| > s(|I|)] &\leq e^{-cd} \cdot \sum_{I \subset J} 2^{|I|} \exp\left(-C|I| \log n \right) \\ &\leq e^{-c'd}. \end{aligned}$$

Assume now that the complementary event occurs:

$$\forall I \subset \{1, \ldots, n\} \ \forall \nu \in \{-1, 1\}^I \ \ \|Wx(I, \nu)\| \leq s(|I|).$$

Let $J \subset \{1, \ldots, n\}$. By Lemma 3.6,

$$\|W|_J\| \leq 2\sqrt{\log n} \cdot \max_{\substack{I \subset J \\ \Gamma \in \{-1,1\}^I}} \sqrt{s(|I|)} = 2\sqrt{\log n} \cdot \sqrt{s(|J|)},$$

which concludes the induction step. □

3.3 Small Ball Probability - Bounds for the Lévy concentration function

Starting from the works of Lévy [25], Kolmogorov [23], and Esséen [17] 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 3.8. *Let $\rho > 0$. Define the Lévy concentration function of a random vector $X \in \mathbb{R}^n$ by*

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

We will use the following standard lemma.

Lemma 3.9. *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

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

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

Lemma 3.10. *Let $m \in \mathbb{N}$ and $x \in \mathbb{R}^m$ be a vector such that $1 \leq |x(j)| \leq 2$ for all $j = 1, \dots, m$. Let $U = (u_{i,j})$ be any $N \times n$ matrix satisfying*

$$\|U\| \leq \frac{1}{8} \|U\|_{HS}.$$

Let $\alpha = (\alpha_1, \dots, \alpha_n)$ be a vector with independent Γ-random coordinates. Then

$$\mathcal{L}\left((U \odot \alpha)x, \frac{1}{8} \|U\|_{HS}\right) \leq \exp\left(-c\frac{\|U\|_{HS}^2}{\|U\|^2}\right).$$

Proof. The proof of this lemma uses Talagrand's inequality for a convex function, in the same way it was used in the proof of Lemma 3.5. Note that $(U \odot \alpha)x = (U \odot x)\alpha$. Let $\alpha'_1, \dots, \alpha'_n$ be independent copies of $\alpha_1, \dots, \alpha_n$. Applying Lemma 3.9, we obtain

$$\mathcal{L}\left((U \odot x)\alpha, \frac{1}{8}\sqrt{Nm}\right) \leq \sqrt{\Pr\left[\|(U \odot x)(\alpha - \alpha')\| \leq \frac{1}{8}\sqrt{Nm}\right]}. \tag{4}$$

Consider a function $F : \mathbb{R}^{2m} \to \mathbb{R}$, defined by

$$F(yzy) = \|(U \odot x)(y - z)\|,$$

where $y, z \in \mathbb{R}^m$. Then F is a convex function with the Lipschitz constant $L \leq 2\|U \odot x\|$. Note that $U \odot x = U \cdot D$, where D is the diagonal matrix: $D = \text{diag}(x_j)_{j \in \{1,\dots,m\}}$. Hence, $L \leq 4\|U\|$.

By Talagrand's measure concentration theorem for convex functions

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

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

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

By the assumption of the lemma,

$$\Big(\mathbb{E}[F^2]\Big)^{1/2} = \sqrt{2}\, \|(U \odot x)\|_{HS} \ge 4L.$$

Hence,

$$\begin{aligned}\Pr\left[\big\|(U \odot x)(\alpha - \alpha')\big\| \le \frac{1}{8}\, \|U\|_{HS}\right] &\le \Pr\left[|F(\alpha, \alpha') - \mathbb{M}(F)| \ge \frac{1}{4}\Big(\mathbb{E}[F^2]\Big)^{1/2}\right] \\ &\le 2\exp\left(-\frac{c\,\mathbb{E}[F^2]}{L^2}\right) \le \exp\left(-c\frac{\|U\|_{HS}^2}{\|U\|^2}\right).\end{aligned}$$

This inequality and (4) finish the proof. □

For the next result we need the following standard lemma.

Lemma 3.11. *Let $s_1, \dots, s_d$ be independent non-negative random variables such that $\Pr[s_j \le K] \le p$ for all j. Then*

$$\Pr\left[\sum_{j=1}^{d} s_j^2 \le \frac{1}{2}K^2 d\right] \le (4p)^{d/2}.$$

Proof. If $\sum_{j=1}^{d} s_j^2 \le \frac{1}{2}K^2 d$, then $s_j \le K$ for at least $d/2$ numbers j. □

Combining Lemma 3.10 with this corollary, we obtain the tensorized version of Lemma 3.10.

Corollary 3.12. *Let $m \in \mathbb{N}$ and $x \in \mathbb{R}^m$ be a vector such that $1 \le |x(j)| \le 2$ for all $j = 1, \dots, m$. Let $U = (u_{i,j})$ be any $N \times n$ matrix satisfying*

$$\|U\| \le \frac{1}{8}\, \|U\|_{HS}.$$

Let V be a $d \times m$ matrix with independent Γ-random entries. Then

$$\mathcal{L}\left((U \odot V)x, \frac{1}{16}\sqrt{d} \cdot \|U\|_{HS}\right) \le \exp\left(-\frac{cd \cdot \|U\|_{HS}^2}{\|U\|^2}\right).$$

Proof. The coordinates of the vector $(U \odot V)x \in \mathbb{R}^{Nd}$ consist of d independent blocks $(U \odot \alpha_1)x, \dots, (U \odot \alpha_d)x$, where $\alpha_1, \dots, \alpha_d$ are the rows of V. The corollary follows from Lemma 3.11, applied to the random variables $s_j = \|(U \odot \alpha_j)x - y_j\|$, where $y_1, \dots, y_d \in \mathbb{R}^N$ are any fixed vectors. □

We can use this corollary to obtain a small ball probability estimate for a set of vectors with commensurate coordinates. Let B_2^n be the unit L_2-ball.

Lemma 3.13. *Let $m \le l \le n$ be natural numbers, and let $t > 0$. Denote by $S(l, m, t)$ the set of all vectors $u \in B_2^n$ with at most l non-zero coordinates, and at least m coordinates satisfying the inequality*

$$t \le |u(j)| \le 2t.$$

Let U' be an $N \times n$ matrix, and denote

$$\left\|U'\right\|_m := \max_J \left\|U'|_J\right\|, \quad \left\|U'\right\|_{HS(m)} := \min_J \left\|U'|_J\right\|_{HS},$$

where the maximum and minimum are taken over all subsets $J \subset \{1, \dots, n\}$ having m elements. Assume that

$$\left\|U'\right\|_m \le \frac{1}{8}\left\|U'\right\|_{HS(m)}$$

Let V' be a $d \times n$ matrix with independent Γ-random entries. Then

$$\Pr[\exists u \in S(l,m,t) \textit{ such that } \left\|(U' \odot V')u\right\| \le ct\sqrt{d}\cdot\left\|U'\right\|_{HS(m)}] \le \left(\frac{cn\left\|U'\right\|_{HS}}{t}\right)^l \cdot \exp\left(-\frac{cd\cdot\left\|U'\right\|_{HS(m)}^2}{\left\|U'\right\|_m^2}\right).$$

Proof. Let $u \in S(l, m, t)$. Choose an m-element subset $J \subset \{1, \dots, n\}$ such that $t \le |x(j)| \le 2t$ for all $j \in J$. Denote $x = u|_J$, $U = U'|_J$, $V = V'|_J$. Condition on the matrix $V'|_{J^c}$. If $\|(U' \odot V')u\| \le ct\sqrt{d}\cdot\|U'\|_{HS(m)}$, then

$$\|(U \odot V)x - y\| \le ct\sqrt{d}\cdot\left\|U'\right\|_{HS(m)},$$

where the vector $y = (U'|_{J^c} \odot V'|_{J^c})u|_{J^c}$ is fixed after conditioning. Hence, by Corollary 3.12,

$$\begin{aligned}&\Pr[\left\|(U' \odot V')u\right\| \le ct\sqrt{d}\cdot\left\|U'\right\|_{HS(m)} \mid V'|_{J^c}]\\ &\le \mathcal{L}((U \odot V)x, ct\sqrt{d}\cdot\|U\|_{HS}) \le \exp\left(-\frac{cd\cdot\left\|U'\right\|_{HS(m)}^2}{\left\|U'\right\|_m^2}\right),\end{aligned}$$

because $\|U'\|_{HS(m)} \le \|U\|_{HS}$ and $\|U\| \le \|U'\|_m$. Taking the expectation with respect to $V'|_{J^c}$ removes conditioning.

Obviously, $S(l, m, t) \subset S_m := \{u \in B_2^n \mid |\text{supp}(u)| \le l\}$. Set $\alpha = \frac{1}{2}(ct\cdot\|U'\|_{HS})^{-1}$. By the volumetric estimate, there exist a α-net $\mathcal{N} \subset S(l, m, t)$ such that

$$|\mathcal{N}| \le \binom{n}{l}\cdot\left(\frac{3}{\alpha}\right)^l \le \left(\frac{Cn}{\alpha}\right)^l.$$

By the union bound,

$$\Pr[\exists u \in \mathcal{N} \text{ such that } \left\|(U' \odot V')u\right\| \le ct\sqrt{d}\cdot\left\|U'\right\|_{HS(m)}] \le \left(\frac{cn}{t}\right)^l\cdot\exp\left(-\frac{cd\cdot\left\|U'\right\|_{HS(m)}^2}{\left\|U'\right\|_m^2}\right).$$

The result of the lemma follows by approximation. Indeed,

$$\left\|U' \odot V'\right\| \le \left\|U' \odot V'\right\|_{HS} \le \sqrt{d}\left\|U'\right\|_{HS}.$$

Let $v \in S(l,m,t)$ and assume that $\|(U' \odot V')v\| \leq \frac{1}{2}ct\sqrt{d} \cdot \|U'\|_{HS(m)}$. Choose $u \in \mathcal{N}$ such that $\|v-u\| < \alpha$. Then

$$\begin{aligned}\left\|(U' \odot V')u\right\| &\leq \left\|(U' \odot V')v\right\| + \left\|U' \odot V'\right\| \cdot \|u-v\| \\ &\leq \frac{1}{2}ct\sqrt{d} \cdot \left\|U'\right\|_{HS(m)} + \sqrt{d}\left\|U'\right\|_{HS} \alpha \leq ct\sqrt{d} \cdot \left\|U'\right\|_{HS(m)}.\end{aligned}$$

□

The next lemma shows that the row product of Γ matrices contains many rows with a lot of ones.

Lemma 3.14. *Let $\Pi_1, \dots, \Pi_s$ be random $d \times m$ matrices with independent Γ-random entries. The ith row of the matrix Π_j is denoted by $\Pi_j(i)$. Denote by I the set of all $(i_1, \dots, i_s) \in \{1, \dots, d\}^s$ such that*

the vector $\Pi_1(i_1) \odot \dots \odot \Pi_s(i_s)$ contains at least $2^{-s-3}m$ non-zero coordinates.

Then

$$\Pr[|I| \leq 2^{-s-4}d^s] \leq \exp\left(-\frac{c2^{-s}d^s m}{s}\right).$$

Proof. Denote $\Pi = (\Pi_1, \dots, \Pi_s)$. For $(i_1, \dots, i_s) \in \{1, \dots, d\}^s$ denote by $F_{i_1,\dots,i_s}(\Pi)$ the sum of the absolute values of all coordinates of the vector $\Pi_1(i_1) \odot \dots \odot \Pi_s(i_s)$, and set

$$F(\Pi) = \sum_{i_1,\dots,i_s=1}^{d} F_{i_1,\dots,i_s}(\Pi)$$

For given $l \in \{1, \dots, s\}, i \in \{1, \dots, d\}, j \in \{1, \dots, m\}$ let $\Pi'_{i,j}$ be another collection of s Γ matrices, all whose entries, except for $\Pi_l(i,j)$ are equal to the corresponding entries of Π. The entry $\Pi_l(i,j)$ will be an independent Γ-random variable. Then

$$|F(\Pi) - F(\Pi_{l,i,j})| \leq d^{s-1},$$

since the change in one entry of the Π affects at most d^{s-1} entries of $\Pi_1 \odot \dots \odot \Pi_s$. By the bounded differences inequality,

$$\Pr[|F(\Pi) - \mathbb{M}F(\Pi)| > t] \leq 2\exp\left(-\frac{ct^2}{M^2}\right), \tag{5}$$

where

$$M^2 = \sum_{l,i,j} |F(\Pi) - F(\Pi_{l,i,j})|^2 \leq sd^s m.$$

The inequality (5) implies that

$$|\,\mathbb{E}[F(\Pi)] - \mathbb{M}F(\Pi)| \leq cM,$$

and since $\mathbb{E}[F(\Pi)] = 2^{-s}d^s m$, we conclude that $\mathbb{M}F(\Pi) \geq 2^{-s-1}d^s m$. Substituting this in (5), we get

$$\Pr[F(\Gamma) \leq 2^{-s-2}d^s m] \leq \exp\left(-\frac{c2^{-s}d^s m}{s}\right).$$

Assume that Π satisfies $F(\Pi) \geq 2^{-s-2}d^s m$. Since for any $(i_1, \dots, i_s) \in \{1, \dots, d\}^s$, $F_{i_1,\dots,i_s}(\Pi) \leq m$, we obtain

$$\begin{aligned} 2^{-s-2}d^s m \leq F(\Pi) = \sum_{i_1,\dots,i_s=1}^{d} F_{i_1,\dots,i_s}(\Pi) &\leq |I| \cdot m + (d^s - |I|) \cdot 2^{-s-3}m \\ &\leq |I|m + 2^{-s-3}d^s m, \end{aligned}$$

which completes the proof. □

The previous lemma implies a lower bound for the Hilbert-Schmidt norm of $\Pi_1 \odot \dots \odot \Pi_s$.

Corollary 3.15. *Let $\Pi_1, \dots, \Pi_s$ be as in Lemma 3.14. Then*

$$\Pr[\exists J \subset \{1, \dots, m\} \textit{ such that } \|\Pi_1|_J \odot \dots \odot \Pi_s|_J\|_{HS} \leq C_s\sqrt{d^s|J|}] \leq m\exp(-c_s d^s).$$

Proof. Applying Lemma 3.14 with $m = 1$ to each column of the matrix $\Pi_1 \odot \dots \odot \Pi_s$, we get

$$\Pr[\exists j \in \{1, \dots, m\} \text{ such that } \left\|\Pi_1|_{\{j\}} \odot \dots \odot \Pi_s|_{\{j\}}\right\|_{HS} \leq C_s\sqrt{d^s}] \leq m\exp(-c_s d^s),$$

from which the result immediately follows. □

3.4 Assembling the Puzzle: Putting Together the Proof of Theorem 3.4

To prove Theorem 3.4 we use the epsilon-net argument again. However, since the small ball probability for

$$\|(\Pi_1 \odot \dots \odot \Pi_K)x\|$$

depends significantly on the vector x, it is impossible to construct one epsilon-net, which would work for the whole sphere. Instead, we partition the sphere in many pieces, and construct a separate epsilon-net for each piece. For each piece we choose a different value of α, which matches the small ball probability. In Lemma 3.16 we introduce an elementary piece $P(l, m, t, r)$ and prove a lower bound for the norm of $(\Pi_1 \odot \dots \odot \Pi_K)x$, which is valid for the whole $P(l, m, t, r)$ with probability close to 1. Then in Lemma 3.17 we show that S^{n-1} can be covered by a few sets $P(l, m, t, r)$. This fact, combined with the union bound proves Theorem 3.4.

We combine Lemmata 3.7, 3.13, 3.14 to prove the small ball probability estimate for the row product of random Γ matrices.

Lemma 3.16. *Let d, K, k, l, m, and n be natural numbers such that $m \leq l \leq n$, $k \leq K$ and $d \geq \log n$, $\leq d^{C_K}$*

$$l \leq \frac{c_k d^k}{\log^{2k-2} n}, \quad \textit{and} \quad m \geq \frac{c_k l}{d} \cdot \log^{k-1} n. \tag{6}$$

Let $\Pi_1, \dots, \Pi_K$ be $d \times n$ matrices with independent Γ-random entries. Let $1/n \leq t \leq 1$. For

$$0 < r < \frac{C_k t\sqrt{m}}{\log^{(k-1)/2} n}. \tag{7}$$

denote by $P(l, m, t, r)$ the set of all vectors $u \in S^{n-1}$ which can be decomposed as $u = v + w$, where the vector v at most l non-zero coordinates, and at least m coordinates satisfying the inequality

$$t \leq |u(j)| \leq 2t,$$

and $\|w\| \le r$. *Then*

$$\Pr[\exists u \in P(l,m,t,r) \text{ such that } \|(\Pi_1 \odot \ldots \odot \Pi_K)u\| \le c_K t \cdot \sqrt{d^K m}] \le \exp\left(-\frac{c_K d}{\log^{2K-3} n}\right).$$

Proof. To prove the lemma, we rewrite $\Pi_1 \odot \ldots \Pi_K$ as $\Pi \odot \Pi_k \odot \Pi'$, where $\Pi = \Pi_1 \odot \ldots \odot \Pi_{k-1}$, $\Pi' = \Pi_{k+1} \odot \ldots \odot \Pi_K$, and use the independence of these three matrices.

Assume first that $1 < k < K$. Set $s = K - k$. Consider first a vector $v \in S(l,m,t)$, and a $d^s \times n$ matrix

$$D = \Pi_{k+1} \odot \ldots \odot \Pi_K \odot v.$$

Denote its rows by $D_{i_1,\ldots,i_s}$, where $(i_1,\ldots,i_s) \in \{1,\ldots,d\}^s$. Then the coordinates of the vector $(\Pi_1 \odot \ldots \Pi_K)v$ consist of d^s blocks $(\Pi_d \odot \ldots \odot \Pi_k)D_{i_1,\ldots,i_s}$.

Fix an m-element set $J \subset \{1,\ldots,n\}$ for which $t \le |v(j)| < 2t$, whenever $j \in J$. Consider a $d^s \times m$ matrix $\Pi_{k+1}|_J \odot \ldots \odot \Pi_K|_J$. Let I be the set of all rows of Π'' which have at least $2^{-s-3}m$ ones. By Lemma 3.14,

$$\Pr[|I| \le 2^{-s-4}d^s] \le \exp\left(-\frac{c2^{-s}dm}{s}\right).$$

For any $(i_1,\ldots,i_s) \in I$, $D_{i_1,\ldots,i_s} \in S(l, 2^{-s-3}m, t)$. Assume that the matrix $\Pi_{k+1}|_J \odot \ldots \odot \Pi_K|_J$ satisfies $|I| \ge 2^{-s-4}d^s$, and condition on the corresponding matrix $\Pi' = \Pi_{k+1} \odot \ldots \odot \Pi_K$.

Set

$$U' = \Pi_1 \odot \ldots \odot \Pi_{k-1}, \quad V' = \Pi_k.$$

By Lemma 3.7 and Corollary 3.15,

$$\Pr[U' \notin \mathcal{W}_{k-1} \text{ or } \|U'\|_{HS(m)} \le C_k\sqrt{d^{k-1}m}] \le e^{-c_k d}.$$

Fix $U' \in \mathcal{W}_{k-1}$ such that $\|U'\|_{HS(m)} \ge C_k\sqrt{d^{k-1}m}$ and condition on it. Set $m' = \lceil 2^{-s-3}m \rceil$. Since $U' \in \mathcal{W}_{k-1}$,

$$\|U'\|_{m'}^2 \le C_k'\left(d^{k-1} + m \cdot \log^{k-1} n\right) \cdot \log^{k-2} n.$$

We use Lemma 3.13 to bound the small ball probability. Denote by $\Pr_k$ the probability with respect to the random matrix Π_k, when the matrix $U' \in \mathcal{W}_{k-1}$, is fixed. Then

$$\begin{aligned} p &:= \Pr_k[\exists u \in S(l,m',t) \text{ such that } \|(\Pi_1 \odot \ldots \odot \Pi_k)u\| < c_k t\sqrt{d^k m}] \\ &\le \Pr_k[\exists u \in S(l,m',t) \text{ such that } \|(U' \odot V')u\| < C_k t\sqrt{d} \cdot \|U'\|_{HS(m')}] \\ &\le \left(\frac{cn\|U'\|_{HS}}{t}\right)^l \cdot \exp\left(-\frac{c_k d^k m}{\|U'\|_{m'}^2}\right). \end{aligned}$$

By the assumption $t \ge 1/n$, and $\|U'\|_{HS} \le nd^k \le n^{C_K}$, so

$$p \le \exp\left(C_K l \log n - \frac{c_k d^k m}{(d^{k-1} + m \cdot \log^{k-1} n) \cdot \log^{k-2} n}\right).$$

Consider two cases. First, assume that

$$d^{k-1} > m \cdot \log^{k-1} n.$$

Then the previous inequality reads

$$p \le \exp\left(C_K l \log n - \frac{c_k dm}{\log^{k-2} n}\right).$$

The assumption on m implies

$$p \le \exp\left(-\frac{c'_k dm}{\log^{k-2} n}\right).$$

Assume now that

$$d^{k-1} \le m \cdot \log^{k-1} n.$$

Then

$$p \le \exp\left(C_K l \log n - \frac{c_k d^k}{\log^{2k-3} n}\right).$$

The assumptions on d and l imply

$$p \le \exp\left(-\frac{c'_k d^k}{\log^{2k-3} n}\right).$$

Also, by Lemma 3.7,

$$\Pr[\Pi_1 \odot \ldots \odot \Pi_k \notin \mathcal{W}_k] \le ke^{-cd}.$$

Assume now that the matrices $\Pi_, \ldots, \Pi_K$ satisfy the following conditions:

1. $\Pi_1 \odot \ldots \odot \Pi_k \in \mathcal{W}_k$;
2. $\forall u \in S(l, 2^{-s-3}m, t)$ $\|(\Pi_1 \odot \ldots \odot \Pi_k)u\| \ge c_k t\sqrt{d^k m}$
3. Denote by I the set of all $(i_{k+1}, \ldots, i_K) \in \{1, \ldots, d\}^{K-k}$ such that the vector $\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K)$ contains at least $2^{-s-3}m$ non-zero entries. Then $|I| \ge 2^{-s-4}d^s$.

The proof above shows that the probability that one of these conditions is violated is at most

$$\begin{aligned}
&\exp(-c_k d) + \exp\left(-\frac{c'_k dm}{\log^{k-2} n}\right) + \exp\left(-\frac{c'_k d^k}{\log^{2k-3} n}\right) + \exp\left(-c_s dm\right) \\
&\le \exp\left(-\frac{c_K d}{\log^{2K-3} n}\right).
\end{aligned}$$

Condition (2) implies that for any $v \in S(l, m, t)$, and any $(i_{k+1}, \ldots, i_K) \in I$

$$\left\|(\Pi_1 \odot \ldots \odot \Pi_k)\Big(\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K) \odot v\Big)\right\| \ge c_k t\sqrt{d^k m}.$$

Let $u \in P(l, m, t, r)$, and let $u = v + w$ be a decomposition from the definition of this set. Since $\|\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K) \odot w\| \le \|w\| \le r$, we can use condition (1) to show that

$$\begin{aligned}
&\|(\Pi_1 \odot \ldots \odot \Pi_k)(\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K) \odot u)\| \\
&\ge \|(\Pi_1 \odot \ldots \odot \Pi_k)(\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K) \odot v)\| - \|(\Pi_1 \odot \ldots \odot \Pi_k)\| \cdot r \\
&\ge c_k t\sqrt{d^k m} - C_k\left(d^{k/2} + \sqrt{m} \cdot \log^{k/2} n\right) \cdot \log^{(k-1)/2} n \cdot r \\
&\ge \frac{1}{2} c_k t\sqrt{d^k m}.
\end{aligned}$$

The last inequality follows from $m \le l \le \frac{C_k d^k}{\log^{2k-2} n}$, and the assumption on r.

Applying condition (3), we get

$$\|(\Pi_1 \odot \ldots \odot \Pi_K)u\| = \left(\sum_{(i_{k+1},\ldots,i_K)\in\{1,\ldots,d\}^{K-k}} \left\| (\Pi_1 \odot \ldots \odot \Pi_k)\Big(\Pi_{k+1}(i_{k+1}) \odot \ldots \odot \Pi_K(i_K) \odot u\Big) \right\|^2 \right)^{1/2}$$
$$\ge \sqrt{|I|} \cdot \frac{1}{2} c_k t \sqrt{d^k m} \ge c_K t \sqrt{d^K m},$$

which completes the proof for $1 < f < K$.

The cases $k = 1$ and $k = K$ are similar, although simpler. If $k = K$, set $D = v$. Then automatically $D \in S(l, m, t)$, which replaces condition (3).

Assume that $k = 1$. Then $l \le c_1 d$. Making the constant c_1 smaller, if necessary, and using a straightforward epsilon-net argument, we can show that

$$\Pr[\exists u \in S^{n-1} \text{ such that} |\text{supp}(u)| \le l \text{ and } \|\Pi_1 u\| \le c\sqrt{d}\, \|u\|] \le e^{-cd}.$$

Since any $u \in S(l, m, t)$ satisfies $\|u\| \ge t\sqrt{m}$, we conclude that

$$\Pr[\exists u \in S(l, m, t) \text{ such that } \|\Pi_1 u\| \le ct\sqrt{dm}] \le e^{-cd}.$$

The complementary event $\forall u \in S(l, m, t)\ \ \|\Pi_1 u\| \ge ct\sqrt{dm}$ replaces condition (2). □

To complete the proof of Theorem 3.4 we show that the sphere can be covered by a small number of sets $P(l, m, t, r)$, where the parameters l, m, t, and r satisfy conditions (6), and (7).

To this end we need another definition. For $x \in \mathbb{R}^n$ and $l \le n$ denote by $x|_l$ the vector x, whose l coordinates with biggest absolute values are replaced by zeros. Note that by Cauchy–Schwartz inequality,

$$\|x|_l\|_\infty \le l^{-1/2} \|x\| . \tag{8}$$

For $l \le n$ and $r > 0$ define sets $Q(l, r)$ by

$$Q(l, r) = \{x \in S^{n-1} \mid \|x|_l\| \le r\}.$$

This definition is similar to that of a set of compressible vectors, which was introduced in [34]. Note that $Q(n, r) = S^{n-1}$ for any $r > 0$. Let $l_1 \le l_2 \le \ldots \le l_{2K-1} = n$ and $r_1 \ge r_2 \ge \ldots \ge r_{2K-1} > 0$ be any sequences. For convenience set $l_0 = 0,\ r_0 = 1$, and $Q(l_0, r_0) = \emptyset$. Then

$$S^{n-1} = \bigcup_{s=1}^{2K-1} Q(l_s, r_s) \setminus Q(l_{s-1}, r_{s-1}).$$

To prove Theorem 3.4 it is enough to choose the sequences $\{l_s\}_{s=1}^{2K-1}$ and $\{r_s\}_{s=1}^{2K-1}$, so that

$$\Pr\left[\exists x \in Q(l_s, r_s) \setminus Q(l_{s-1}, r_{s-1}) \text{ such that } \|(\Pi_1 \odot \ldots \odot \Pi_K)x\| \le \rho_s\right]$$

is small if $\rho_s > 0$ is appropriately chosen. We make these choices in the following lemma.

Lemma 3.17. *For $s = 1, \dots, (2K-1)$ denote $k_s = \lceil \frac{s+1}{2} \rceil$. Let C_k, c_k be constants from Lemma 3.16. Let*

$$l_s = \frac{c_{k_s} d^{(s+1)/2}}{\log^{2k_s - 2} n}.$$

Set $r_1 = \frac{1}{2} C_1 \log^{-1/2} n$, and define the sequence $\{r_s\}_{s=1}^{2K-1}$ inductively:

$$r_s = \frac{1}{2} C_{k_s} r_{s-1} \cdot \log^{-k_s/2} n.$$

Then

$$\Pr \left[\exists x \in Q(l_s, r_s) \setminus Q(l_{s-1}, r_{s-1}) \text{ such that } \|(\Pi_1 \odot \dots \odot \Pi_K)x\| \le \frac{c_K \sqrt{d^K}}{\log^{(K^2+K+1)} n} \right] \le 2 \exp\left(-\frac{C_K d}{\log^{2K-3} n} \right).$$

The choice of l_s is dictated by the first condition in (6). The choice of r_s is defined by condition (7). The requirement $l_{2K-1} = n$ leads to the bound on n in the formulations of Theorems 3.1 and 3.4.

Proof. Let j_1 be the minimal number such that $2^{-j_1} \le l_{s-1}^{-1/2}$, and let j_2 be the minimal number such that $2^{-j_2} \le l_s^{-1/2}$. Then $j_2 - j_1 \le c \log d \le c \log n$. For $j = j_1, \dots, j_2$ set

$$t_j = 2^{-j}, \quad \text{and } m_j = t_j^{-2} r_{s-1}^2 \log^{-1} n.$$

For $k = k_s$ the first condition in (6) holds for $l = l_s$. Moreover, the inductive definition of r_s implies that $r_s \ge r_{2K} \ge c''_{2K} \log^{-1/2 - K(K+1)} n$. Since $t_j \ge 2^{-j_1} \le \frac{1}{2} l_{s-1}^{-1/2}$,

$$m_j \ge \frac{l_{s-1}}{2} \cdot (c''_{2K})^2 \log^{-1-2K(K+1)} n \ge \frac{l_s}{2d^{1/2}} \cdot (c''_{2K})^2 \log^{-1-2K(K+1)} n.$$

Since by assumption $d \ge n^{1/K}$, this inequality implies the second condition in (6). Also, for any $j \in \{j_1, \dots, j_2\}$

$$r_s < \frac{C_{k_s} t_j \sqrt{m_j}}{\log^{(k_s - 1)/2} n} = C_{k_s} r_{s-1} \cdot \log^{-k_s/2} n.$$

Thus, condition (7) is satisfied for $r = r_s$. By Lemma 3.16,

$$\Pr[\exists u \in P(l_s, m_j, t_j, r_s) \text{ such that } \|(\Pi_1 \odot \dots \odot \Pi_K)u\| \le c_K t_j \cdot \sqrt{d^K m_j}] \le \exp\left(-\frac{c_K d}{\log^{2K-3} n} \right).$$

Taking into account that

$$t_j \sqrt{m_j} = r_{s-1}^2 \log^{-1/2} n \ge r_{2K-1} \log^{-1/2} n \ge \log^{-(K^2+K+1)} n,$$

we obtain

$$\Pr[\exists u \in P(l_s, m_j, t_j, r_s) \text{ such that } \|(\Pi_1 \odot \dots \odot \Pi_K)u\| \le c_K \sqrt{d^K} \log^{-(K^2+K+1)} n] \le \exp\left(-\frac{c_K d}{\log^{2K-3} n} \right).$$

The proof will be complete if we show that

$$Q(l_s, r_s) \setminus Q(l_{s-1}, r_{s-1}) \subset \bigcup_{j=j_1}^{j_2} P(l_s, m_j, t_j, r_s).$$

Let $x \in Q(l_s, r_s) \setminus Q(l_{s-1}, r_{s-1})$. Then

$$\left\| x|_{l_{s-1}} - x|_{l_s} \right\| \geq r_{s-1} - r_s \geq r_{s-1}/2,$$

and by (8), $\left\| x|_{l_{s-1}} \right\|_\infty \leq l_{s-1}^{-1/2}$. For $j = j_1, \ldots, j_2$, let

$$I_j = \{i \in \{1, \ldots, n\} \mid 2^{-j} \leq |x(i)| < 2^{-j+1}\}.$$

Set $y_j = \sum_{i \in I_j} x(i) e_i$. Then $x|_{l_{s-1}} - x|_{l_s} = \sum_{j=j_1}^{j_2} y_j$, so there exists $j \in \{j_1, \ldots, j_2\}$ such that

$$\|y_j\|^2 \geq \frac{r_{s-1}^2}{4(j_2 - j_1)} \geq \frac{r_{s-1}^2}{c \log n}.$$

Hence, $|I_j| \geq 4^j \cdot \|y_j\|^2 \geq m_j$. This shows that $x \in P(l_s, m_j, t_j, r_s)$. □

This completes the proof of Theorem 3.4. □

4 Lower Bounds on Noise for Differential Privacy

In this section, we establish a lower bound on noise needed for releasing k-attribute marginal tables under the notion of (ϵ, δ)-differential privacy. Let D be a database. 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 4.1 ((ϵ, δ)-differential privacy [14, 12]). *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 $\mathcal{A}$. If $\mathcal{A}$ is $(\epsilon, 0)$-differentially private (i.e., $\delta = 0$), then we say it is ϵ-differentially private.*

Differential privacy composes well as shown by the following claim.

Claim 4.2 (Composition and Post-processing [13, 29]). *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,*

$$\mathcal{A}(\mathrm{z}) = G(\mathcal{A}_1(\mathrm{z}), \mathcal{A}_2(\mathrm{z}), \ldots, \mathcal{A}_k(\mathrm{z}))$$

for some probabilistic algorithm G), then $\mathcal{A}$ is $(\sum_{i=1}^k \epsilon_i, \sum_{i=1}^k \delta_i)$-differentially private.

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.,

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

We also use $X \approx_\epsilon Y$ to indicate that random variables X and Y are $(\epsilon, 0)$-indistinguishable.

Our bounds are tight under a natural and popular class of differentially private algorithms based on adding instance-independent noise. This class contains algorithms that for all inputs add noise from a fixed distribution (i.e., the noise distribution is independent of the input). Formally, if an algorithm $\mathcal{A}$ for a function class $\mathcal{F}$ adds instance-independent noise from a distribution Z 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 *et al.* [5] is an example of an algorithm that adds instance-independent noise.

In Section 4.1, we consider (ϵ, δ)-differentially private algorithms for $\mathcal{C}_k$ that add instance-independent noise. For an instance-independent differentially private algorithm $\mathcal{A}$, we can measure the perturbation introduced by $\mathcal{A}$ either by using the mean squared error matrix

$$\Sigma_{\mathcal{A}}(D) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{C}_k(D))(\mathcal{A}(D) - \mathcal{C}_k(D))^\top]$$

or the covariance matrix of $\mathcal{A}(D)$, and the results are the same with either choice. Define the average mean squared error of $\mathcal{A}(D)$ as the trace of $\Sigma_{\mathcal{A}}(D)$ divided by the size of $\mathcal{A}(D)$. Let $m_k = \binom{d}{k}$. We show if for every database D, $\mathcal{A}(D)$ has an average mean squared error (or variance) of $o(m_k(1-\delta/\epsilon)^2/(2^{2k}\epsilon^2))$, then $\mathcal{A}$ is *not* (ϵ, δ)-differentially private. To do so, we analyze projections onto various directions. The idea is to show that for any neighboring databases D and D' with $\mathcal{C}_k(D') - \mathcal{C}_k(D) = \Delta$, the indistinguishability requirement of differential privacy forces both the expected squared length of the projection of $\mathcal{A}(D) - \mathcal{C}_k(D)$ on Δ and the expected squared length of the projection of $\mathcal{A}(D') - \mathcal{C}_k(D')$ on Δ to be at least square of the length of Δ. Of particular interest to us are the direction vectors Δ's with large lengths (close to the largest possible length of $\sqrt{m_k}$). 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 $m_k^2(1-\delta/\epsilon)^2/(2^k\epsilon^2)$. The result follows by dividing the trace by the size of $\mathcal{A}(D^*)$.

In Section 4.2, we consider general (ϵ, δ)-differentially private algorithms for $\mathcal{C}_k$. For a "general" differentially private algorithm $\mathcal{A}$, we *need*[4] to use the mean squared error matrix to measure the perturbation. We show that if for every database D, $\mathcal{A}(D)$ has an average mean squared error of $o(\min\{m_k(1-\delta/\epsilon)^2/(2^{2k}\epsilon^2), n(1-\delta/\epsilon)^2/(2^{2k}\epsilon\log m_k)\})$, then $\mathcal{A}$ is *not* (ϵ, δ)-differentially private. Again for neighboring databases D and D' with $\mathcal{C}_k(D') - \mathcal{C}_k(D) = \Delta$, we investigate the expected squared length of projections of $\mathcal{A}(D) - \mathcal{C}_k(D)$ and $\mathcal{A}(D') - \mathcal{C}_k(D')$ on Δ. The analysis of the general case is harder, because now the indistinguishability requirement forces only one among these two projection lengths to be greater than squared length of Δ. Our proof looks at random databases and shows that for a random database D_r with high probability the trace of $\Sigma_{\mathcal{A}}(D_r)$ is at least $\min\{m_k^2(1-\delta/\epsilon)^2/(2^k\epsilon^2), nm_k(1-\delta/\epsilon)^2/(2^k\epsilon\log m_k)\}$.

In our analysis, (for simplicity) instead of conjunctions, we consider inner products over the domain $\{-1,1\}^d$. 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 $v_i = 0$). Similar to $\mathcal{C}_k$, let $\mathcal{I}_k$ be the class of all k-way inner product predicates. Let D be a database from $(\{-1,1\}^d)^n$. For a predicate $i_v \in \mathcal{I}_k$, define $i_v(D) = \sum_{x \in D} i_v(x)$. Let $\mathcal{I}_k(D)$ be the vector of all predicates in $\mathcal{I}_k$ evaluated on D. In Appendix C, we provide the relationship between releasing conjunctions and inner products.

4.1 Lower Bounds for the Instance-independent Additive Case

For simplicity, we set $\delta = 0$ in the following analysis (see the proof of Theorem 4.9 for $\delta > 0$). We start by proving a very useful property about differential privacy. We state the lemma in terms of a general function class $\mathcal{F}$ and later use if for our specific function class $\mathcal{I}_k$. Let $\mathcal{A}$ be an ϵ-differentially private algorithm for $\mathcal{F}$ that adds instance-independent noise. The lemma shows that both $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{F}(D), \Delta \rangle^2]$ and $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{F}(D'), \Delta \rangle^2]$ are $\Omega(\langle \Delta, \Delta \rangle^2/\epsilon^2)$ where $\Delta = \mathcal{F}(D') - \mathcal{F}(D)$. The proof uses the fact that projections onto direction Δ need to be ϵ-indistinguishable for $\mathcal{A}(D)$ and $\mathcal{A}(D')$.

[4] This is because an algorithm could always add noise such that the output released is always a 0 vector for every database. This is clearly *not* a good algorithm as the deviation from the true answer could be quite big. But this algorithm clearly satisfies all the privacy requirements and also the variance in each coordinate of $\mathcal{A}(D)$ is 0.

Lemma 4.3. *Let $\mathcal{F}$ be a function class of Boolean predicates, and let $\mathcal{A}$ be an ϵ-differentially private algorithm for $\mathcal{F}$ that adds instance-independent noise. Let $\mathcal{A}(D) = \mathcal{F}(D) + Z$. Let $\Delta = \mathcal{F}(D') - \mathcal{F}(D)$, and let $\mathcal{A}(D) \approx_\epsilon \mathcal{A}(D') = \mathcal{A}(D) + \Delta$. Then,*

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

Proof. Since, we are measuring expected squared loss, we can assume without loss of generality that each coordinate of the noise distribution Z of $\mathcal{A}$ is centered at 0 (otherwise, we can shift the noise distribution to satisfy this property without increasing the mean squared error). Therefore, we can assume without loss of generality that for all databases D, $\mathbb{E}[\mathcal{A}(D)] = \mathcal{F}(D)$. Hence, in the instance-independent case for all databases D the mean-squared error matrix is same as the covariance matrix, i.e.,

$$\Sigma_{\mathcal{A}}(D) = \mathbb{E}[(\mathcal{A}(D) - \mathcal{F}(D))(\mathcal{A}(D) - \mathcal{F}(D))^\top] = \mathbb{E}[(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])(\mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)])^\top] = \mathbb{E}[ZZ^\top].$$

We start by proving a simplification of the lemma that illustrates the key ideas.

Lemma 4.4. *Let X' and Y' be two random variables over $\mathbb{R}$. Let $X' \approx_\epsilon Y'$. Let $\mathbb{E}[X'] = p$ and $\mathbb{E}[Y'] = p + 1$. Then, $\mathbb{E}[X'^2] = \Omega(1/\epsilon^2) + p^2$ and $\mathbb{E}[Y'^2] = \Omega(1/\epsilon^2) + p^2$. Therefore, $Var[X'] = \Omega(1/\epsilon^2)$ and $Var[Y'] = \Omega(1/\epsilon^2)$.*

Proof of Lemma 4.4. Define $X = X' - p$ and $Y = Y' - p$. 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 the differential privacy, we have $e^{-\epsilon} b_i \leq a_i \leq e^\epsilon b_i$ 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, from the above inequality 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

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

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

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

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

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

Now, define $\Pr[X_+ = z] = \Pr[X = z \,|\, X \geq 1]$. 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,

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

In particular,

$$\mathbb{E}[X_+^2] \leq \frac{\mathbb{E}[X^2]}{\Pr[X \geq 1]}.$$

Similarly, we get that

$$\mathbb{E}[Y_-^2] \leq \frac{\mathbb{E}[Y^2]}{\Pr[Y \leq 0]}.$$

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 X' and Y' completes the proof of Lemma 4.4. □

We now extend Lemma 4.4. The idea is as follows: Define

$$X = \langle \mathcal{A}(D) - \mathbb{E}[\mathcal{A}(D)], \Delta \rangle = \langle \mathcal{A}(D) - \mathcal{F}(D), \Delta \rangle, \text{ and}$$
$$Y = \langle \mathcal{A}(D') - \mathbb{E}[\mathcal{A}(D)], \Delta \rangle = \langle \mathcal{A}(D') - \mathcal{F}(D), \Delta \rangle.$$

Repeating same arguments as in the previous lemma, we get that

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

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

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

As in Lemma 4.4, we define random variables Y_- and X_+. By arguments similar to Lemma 4.4 we can show that

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

As in Lemma 4.4,

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

Therefore, we now get that

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

As in Lemma 4.4, we can argue that $\mathbb{E}[X^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2)$ and $\mathbb{E}[Y^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2)$. Therefore,

$$\mathbb{E}[\langle\mathcal{A}(D) - \mathcal{F}(D), \Delta\rangle^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2) \quad \text{and} \quad E[\langle\mathcal{A}(D') - \mathcal{F}(D), \Delta\rangle^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2).$$

Also,

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

The last line follows because $\mathbb{E}[Y^2] = \Omega(\langle\Delta,\Delta\rangle^2/\epsilon^2)$. □

One-way Inner products. If $k = 1$ (i.e., 1-way inner products), then the remaining analysis is quite simple. Let D_e be any database which has at least a row of both $(-1)^d$ and $(1)^d$. Consider a vector $\Delta \in \{-2,2\}^d$, construct D_Δ from D_e by replacing the row $(-1)^d$ is replaced by $\Delta/2$ and the row $(1)^d$ by $\Delta/2$. The Hamming distance between D_e and D_Δ is 2 and $\mathcal{I}_1(D_\Delta) - \mathcal{I}_1(D_e) = \Delta$. Also, $\langle\Delta,\Delta\rangle^2 = 16d^2$. We use the above construction to create for every $\Delta = \{-2,2\}^d$ a corresponding database D_Δ. The idea now is to use the fact that this set of Δ's (which contains every vector from $\{-2,2\}^d$) contains an orthogonal (Hadamard) basis $\Delta_1,\ldots,\Delta_d$, and therefore, by invoking Lemma 4.3 for $\mathcal{I}_1$, and noting that $\sum_{i=1}^d \Delta_i\Delta_i^\top = 4d \cdot \mathbb{I}_d$, we have

$$\begin{aligned}
\Omega\left(\frac{d^3}{\epsilon^2}\right) &= \sum_{\Delta_i} tr(\mathbb{E}[\langle\mathcal{A}(D_e) - \mathcal{I}_1(D_e), \Delta_i\rangle^2]) = \sum_{\Delta_i} tr(\Delta_i^\top \Sigma_{\mathcal{A}}(D_e)\Delta_i) \\
&= \sum_{\Delta_i} tr(\Sigma_{\mathcal{A}}(D_e)\Delta_i\Delta_i^\top) = tr(\Sigma_{\mathcal{A}}(D_e) \cdot 4d \cdot \mathbb{I}_d) = 4d \cdot tr(\Sigma_{\mathcal{A}}(D_e)).
\end{aligned}$$

Proposition 4.5 (Instance-independent additive case: 1-way inner products)**.** *Let $\mathcal{A} : (\{-1,1\}^d)^n \to \mathbb{R}^d$ be an ϵ-differentially private algorithm for $\mathcal{I}_1$ that adds instance-independent noise. Let D_e be any database which has at least a row of both $(-1)^d$ and $(1)^d$. Then, $tr(\Sigma_{\mathcal{A}}(D_e)) = \Omega(d^2/\epsilon^2)$.*

Proof. Consider any vector $\Delta \in \{-2,2\}^d$. Since $\mathcal{A}$ is differentially private, by Claim 4.2, $\mathcal{A}(D_e) \approx_{2\epsilon} \mathcal{A}(D_\Delta)$ (where D_Δ is as defined above).

From Lemma 4.3[5], we know

$$\mathbb{E}[\langle \mathcal{A}(D_e) - \mathcal{I}_1(D_e), \Delta \rangle^2] = \Omega(\langle \Delta, \Delta \rangle^2/\epsilon^2) = \Omega(d^2/\epsilon^2),$$

therefore, $\mathbb{E}[\langle \mathcal{A}(D_e) - \mathcal{I}_1(D_e), u_\Delta \rangle^2] = \Omega(d/\epsilon^2)$ (where u_Δ is the unit vector corresponding to Δ). This holds for every $\Delta \in \{-2,2\}^d$. Consider an orthonormal basis $u_1, \dots, u_d$ (one such example is the Hadamard basis) such that $u_i^\top \Sigma_{\mathcal{A}}(D_e) u_i = \Omega(d/\epsilon^2)$ for all $i \in [d]$. Identity matrix $\mathbb{I}_d = \sum_{i=1}^d u_i u_i^\top$. Now,

$$\begin{aligned}
\Omega\left(\frac{d^2}{\epsilon^2}\right) &= \sum_{i=1}^d tr(\mathbb{E}[\langle \mathcal{A}(D_e) - \mathcal{I}_1(D_e), u_i \rangle^2]) \\
&= \sum_{i=1}^d tr(\mathbb{E}[u_i^\top (\mathcal{A}(D_e) - \mathcal{I}_1(D_e))(\mathcal{A}(D_e) - \mathcal{I}_1(D_e))^\top u_i]) \\
&= \sum_{i=1}^d tr(u_i^\top \Sigma_{\mathcal{A}}(D_e) u_i) = \sum_{i=1}^d tr(\Sigma_{\mathcal{A}}(D_e) u_i u_i^\top) \\
&= tr(\Sigma_{\mathcal{A}}(D_e) \cdot \mathbb{I}_d) = tr(\Sigma_{\mathcal{A}}(D_e)).
\end{aligned}$$

The second equality $tr(u_i^\top \Sigma_{\mathcal{A}}(D_e) u_i) = tr(\Sigma_{\mathcal{A}}(D_e) u_i u_i^\top)$ holds because trace is cyclically invariant. This implies that $tr(\Sigma_{\mathcal{A}}(D_e)) = \Omega(d^2/\epsilon^2)$. □

Extension to the k-way inner products. The analysis for $k > 1$ is trickier, as we don't get a set of orthogonal Δ vectors. Here, we start with a special database $D_c = ((1)^d)^n$ (a database of all 1's), and look at the neighbors of D_c obtained by replacing a row of D_c by a vector from $\{-1,1\}^d$. Let D'_c be a neighbor of D_c, and let $\tilde{z} = \mathcal{I}_k(D'_c) - \mathcal{I}_k(D_c)$. Assume that D'_c is obtained from D_c by replacing the jth row of D_c by a vector $d'_c \in \{-1,1\}^d$. Therefore, $\tilde{z} = (n)^{m_k} - ((n-1)^{m_k} + \mathcal{I}_k(d'_c)) = (1)^{m_k} - \mathcal{I}_k(d'_c)$ has lots of 0 entries making $\|\tilde{z}\|$ "small". This is true for many different choices of D'_c. To overcome this problem we analyze projections of $\mathcal{A}(D) - \mathcal{I}_k(D)$ onto direction $\pi\tilde{z}$ where $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o \rangle$ ($\pi\tilde{z}$ is the orthogonal projection of $\tilde{z}$ onto the orthogonal complement of $o = (1)^{m_k}$). These kind of projections have the advantage that $\|\pi\tilde{z}\| = \|\pi \cdot \mathcal{I}_k(d'_c)\|$ is "big" with probability at least $1/2$ over random choices of $\tilde{z}$. The idea now is to use Lemma 4.3 to show that $\mathbb{E}[\langle \mathcal{A}(D_c) - \mathcal{I}_k(D_c), \pi\tilde{z} \rangle^2] = \Omega(\langle \pi\tilde{z}, \pi\tilde{z} \rangle^2/\epsilon^2)$.

Lemma 4.6. *Let $\mathcal{A}$ be an ϵ-differentially private algorithm for $\mathcal{I}_k$ that adds instance-independent noise. Let $D_c = ((1)^d)^n$. Let $D'_c \in (\{-1,1\}^d)^n$ be a neighbor of D_c. Let $\tilde{z} = \mathcal{I}_k(D'_c) - \mathcal{I}_k(D_c)$, $z = \mathcal{I}_k(D'_c) - (n-1)^{m_k}$, and $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o \rangle$ where $o = (1)^{m_k}$. Then,*

$$\mathbb{E}[\langle \mathcal{A}(D_c) - \mathcal{I}_k(D_c), \pi\tilde{z} \rangle^2] = \mathbb{E}[\langle \mathcal{A}(D_c) - \mathcal{I}_k(D_c), \pi z \rangle^2] = \Omega(\langle \pi\tilde{z}, \pi\tilde{z} \rangle^2/\epsilon^2) = \Omega(\langle \pi z, \pi z \rangle^2/\epsilon^2).$$

Proof. Arguments similar to Lemma 4.3 shows that $(\pi\tilde{z})^\top \Sigma_{\mathcal{A}}(D_c)(\pi\tilde{z}) = \Omega(\langle \tilde{z}, \pi\tilde{z} \rangle^2/\epsilon^2) = \Omega(\langle \pi\tilde{z}, \pi\tilde{z} \rangle^2/\epsilon^2)$. Since $\pi\tilde{z} = \pi z$, we get the desired result. □

[5] Substitute D_Δ for D' in Lemma 4.3. As the Hamming distance between D_Δ and D_e is two, ϵ gets replaced by 2ϵ.

In the proof, we use random directions z. The following lemma analyzes the structure of $\mathbb{E}_z[zz^\top]$ (where the randomness is over the choice of z).

Lemma 4.7. *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_r of length m_k as $z_r = \mathcal{I}_k(r)$. Define a matrix $B = \mathbb{E}_{z_r}[z_r z_r^\top]$ where the randomness is over z_r. Then, $B = \mathbb{I}_{m_k}$ where $\mathbb{I}_{m_k}$ is an identity matrix of dimension m_k.*

Proof. We prove the lemma for $k = 2$ (2-way inner products). The proofs for higher k's follow similarly. Let r_i denote the ith entry in 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). Now, $z_r = (z_{1,1}, z_{1,2}, \ldots, z_{d-1,d})$ where $z_{i,j} = r_i r_j$. We now show that for $e, f, g, h \in [d]$, $e \neq f$, and $g \neq h$,

$$\mathop{\mathbb{E}}_{z_r}[z_{e,f} z_{g,h}] = \begin{cases} 1 \text{ if } \{e,f\} = \{g,h\}, \\ 0 \text{ otherwise.} \end{cases}$$

Note that if $\{e,f\} = \{g,h\}$, then $\mathbb{E}_{z_r}[z_{e,f} z_{g,h}] = \mathbb{E}_{z_r}[z_{e,f}^2] = 1$. If $\{e,f\} \neq \{g,h\}$, then there are three cases: if e, f, g, h are all disjoint then $\mathbb{E}_{z_r}[z_{e,f} z_{g,h}] = \mathbb{E}_{z_r}[z_{e,f}]\,\mathbb{E}_{z_r}[z_{g,h}] = 0$, if $e = g$ then $\mathbb{E}_{z_r}[z_{e,f} z_{g,h}] = \mathbb{E}_{z_r}[(r_e^2)(r_f)(r_h)] = \mathbb{E}_{z_r}[r_e^2]\,\mathbb{E}_{z_r}[r_f]\,\mathbb{E}_{z_r}[r_h] = 0$, end if $f = h$ then $\mathbb{E}_{z_r}[z_{e,f} z_{g,h}] = \mathbb{E}_{z_r}[r_e]\,\mathbb{E}_{z_r}[r_g]\,\mathbb{E}_{z_r}[r_h^2] = 0$. Therefore, $\mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_2}$. □

The following proposition uses Lemmata 4.6 and 4.7 to show that there exists a database D_c such that every instance-independent differentially private algorithm needs to add a lot of noise to $\mathcal{I}_k(D_c)$.

Proposition 4.8 (Instance-independent additive case: k-way inner products)**.** *Let $\mathcal{A} : (\{-1,1\}^d)^n \to \mathbb{R}^{m_k}$ be an ϵ-differentially private algorithm for $\mathcal{I}_k$ that adds instance-independent noise. Let $D_c = ((1)^d)^n$. Then, $tr(\Sigma_{\mathcal{A}}(D_c)) = \Omega(m_k^2/\epsilon^2)$.*

Proof. Let $D_c = ((1)^d)^n$. The set of vectors $z = \mathcal{I}_k(D'_c)$ is exactly $\mathrm{supp}(z_r)$ (where z_r is defined in Lemma 4.7 and $\mathrm{supp}(z_r)$ denotes the support of z_r). With expectation over random z_r,

$$\begin{aligned}
\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma_{\mathcal{A}}(D_c)(\pi z_r)] &= \mathop{\mathbb{E}}_{z_r}[tr((\pi z_r)^\top \Sigma_{\mathcal{A}}(D_c)(\pi z_r))] \\
&= \mathop{\mathbb{E}}_{z_r}[tr(\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi z_r z_r^\top)] \\
&= tr(\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi B). \qquad \text{(since trace and expectation commute)}
\end{aligned}$$

From Lemma 4.7, $B = \mathbb{E}_{z_r}[z_r z_r^\top] = \mathbb{I}_{m_k}$. Note that $\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi$ and $B = \mathbb{I}_{m_k}$ are both positive semidefinite. Therefore, $tr(\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi B) \leq tr(\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi)\|B\|_\infty$. Also, since π is an orthogonal projection matrix

$$tr(\pi^\top \Sigma_{\mathcal{A}}(D_c)\pi) = tr(\Sigma_{\mathcal{A}}(D_c)\pi) \leq tr(\Sigma_{\mathcal{A}}(D_c))\|\pi\|_\infty = tr(\Sigma_{\mathcal{A}}(D_c)).$$

Therefore,

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

From Lemma 4.6, $\forall z \in \mathrm{supp}(z_r)$,

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

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 $z \in \text{supp}(z_r)$, $z^\top \pi z \leq z^\top z = m_k$. 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

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

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

$$\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma_{\mathcal{A}}(D_c)(\pi z_r)] = \mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma_{\mathcal{A}}(D_c)(\pi z_r) \mid z_r^\top \pi z_r \geq m_k - 2] \Pr[z_r^\top \pi z_r \geq m_k - 2] +$$
$$\mathop{\mathbb{E}}_{z_r}[(\pi z_r)^\top \Sigma_{\mathcal{A}}(D_c)(\pi z_r) \mid z_r^\top \pi z_r \leq m_k - 2] \Pr[z_r^\top \pi z_r \leq m_k - 2].$$

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

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

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

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

Theorem 4.9 (Instance-independent additive case)**.** *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that adds instance-independent noise and that for every database $D \in (\{0,1\}^d)^n$ has a root mean squared error (or standard deviation) of $o(\sqrt{m_k}(1-\delta/\epsilon)/(2^k\epsilon))$ for each entry of $\mathcal{A}(D)$ is* not *(ϵ, δ)-differentially private.*

Proof. In the case of ϵ-differential privacy ($\delta = 0$), Proposition 4.8 and Corollary C.2 put together give the claimed result. If $\delta > 0$, we use the following lemma which generalizes Lemma 4.6.

Lemma 4.10. *Let $\mathcal{A}$ be a $(1/2, \delta)$-differentially private algorithm for $\mathcal{I}_k$ that adds instance-independent noise. Let $D_c = ((1)^d)^n$. Let $D'_c \in (\{-1,1\}^d)^n$ be a neighbor of D_c. Let $z = \mathcal{I}_k(D'_c) - (n-1)^{m_k}$ and $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o\rangle$. Then, $\mathbb{E}[\langle \mathcal{A}(D_c) - \mathcal{I}_k(D_c), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2)$.*

Proof. Let $\tilde{z} = \mathcal{I}_k(D'_c) - \mathcal{I}_k(D_c)$. In Lemma 4.12, we set $X = \langle \mathcal{A}(D_c) - \mathcal{I}_k(D_c), \pi\tilde{z}\rangle$, $Y = \langle \mathcal{A}(D'_c) - \mathcal{I}_k(D_c), \pi\tilde{z}\rangle$, and $a = \langle \pi\tilde{z}, \pi\tilde{z}\rangle$. Note that $Y = X + a$ and $\pi z = \pi\tilde{z}$. □

Now, using the above lemma (instead of Lemma 4.6) in the proof of Proposition 4.8 shows that $tr(\Sigma_{\mathcal{A}}(D_c)) = \Omega(m_k^2(1-\delta)^2)$. Using the trick explained in Section 4.3 we can introduce ϵ into the lower bound. Finally, using Corollary C.2 to convert the result about inner products to conjunctions proves the claim. Remember that $|\mathcal{C}_k| = 2^k\binom{d}{k}$, whereas $|\mathcal{I}_k| = \binom{d}{k}$. Therefore, we can establish that any algorithm $\mathcal{A}$ for $\mathcal{C}_k$ that adds instance-independent noise and that for every database D has an average mean squared error of $o(m_k(1-\delta/\epsilon)^2/(2^{2k}\epsilon^2))$ for $\mathcal{A}(D)$ is not (ϵ, δ)-differentially private. By taking a square root we get the claimed result. □

4.2 Lower Bounds for the General Case

Again our analysis looks at the related problem of releasing inner products. We initially prove the lower bound by fixing ϵ to $1/2$. We start by proving an extension of Lemma 4.3 to general differentially private algorithms. Let $\mathcal{F}$ be a function class, and let $\mathcal{A}$ be an (ϵ, δ)-differentially private algorithm for $\mathcal{F}$. Let $\mathcal{A}(D) \approx_{1/2,\delta} \mathcal{A}(D')$, and let $\Delta = \mathcal{F}(D') - \mathcal{F}(D)$. Unlike in the instance-independent case (Lemma 4.3) both $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{F}(D), \Delta \rangle^2]$ and $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{F}(D'), \Delta \rangle^2]$ needn't be $\|\Delta\|^4$, but the following lemma shows that at least one of them is $\|\Delta\|^4$ (this one-sided behavior is in fact unavoidable and is explained in Appendix D).

Lemma 4.11. *Let $\mathcal{F}$ be a function class of Boolean predicates, and let $\mathcal{A}$ be a $(1/2, \delta)$-differentially private algorithm for $\mathcal{F}$. Let $\mathcal{A}(D) \approx_{1/2,\delta} \mathcal{A}(D')$. Let $\Delta = \mathcal{F}(D') - \mathcal{F}(D)$. Then, at least one of $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{F}(D), \Delta \rangle^2]$ or $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{F}(D'), \Delta \rangle^2]$ is $\Omega(\langle \Delta, \Delta \rangle^2 (1-\delta)^2)$.*

Proof. Lemma 4.11 follows from setting $X = \langle \mathcal{A}(D) - \mathcal{F}(D), \Delta \rangle$, $Y = \langle \mathcal{A}(D') - \mathcal{F}(D), \Delta \rangle$, and $a = \Delta^\top \Delta$ in the following lemma. If two random variables, X and Y are $(1/2, \delta)$-indistinguishable, then the statistical difference[6] between X and Y is at most $e^{1/2} - 1 + \delta$.

Lemma 4.12 (Lemma 4.11, 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 [39] 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 $a = 0$. □

This completes the proof of Lemma 4.11. □

For a database $D \in (\{-1,1\}^d)^n$, consider the n neighboring databases[7] $\widetilde{D}_1, \ldots, \widetilde{D}_n$ where $\widetilde{D}_i$ is obtained by replacing ith row of D by $(1)^d$. 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 $z_i = o - \tilde{z}_i$. Let $\pi = \mathbb{I}_{m_k} - oo^\top / \langle o, o \rangle$ be an orthogonal projection matrix. Notice that, $\pi \tilde{z}_i = -\pi z_i$. For the reasons same as in the instance-independent case, we analyze projections onto $\pi \tilde{z}_i$ (or equivalently $-\pi z_i$). Let u_{z_i} be the unit vector corresponding to z_i. Define

$$S_k(D) = \{ z \in T_k(D) \mid \mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z \rangle^2] = \Omega(m_k^2 (1-\delta)^2) \},$$
$$U_k(D) = \textstyle\sum_{z \in S_k(D)} u_z u_z^\top \quad \text{and} \quad V_k(D) = \textstyle\sum_{z \in T_k(D)} u_z u_z^\top.$$

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

[7] 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.

Because of the Lemma 4.11 for a database D it is possible that the expected squared length of projection of $\mathcal{A}(D) - \mathcal{I}_k(D)$ onto $\mathcal{I}_k(D') - \mathcal{I}_k(D)$ is small (i.e., $o(\|\mathcal{I}_k(D') - \mathcal{I}_k(D)\|^2)$) for all neighbors D' of D. To overcome this problem we use random databases. Let D_r be a database drawn uniformly at random from $(\{-1,1\}^d)^n$. We use $\sum_{z \in S_k(D_r)} \mathbb{E}[\langle \mathcal{A}(D_r) - \mathcal{I}_k(D_r), \pi z\rangle^2]$ to bound the trace. The idea is to use properties of semidefinite matrices and projection matrices to derive the following inequality,

$$\begin{aligned}\Omega(m_k^2(1-\delta)^2|S_k(D_r)|) = \sum_{z \in S_k(D_r)} \mathbb{E}[\langle \mathcal{A}(D_r) - \mathcal{I}_k(D_r), \pi z\rangle^2] &= tr(\pi^\top \Sigma_{\mathcal{A}}(D_r) \pi m_k U_k(D_r)) \\ &\leq m_k\, tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi) \|U_k(D_r)\|_\infty \leq m_k\, tr(\Sigma_{\mathcal{A}}(D_r)) \|U_k(D_r)\|_\infty.\end{aligned}$$

Therefore, to lower bound $tr(\Sigma_{\mathcal{A}}(D_r))$, we need good upper bound on the largest eigenvalue of $U_k(D_r)$. Lemma 4.14 does that by using the matrix-valued Chernoff bound from Ahlswede and Winter [2].

Facts used in the proof Lemma 4.14. We need some simple definitions to prove Lemma 4.14. 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 $\exp(M)$.

Claim 4.13. *For all $D \in (\{-1,1\}^d)^n$, $\|U_k(D)\|_\infty \leq \|V_k(D)\|_\infty$.*

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

Lemma 4.14. *For all $D \in (\{-1,1\}^d)^n$, $\|U_k(D)\|_\infty \leq \|V_k(D)\|_\infty$, and with probability at least $1 - 1/n$ over the choice of D_r, $\|V_k(D_r)\|_\infty = O(\max\{n/m_k, \log m_k\})$.*

Proof. To prove the lemma we will show that with high probability

$$\|V_k(D_r)\|_\infty = O(\max\{n/m_k, \log m_k\}).$$

Then, by using Claim 4.13 we get the desired result. To prove the bound on $\|V_k(D_r)\|_\infty$ we use the following matrix-valued Chernoff bound of Ahlswede and Winter [2].

Theorem 4.15 ([2, 41])**.** *Suppose $f : [\ell] \to [-\mathbb{I}_{m_k}, \mathbb{I}_{m_k}]$ and let $X_1, \ldots, X_l$ be arbitrary independent random variables distributed over $[\ell]$. Then, for all $\gamma \in \mathbb{R}$ and $t > 0$:*

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

Let $T_k(D_r) = \{z_1, \ldots, z_n\}$. Now $u_z u_z^\top \in [-\mathbb{I}_{m_k}, \mathbb{I}_{m_k}]$ for $z \in \{-1,1\}^{m_k}$. Restating the above theorem:

Corollary 4.16. *Let $z_j \in \{-1,1\}^{m_k}$ for $j \in [n]$. For all $\gamma \in \mathbb{R}$ and $t > 0$,*

$$\Pr_{D_r}\left[\frac{1}{n}\sum_{j=1}^{n} u_{z_j}u_{z_j}^\top \npreceq \gamma \mathbb{I}_{m_k}\right] \leq m_k \exp(-t\gamma n)\prod_{j=1}^{n}\left\|\mathop{\mathbb{E}}_{D_r}[\exp(t\, u_{z_j}u_{z_j}^\top)]\right\|_\infty .$$

Note that $\frac{1}{n}\sum_{j=1}^{n} u_{z_j}u_{z_j}^\top \npreceq \gamma\mathbb{I}_{m_k} \equiv \|\frac{1}{n}V_k(D_r)\|_\infty \geq \gamma$. Also, since $u_{z_1}u_{z_1}^\top, \ldots, u_{z_n}u_{z_n}^\top$ are all independent and identically distributed we can restate the corollary in a more useful form as (where $z = z_1$):

$$\Pr_{D_r}\left[\left\|\tfrac{1}{n}V_k(D_r)\right\|_\infty \geq \gamma\right] \leq m_k \exp(-t\gamma n)\left(\left\|\mathbb{E}_z[\exp(t\, u_z u_z^\top)]\right\|_\infty\right)^n . \tag{9}$$

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. Note that $u_z u_z^\top$ is a rank 1 projection matrix, so it has a single eigenvalue of value 1 and the remaining eigenvalues are all 0. We diagonalize $u_z u_z^\top$ in the basis where u_z is the first eigenvector, then $u_z u_z^\top = P^\top \cdot diag(1, 0, \ldots, 0) \cdot P$, where P is an orthogonal matrix whose first row is u_z and the all the remaining rows are 0's.

Consider $\|\,\mathbb{E}_z[\exp(t\, u_z u_z^\top)]\|_\infty$,

$$\begin{aligned}
&\left\|\mathbb{E}_z[\exp(t\, u_z u_z^\top)]\right\|_\infty = \|\,\mathbb{E}_z[\exp(t \cdot P^\top \cdot diag(1,0,\ldots,0)\cdot P)]\|_\infty \\
&= \|\,\mathbb{E}_z[P^\top \cdot \exp(diag(t,0,\ldots,0))\cdot P]\|_\infty \\
&= \|\,\mathbb{E}_z[P^\top \cdot diag(e^t,1,\ldots,1)\cdot P]\|_\infty \qquad \text{(as } \exp(diag(c_1,\ldots,c_n)) = diag(e^{c_1},\ldots,e^{c_n})\text{)} \\
&= \|\,\mathbb{E}_z[e^t\, u_z u_z^\top]\|_\infty = \left\|e^t\,\tfrac{\mathbb{I}_{m_k}}{m_k}\right\|_\infty = \tfrac{e^t}{m_k}.
\end{aligned}$$

The second last equality follows because $\mathbb{E}_z[u_z u_z^\top] = \mathbb{I}_{m_k}/m_k$ (from Lemma 4.7). Setting $t = 1$, the right hand of Equation 9 simplifies to

$$\begin{aligned}
m_k \exp(-\gamma n)\left(\left\|\mathbb{E}_z[\exp(u_z u_z^\top)]\right\|_\infty\right)^n &= m_k \exp(-\gamma n)\left(\frac{e}{m_k}\right)^n \\
&\leq m_k \exp(-\gamma n)\exp(en/m_k).
\end{aligned}$$

The last inequality uses the fact that $e/m_k \leq \exp(e/m_k)$. We consider two cases:

Case 1: $n \geq 8m_k \log m_k$**.** Setting $\gamma = 3/m_k$, implies that

$$\Pr_{D_r}\left[\left\|\frac{1}{n}V_k(D_r)\right\|_\infty \geq \frac{3}{m_k}\right] \leq m_k \exp\left(\frac{-n(3-e)}{m_k}\right).$$

which simplifies to $\Pr_{D_r}[(1/n)\|V_k(D_r)\|_\infty \geq 3/m_k] \leq 1/n$ as $n \geq 8m_k \log m_k$.

Case 2: $n < 8m_k \log m_k$. Setting $\gamma = (8\log m_k)/n$, implies that

$$\Pr_{D_r}\left[\left\|\frac{1}{n}V_k(D_r)\right\|_\infty \geq \frac{8\log m_k}{n}\right] \leq \frac{1}{n}.$$

Rewriting the above inequalities proves the desired statement. □

Let $\widetilde{\mathcal{D}}$ be the set of all databases from $(\{-1,1\}^d)^n$ which have at least one row of $(1)^d$. If with high probability, $|S_k(D_r)|$ is $\Omega(n)$, then using the upper bound on $\|U_k(D_r)\|_\infty$ from Lemma 4.14 will give us the lower bound on $tr(\Sigma_{\mathcal{A}}(D_r))$. But it is not necessary that $|S_k(D_r)| = \Omega(n)$ (as in constructing $S_k(D_r)$ we only consider the neighbors of D_r belonging to $\widetilde{\mathcal{D}}$). In that case, we show that one could pick a database $\widetilde{D}_r$ uniformly at random from $\widetilde{\mathcal{D}}$, and use a similar analysis to lower bound $tr(\Sigma_{\mathcal{A}}(\widetilde{D}_r))$.

Let $\widetilde{\mathcal{D}}$ be the set of all databases from $(\{-1,1\}^d)^n$ which have 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 $D_1,\dots,D_n$ drawn independently at random from $Neig(\widetilde{D})$. Let $\widetilde{T}_k(\widetilde{D}) = \{z_1,\dots,z_n\}$ denote the (multi) set such that $\mathcal{I}_k(D_i) - \mathcal{I}_k(\widetilde{D}) = \tilde{z}_i$ and $z_i = \tilde{z}_i + o$. $\pi = \mathbb{I}_{m_k} - oo^\top/\langle o, o\rangle$. Define, $\widetilde{S}_k(\widetilde{D}) \subseteq \widetilde{T}_k(\widetilde{D})$, $\widetilde{U}_k(\widetilde{D})$, and $\widetilde{V}_k(\widetilde{D})$ as

$$\widetilde{S}_k(\widetilde{D}) = \{z \in \widetilde{T}_k(\widetilde{D}) \mid \mathbb{E}[\langle \mathcal{A}(\widetilde{D}) - \mathcal{I}_k(\widetilde{D}), \pi z\rangle^2] = \Omega(m_k^2(1-\delta)^2)\},$$
$$\widetilde{U}_k(\widetilde{D}) = \textstyle\sum_{z \in \widetilde{S}_k(\widetilde{D})} u_z u_z^\top \quad \text{and} \quad \widetilde{V}_k(\widetilde{D}) = \textstyle\sum_{z \in \widetilde{T}_k(\widetilde{D})} u_z u_z^\top.$$

The proof of the following lemma follows identical to Lemma 4.14.

Lemma 4.17. *For every $\widetilde{D} \in \widetilde{\mathcal{D}}$, $\|\widetilde{U}_k(\widetilde{D})\|_\infty \leq \|\widetilde{V}_k(\widetilde{D})\|_\infty$ and with probability (over the random choices of $D_1,\dots,D_n$) greater than $1 - 1/n$, $\|\widetilde{V}_k(\widetilde{D})\|_\infty = O(\max\{n/m_k, \log m_k\})$.*

The proof of the following lemma follows as in Lemma 4.6 by using Lemma 4.11 instead of Lemma 4.3 in the proof.

Lemma 4.18. *Let $\mathcal{A}$ be a $(1/2,\delta)$-differentially private algorithm for $\mathcal{I}_k$. Let $\mathcal{A}(D) \approx_{1/2,\delta} \mathcal{A}(D')$. Let $z_r = \mathcal{I}_k(r)$ for a random vector $r \in \{-1,1\}^d$. Let $\tilde{z}_r = \mathcal{I}_k((1)^d) - z_r = o - z_r$, where $o = (1)^{m_k}$. Let $\mathcal{I}_k(D') = \mathcal{I}_k(D) + \tilde{z}$ for some $\tilde{z} \in \operatorname{supp}(\tilde{z}_r)$ 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 $\Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)$.*

The following lemma shows that if expected size of $S_k(D_r)$ is small (less than $n/4$), then the expected size of $\widetilde{S}_k(\widetilde{D}_r)$ is greater than $n/4$ (i.e., at least one of the two expected sizes is greater than $n/4$).

Lemma 4.19. *Let D_r be a database chosen uniformly at random from $(\{-1,1\}^d)^n$ 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 $\forall z \in \operatorname{supp}(z_r)$ (where z_r is defined in Lemma 4.18),

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

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

For databases $D \in (\{-1,1\}^d)^n$ and $\widetilde{D} \in \widetilde{\mathcal{D}}$ define,

$$R_k(D) = \{z \in T_k(D) \mid \mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)\}$$
$$\widetilde{R}_k(\widetilde{D}) = \{z \in \widetilde{T}_k(\widetilde{D}) \mid \mathbb{E}[\langle \mathcal{A}(\widetilde{D}) - \mathcal{I}_k(\widetilde{D}), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)\}.$$

Let D_r and $\widetilde{D}_r$ be random databases from $(\{-1,1\}^d)^n$ 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 $1/2$. By using this along with the definitions of $S_k(D_r)$ and $R_k(D_r)$ implies

$$\mathop{\mathbb{E}}_{D_r}[|S_k(D_r)|] \geq \frac{\mathbb{E}_{D_r}[|R_k(D_r)|]}{2}.$$

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 $1/2$, therefore,

$$\mathop{\mathbb{E}}_{\widetilde{D}_r}[|\widetilde{S}_k(\widetilde{D}_r)|] \geq \frac{\mathbb{E}_{\widetilde{D}_r}[|\widetilde{R}_k(\widetilde{D}_r)|]}{2}.$$

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

For any database $D_a \in Neig(\widetilde{D}_r)$, we know (from Lemma 4.18) that at least one of $\mathbb{E}[\langle \mathcal{A}(\widetilde{D}_r) - \mathcal{I}_k(\widetilde{D}_r), \pi z, \rangle^2]$ or $\mathbb{E}[\langle \mathcal{A}(D_a) - \mathcal{I}_k(D_a), \pi z\rangle^2]$ is $\Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)$ (where $z = \mathcal{I}_k(D_a) - \mathcal{I}_k(\widetilde{D}_r) + o$). Now, if $\mathbb{E}_{D_r}[|R_k(D_r)|] < n/2$, then

$$\mathop{\mathbb{E}}_{\widetilde{D}_r}\left[\left|\left\{D_a \in Neig(\widetilde{D}_r) \,:\, \mathbb{E}[\langle \pi z, \mathcal{A}(\widetilde{D}_r) - \mathcal{I}_k(\widetilde{D}_r)\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2)\right\}\right|\right] \geq \frac{|Neig(\widetilde{D}_r)|}{2}.$$

Therefore, for $\widetilde{D}_r$ chosen uniformly at random from $\widetilde{\mathcal{D}}$ and D_b chosen uniformly at random $Neig(\widetilde{D}_r)$, with probability at least $1/2$,

$$\mathbb{E}[\langle \mathcal{A}(\widetilde{D}_r) - \mathcal{I}_k(\widetilde{D}_r), \pi z\rangle^2] = \Omega(\langle \pi z, \pi z\rangle^2(1-\delta)^2) \text{ where } z = \mathcal{I}_k(D_b) - \mathcal{I}_k(\widetilde{D}_r) + o.$$

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 $(1/2)\cdot(n/2) \geq n/4$. □

The following proposition uses Lemmata 4.14, 4.17, and 4.19 to show that every differentially private algorithm with probability $\Omega(1-1/n)$ needs to add a lot of noise to either $\mathcal{I}_k(D_r)$ or $\mathcal{I}_k(\widetilde{D}_r)$.

Proposition 4.20 (General case: k-way inner products)**.** *Let $\mathcal{A} : (\{-1,1\}^d)^n \to \mathbb{R}^{m_k}$ be a $(1/2,\delta)$-differentially private algorithm for $\mathcal{I}_k$. Let D_r be a database chosen uniformly at random from $(\{-1,1\}^d)^n$ and $\widetilde{D}_r$ be a database chosen uniformly at random from $\widetilde{\mathcal{D}}$. Then, with probability $\Omega(1-1/n)$, at least one of $tr(\Sigma_{\mathcal{A}}(D_r))$ or $tr(\Sigma_{\mathcal{A}}(\widetilde{D}_r))$ is $\Omega(\min\{m_k^2(1-\delta)^2, nm_k(1-\delta)^2/(\log m_k)\})$.*

Proof. We divide the proof into two cases based on Lemma 4.19. Let $o = (1)^{m_k}$.

Case 1: $\mathbb{E}_{D_r}[|S_k(D_r)|] \geq n/4$. Let $\Sigma_{\mathcal{A}}(D_r) = \mathbb{E}[(\mathcal{A}(D_r) - \mathcal{I}_k(D_r))(\mathcal{A}(D_r) - \mathcal{I}_k(D_r))^\top]$ be the mean squared error matrix. By definition of $S_k(D_r)$,

$$\sum_{z \in S_k(D_r)} (\pi z)^\top \Sigma_{\mathcal{A}}(D_r)(\pi z) = \sum_{z \in S_k(D_r)} \mathbb{E}[\langle \mathcal{A}(D_r) - \mathcal{I}_k(D_r), \pi z\rangle^2] = \Omega(m_k^2(1-\delta)^2|S_k(D_r)|).$$

On the other hand,

$$\begin{aligned}\sum_{z \in S_k(D_r)} (\pi z)^\top \Sigma_{\mathcal{A}}(D_r)(\pi z) &= \sum_{z \in S_k(D_r)} tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi z z^\top) = tr\left(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi \sum_{z \in S_k(D_r)} z z^\top\right) \\ &= m_k\, tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi U_k(D_r)).\end{aligned}$$

Note that $\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi$ is a positive semidefinite matrix and so is $U_k(D_r)$, therefore, $tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi U_k(D_r)) \leq tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi)\|U_k(D_r)\|_\infty$. Also, since π is an orthogonal projection matrix

$$tr(\pi^\top \Sigma_{\mathcal{A}}(D_r)\pi) = tr(\Sigma_{\mathcal{A}}(D_r)\pi) \leq tr(\Sigma_{\mathcal{A}}(D_r))\|\pi\|_\infty = tr(\Sigma_{\mathcal{A}}(D_r)).$$

Therefore,

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

Let $H_k(D_r) = U_k(D_r)/|S_k(D_r)|$. 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,

$$\Omega(m_k(1-\delta)^2) = tr(\Sigma_{\mathcal{A}}(D_r))\|H_k(D_r)\|_\infty.$$

From our assumption, we know that $\mathbb{E}_{D_r}[|S_k(D_r)|] \geq n/4$. Let E_1 be the random event that $|S_k(D_r)| \leq n/8$. Since $\mathbb{E}_{D_r}[|S_k(D_r)|] \geq n/4$, $\Pr[E_1] \leq 6/7$. Let E_2 be the random event that

$$\|H_k(D_r)\|_\infty \geq \max\{cn/(m_k|S_k(D_r)|), (c\log m_k)/|S_k(D_r)|\}$$

for some constant c. From Lemma 4.14, we know that, with probability at least $1 - 1/n$ over D_r,

$$\|H_k(D_r)\|_\infty \leq \max\left\{\frac{cn}{m_k|S_k(D_r)|}, \frac{c\log m_k}{|S_k(D_r)|}\right\}.$$

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

$$tr(\Sigma_{\mathcal{A}}(D_r)) \geq \min\left\{\frac{m_k^2(1-\delta)^2}{c'n}|S_k(D_r)|, \frac{m_k(1-\delta)^2}{c'\log m_k}|S_k(D_r)|\right\}.$$

Since with probability at least $1-\Pr[E_1]$, $|S_k(D_r)| \geq n/4$, we get that with probability at least $1-\Pr[E_1]-\Pr[E_2] \geq 1 - 6/7 - 1/n$,

$$tr(\Sigma_{\mathcal{A}}(D_r)) = \Omega\left(\min\left\{m_k^2(1-\delta)^2, \frac{nm_k(1-\delta)^2}{\log m_k}\right\}\right).$$

Case 2: $\mathbb{E}_{\widetilde{D}_r}[|\widetilde{S}_k(\widetilde{D}_r)|] \geq n/4$**.** The proof of this case goes similar to the previous case. We define $\widetilde{H}_k(\widetilde{D}_r) = \widetilde{U}_k(\widetilde{D}_r)/|\widetilde{S}_k(\widetilde{D}_r)|$. In this case, we use Lemma 4.17 to bound $\|\widetilde{U}_k(\widetilde{D}_r)\|_\infty$.

Since by Lemma 4.19 at least one of the cases hold, we get that with probability $\Omega(1-1/n)$ there exists a database such that trace of its mean squared error matrix is $\Omega(\min\{m_k^2(1-\delta)^2, \frac{nm_k(1-\delta)^2}{\log m_k}\})$. □

In Section 4.3, we show how this lower bound for $(1/2,\delta)$-differentially private algorithms can be converted into a lower bound for (ϵ,δ)-differentially private algorithms. We now summarize the main result.

Theorem 4.21 (General Case)**.** *Let $m_k = \binom{d}{k}$. Any algorithm $\mathcal{A}$ for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that for every database $D \in (\{0,1\}^d)^n$ has a root mean squared error of*

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

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

Proof. Proof follows from Proposition 4.20 and Lemma 4.24. Corollary C.2 can be used to convert the result on inner products to conjunctions. Remember that $|\mathcal{C}_k| = 2^k m_k$, whereas $|\mathcal{I}_k| = m_k$. □

4.3 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 4.22. *If $\mathcal{A}$ is (ϵ,δ)-differentially private then $\mathcal{A}'$ is $(1/2, \delta/(2\epsilon))$-differentially private.*

Proof. 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 4.2), for every output set $\mathcal{S}$

$$\Pr[\mathcal{A}(R(D_v)) \in \mathcal{S}] \le e^{1/2}\Pr[\mathcal{A}(R(D'_v)) \in \mathcal{S}] + \frac{\delta}{2\epsilon} \Rightarrow \Pr[\mathcal{A}'(D_v) \in \mathcal{S}] \le e^{1/2}\Pr[\mathcal{A}'(D'_v) \in \mathcal{S}] + \frac{\delta}{2\epsilon}.$$

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

Claim 4.23. *There exists a database $D_v \in (\{-1,1\}^d)^{2\epsilon n}$ such that*

$$tr(\Sigma_{\mathcal{A}'}(D_v)) = \Omega(\min\{m_k^2(1-\delta/\epsilon)^2, (n\epsilon m_k(1-\delta/\epsilon)^2)/\log m_k\}).$$

Proof. Since $\mathcal{A}'$ is $(1/2, \delta/(2\epsilon))$-differentially private (Claim 4.22), means that we can apply Proposition 4.20 to conclude that there exists a database D_v of size $2\epsilon n$ such that $tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top]) = \Omega(\min\{m_k^2(1-\delta/\epsilon)^2, (n\epsilon m_k(1-\delta/\epsilon)^2)/\log m_k\})$. □

Lemma 4.24. *Let $\mathcal{A}$ be an (ϵ,δ)-differentially private algorithm for $\mathcal{I}_k$. Let D_v be the database such that $tr(\Sigma_{\mathcal{A}'}(D_v)) = \Omega(\min\{m_k^2(1-\delta/\epsilon)^2, (n\epsilon m_k(1-\delta/\epsilon)^2)/\log m_k\})$. Then,*

$$tr(\Sigma_{\mathcal{A}}(R(D_v))) = \Omega\left(\min\left\{\frac{m_k^2(1-\delta/\epsilon)^2}{\epsilon^2}, \frac{(nm_k(1-\delta/\epsilon)^2)}{\epsilon\log m_k}\right\}\right).$$

Proof. We equate $tr(\Sigma_{\mathcal{A}}(R(D_v)))$ in terms of $tr(\Sigma_{\mathcal{A}'}(D_v))$.

$$\begin{aligned}
&tr(\Sigma_{\mathcal{A}}(R(D_v))) = tr(\mathbb{E}[(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))(\mathcal{A}(R(D_v)) - \mathcal{I}_k(R(D_v)))^\top])\\
&= tr\left(\mathbb{E}\left[\left(\frac{\mathcal{A}'(D_v)}{2\epsilon} - \frac{\mathcal{I}_k(D_v)}{2\epsilon}\right)\left(\frac{\mathcal{A}'(D_v)}{2\epsilon} - \frac{\mathcal{I}_k(D_v)}{2\epsilon}\right)^\top\right]\right) \qquad \text{(by definition of the algorithm } \mathcal{A}')\\
&= \frac{1}{4\epsilon^2} tr(\mathbb{E}[(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))(\mathcal{A}'(D_v) - \mathcal{I}_k(D_v))^\top])\\
&= \frac{1}{4\epsilon^2} tr(\Sigma_{\mathcal{A}}(R(D_v))) = \Omega\left(\min\left\{\frac{m_k^2(1-\delta/\epsilon)^2}{\epsilon^2}, \frac{nm_k(1-\delta/\epsilon)^2}{\epsilon\log m_k}\right\}\right).
\end{aligned}$$

The last equality follows from Claim 4.23. □

References

[1] Agrawal, R., and Srikant, R. Privacy-preserving data mining. In *SIGMOD* (2000), pp. 439–450.

[2] Ahlswede, R., and Winter, A. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory 48*, 3 (2002), 569–579.

[3] 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.

[4] Bishop, Y., Fienberg, S., and Holland, P. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge MA, 1975.

[5] Blum, A., Dwork, C., McSherry, F., and Nissim, K. Practical privacy: The SuLQ framework. In *PODS* (2005), ACM, pp. 128–138.

[6] Blum, A., Ligett, K., and Roth, A. A learning theory approach to non-interactive database privacy. In *STOC* (2008), ACM, pp. 609–618.

[7] Chen, B.-C., Ramakrishnan, R., and LeFevre, K. Privacy skyline: Privacy with multidimensional adversarial knowledge. In *VLDB* (2007), VLDB Endowment, pp. 770–781.

[8] Dinur, I., Dwork, C., and Nissim, K. Revealing information while preserving privacy, full version of [9], in preparation, 2009.

[9] Dinur, I., and Nissim, K. Revealing information while preserving privacy. In *PODS* (2003), ACM, pp. 202–210.

[10] Dwork, C. Differential privacy: A survey of results. In *TAMC* (2008), pp. 1–19.

[11] Dwork, C. The differential privacy frontier (extended abstract). In *TCC* (2009), pp. 496–502.

[12] 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.

[13] Dwork, C., and Lei, J. Differential privacy and robust statistics. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing* (2009), pp. 371–380.

[14] 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.

[15] Dwork, C., McSherry, F., and Talwar, K. The price of privacy and the limits of LP decoding. In *STOC* (2007), ACM, pp. 85–94.

[16] Dwork, C., and Yekhanin, S. New efficient attacks on statistical disclosure control mechanisms. In *CRYPTO* (2008), Springer, pp. 469–480.

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

[18] Evfimievski, A., Srikant, R., Agrawal, R., and Gehrke, J. Privacy preserving mining of association rules. In *KDD* (2002), pp. 217–228.

[19] Fienberg, S., and Willenborg, L. C. R. J. Special issue on disclosure control. *Journal of Official Statistics 14*, 4 (1998).

[20] Ghosh, A., Roughgarden, T., and Sundararajan, M. Universally utility-maximizing privacy mechanisms. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing* (2009), pp. 351–360.

[21] Hardt, M., and Talwar, K. On the Geometry of Differential Privacy. *Arxiv preprint arXiv:0907.3754* (2009).

[22] 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.

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

[24] LEDOUX, M. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. *American Mathematical Society, Providence, RI 208* (2001), 2005–2006.

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

[26] 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.

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

[28] MARTIN, D. J., KIFER, D., MACHANAVAJJHALA, A., GEHRKE, J., AND HALPERN, J. Y. Worst-case background knowledge for privacy-preserving data publishing. In *ICDE* (2007), IEEE Computer Society, pp. 126–135.

[29] MCSHERRY, F., AND MIRONOV, I. Differentially private recommender systems: building privacy into the net. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), ACM New York, NY, USA, pp. 627–636.

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

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

[32] RASTOGI, V., HONG, S., AND SUCIU, D. The boundary between privacy and utility in data publishing. In *VLDB* (2007), pp. 531–542.

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

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

[35] RUDELSON, M., AND VERSHYNIN, R. The smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics* (2009), 1707 – 1739.

[36] SKINNER, C., AND SHLOMO, N. Assessing identification risk in survey microdata using log-linear models. *Journal of the American Statistical Association 103*, 483 (2008), 989–1001.

[37] SMITH, A. Efficient, differentially private point estimators. *CoRR abs/0809.4794* (2008).

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

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

[40] WARNER, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*, 309 (1965), 63–69.

[41] 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.

[42] WONG, R. C.-W., FU, A. W.-C., WANG, K., AND PEI, J. Minimality attack in privacy preserving data publishing. In *VLDB* (2007), VLDB Endowment, pp. 543–554.

[43] XIAO, X., AND TAO, Y. M-invariance: towards privacy preserving re-publication of dynamic datasets. In *SIGMOD* (2007), ACM Press, pp. 689–700.

A Interpreting the Blum-Ligett-Roth [6] Upper Bound for MSE

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 A.1 ((α, β)-usefulness [6])**.** *An algorithm $\mathcal{A}$ is (α, β)-useful for class of predicates $\mathcal{F}$ and a database D if, with probability at least $1 - \beta$, $\mathcal{A}(D)$ outputs a database $\widehat{D}$ that satisfies*

$$\left| f_j(\widehat{D})/|\widehat{D}| - f_j(D)/|D| \right| \leq \alpha$$

for every $f_j \in \mathcal{F}$.

For a function class $\mathcal{F}$, let $VCDIM(\mathcal{F})$ represent the VC-dimension of $\mathcal{F}$.

Theorem A.2 ([6])**.** *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 A.3 (BLR [6] 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 $\widetilde{O}((n^2 \cdot VCDIM(\mathcal{F}) \cdot d/\epsilon)^{2/3})$.*

Proof. Let D be a database. Theorem A.2 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 mean squared 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 if

$$n \geq \frac{Cd \cdot VCDIM(\mathcal{F}) \cdot \log(1/\alpha)}{\alpha^3 \epsilon} \tag{10}$$

then the mean squared error is $2\alpha^2 n^2$. Isolating α from Equation 10, and substituting it in $2\alpha^2 n^2$, we get that the expected error for each count in the Blum, Ligett, and Roth mechanism is $\widetilde{O}((n^2 \cdot VCDIM(\mathcal{F}) \cdot d/\epsilon)^{2/3})$. □

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

Corollary A.4. *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 $\widetilde{O}((n^2 dk/\epsilon)^{2/3})$.*

B Upper Bounds for (Not) Row Non-Privacy and (Not) Attribute Non-Privacy

Proposition B.1 (Row Non-Privacy: upper bound)**.** *There exist an algorithm for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that is* not *row non-private, and that for every database $D \in (\{0,1\}^d)^n$ with constant probability adds $O(\min\{\sqrt{nk\log(d/k)}, \sqrt{m_k k\log(d/k)}\})$ noise to each query in $\mathcal{C}_k(D)$.*

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

Random Sampling. Let $D \in (\{0,1\}^d)^n$ be a database. 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 $2 \cdot \mathcal{C}_k(D_{sam})$.

Let $\mathbb{D}$ be a good distribution and let S be a good set. Consider $D_p \sim \mathbb{D}$. We give D_p as input to $\mathcal{A}_{sam}$. Now, given $\mathcal{A}_{sam}(D_p)$ an adversary can output the ith row of D_p only if: (a) if the ith row is included in the sampling done by $\mathcal{A}_{sam}$, or (b) with probability at most $2/3$ if the ith is discarded in the sampling. From the cases (a) and (b), it follows that the probability that an adversary can output the ith row is at most $1/2 + 1/2 \times 2/3 = 5/6$. Since the rows are independent of each other, no adversary (even with unbounded time) can output $1 - o(1)$ fraction of the rows indexed by the elements of S with $1 - negl(d)$ probability. This shows that random sampling is *not* row non-private.

We now invoke Chernoff bound to argue about the noise. Consider some conjunction predicate $c_v \in \mathcal{C}_k$. Now, for some constants t, t',

$$\Pr\left[|2 \cdot c_v(D_{sam}) - c_v(D)| \geq \sqrt{tn\log(2^k m_k)}\right] \leq \exp\left(-2 \cdot n \cdot \frac{t\log(2^k m_k)}{n}\right) \leq \frac{1}{t' 2^k m_k}.$$

By applying a union bound it follows that the probability that

$$\forall c_v \in \mathcal{C}_k,\ \Pr\left[|2 \cdot c_v(D_{sam}) - c_v(D)| \geq \sqrt{tn\log(2^k m_k)}\right] \leq \frac{1}{t'}.$$

Therefore, for every database D the above random sampling procedure with constant probability adds $O(\sqrt{nk\log(d/k)})$ noise to each query in $\mathcal{C}_k(D)$.

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

Lemma B.2. *Any (ϵ, δ)-differentially private algorithm with $(2/3)e^\epsilon + \delta$ is bounded away from 1 is* not *row non-private.*

Proof. Let $\mathcal{A}_{dp}$ be an (ϵ, δ)-differentially private algorithm satisfying the conditions of the lemma, we argue that $\mathcal{A}_{dp}$ is *not* row non-private. Let $\mathbb{D}$ be a good distribution and S be a good set. Consider $D_p \sim \mathbb{D}$. Because of the guarantees of (ϵ, δ)-differential privacy, given $\mathcal{A}_{dp}(D_p)$, no adversary (even with unbounded time) can predict any row of D_p 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_p indexed by the elements of S is small. □

The SuLQ mechanism of Blum *et al.* [5] adds independent noise drawn according to the normal distribution (with mean 0 and standard deviation $\sqrt{2m_k \log(1/\delta)}/\epsilon$) to each entry in $\mathcal{C}_k(D)$. We set $\epsilon = 0.1$ and $\delta = 0.1$. By Lemma B.2, the SuLQ mechanism for these values of ϵ and δ is *not* row non-private. A simple analysis of the c.d.f. of the normal distribution and an application of the union bound shows that for every database D the SuLQ mechanism with constant probability adds $O(\sqrt{m_k \log(2^k m_k)}) = O(\sqrt{m_k k \log(d/k)})$ noise to each query in $\mathcal{C}_k(D)$.

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 follows that $\mathcal{A}$ is *not* row non-private and has the claimed noise bounds. □

Proposition B.3 (Attribute Non-Privacy: upper bound)**.** *There exist an algorithm for releasing all k-attribute marginal tables (or all k-way conjunction predicates) that is* not *attribute non-private, and that for every database $D \in (\{0,1\}^d)^n$ with constant probability adds $O(\min\{\sqrt{nk\log(d/k)}, \sqrt{m_k k \log(d/k)}\})$ noise to each query in $\mathcal{C}_k(D)$.*

Proof. The proof is very similar to Proposition B.1. To show an algorithm $\mathcal{A}$ is *not* attribute non-private we show that there exists a $s \in \{0,1\}^n$ such that for all databases $D(s)$ whose last column is s, no adversary given as input $\mathcal{A}(D(s))$ and the first $d-1$ columns of $D(s)$ can with probability $1 - negl(d)$ reconstruct $\widetilde{\Omega}(\min\{n, d^{k-1}\})$ entries of s. We again construct two different algorithms, both of which are *not* attribute non-private and together they give the claimed noise bound.

Random Sampling. Let s_r be a random vector from $\{0,1\}^n$ (each entry in s_r is 0 or 1 independently with probability $1/2$). Let $D(s_r)$ be a database whose last column is s_r. Consider the algorithm $\mathcal{A}_{sam}$ from Proposition B.1. Given, the first $d-1$ columns of $D(s_r)$ and $\mathcal{A}_{sam}(D(s_r))$, the probability that an adversary can guess an entry of s_r is at most $1/2 + 1/2 \times 1/2 = 3/4$. Since the entries of s_r are independent of each other, no adversary (even with unbounded time) can reconstruct $\widetilde{\Omega}(\min\{n, d^{k-1}\})$ entries of s_r with $1 - negl(d)$ probability. This shows that random sampling is *not* attribute non-private.

The noise analysis is same as in Proposition B.1. It shows that for every database D the random sampling procedure with constant probability adds $O(\sqrt{nk\log(d/k)})$ noise to each query in $\mathcal{C}_k(D)$.

Adapting Differential Privacy. Using a random $s \in \{0,1\}^n$ and a database $D(s)$, we can show similar to Lemma B.2 that any (ϵ, δ)-differentially private algorithm (for reasonable values of ϵ and δ) is *not* attribute non-private.

The noise analysis is same as in Proposition B.1. We get that for every database D the SuLQ mechanism with constant probability adds $O(\sqrt{m_k \log(2^k m_k)}) = O(\sqrt{m_k k \log(d/k)})$ noise to each query in $\mathcal{C}_k(D)$.

Putting Together. As in Proposition B.1. We get an algorithm that is *not* attribute non-private and has the claimed noise bounds. □

C Going From Inner Products to Conjunctions

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

Let us consider the case of 2-way conjunctions. Now, consider all the 4 possible conjunctions on any two variables x_i and x_j. The conjunction predicates $(c_{x_i x_j}, c_{\bar{x}_i \bar{x}_j}, c_{\bar{x}_i x_j}, c_{x_i \bar{x}_j})$ and the inner product predicates $(i_{y_j}, i_{y_i y_j}, i_{y_i})$ can be related using a Hadamard matrix as,

$$\underbrace{\begin{pmatrix} n \\ i_{y_j}(D_o) \\ i_{y_i y_j}(D_o) \\ i_{y_i}(D_o) \end{pmatrix}}_{F} = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}}_{H_2} \underbrace{\begin{pmatrix} c_{x_i x_j}(D_z) \\ c_{\bar{x}_i \bar{x}_j}(D_z) \\ c_{\bar{x}_i x_j}(D_z) \\ c_{x_i \bar{x}_j}(D_z) \end{pmatrix}}_{G} \qquad (11)$$

Therefore, $F = H_2 G$ where F, G, H_2 are defined in Equation 11.

Let $m_2 = \binom{d}{2}$. Now, consider the vectors $\mathcal{I}_2(D_o)$ (defined as all 2-way inner product predicates evaluated on D_o), $\mathcal{I}_1(D_o)$ (defined as all 1-way inner product predicates evaluated on D_o), and $\mathcal{I}_0(D_o) = (n)^{m_2}$, and let $\mathcal{I}_{\leq 2}(D_o)$ be a vector obtained by concatenating the entries of $\mathcal{I}_2(D_o), \mathcal{I}_1(D_o), \mathcal{I}_1(D_o)$, and $\mathcal{I}_0(D_o)$. A simple extension of Equation 11, shows that

$$\mathcal{I}_{\leq 2}(D_o) = diag(H_2,\ldots,H_2)\cdot \mathcal{C}_2(D_z),$$

where $diag(H_2,\ldots,H_2)$ is a block diagonal matrix. By using a suitable projection matrix Π_2 to cancel out all but 2-way inner products we get,

$$\mathcal{I}_2(D_o) = \Pi_2 \cdot diag(H_2,\ldots,H_2)\cdot \mathcal{C}_2(D_z).$$

The following proposition generalizes this observation.

Proposition C.1. *If there exists an (ϵ,δ)-differentially private algorithm $\mathcal{A}$ for $\mathcal{C}_k$ that has $tr(\Sigma_{\mathcal{A}}(D_z)) \leq T$, then there exists an (ϵ,δ)-differentially private algorithm $\mathcal{B}$ for $\mathcal{I}_k$ that has $tr(\Sigma_{\mathcal{B}}(D_o)) \leq 2^k T$.*

Proof. Let H_k be a $2^k \times 2^k$ Hadamard matrix. Let $D_o \in (\{-1,1\}^d)^n$ and D_z is a database obtained by replacing all the -1's in D_o by 0. Then,

$$\mathcal{I}_k(D_o) = \Pi_k \cdot diag(H_k,\ldots,H_k)\cdot \mathcal{C}_k(D_z),$$

where Π_k is a suitable projection matrix to cancel out all but k-way inner products.

Now, given an (ϵ,δ)-differentially private algorithm $\mathcal{A}$, we define an algorithm $\mathcal{B}$ as follows: $\mathcal{B}(D_o) = \Pi_k \cdot diag(H_k,\ldots,H_k)\cdot \mathcal{A}(D_z)$. By Claim 4.2 (described below), $\mathcal{B}$ is also (ϵ,δ)-differentially private. The largest eigenvalue of the Hadamard matrix H_k is $2^{k/2}$, therefore, the largest eigenvalue of $diag(H_k,\ldots,H_k)$ is $2^{k/2}$. Since all the eigenvalues of the projection matrix Π_k are either 0 or 1, therefore, the largest eigenvalue (operator norm) of $\Pi_k \cdot diag(H_k,\ldots,H_k)$ is $2^{k/2}$ as

$$\|\Pi_k \cdot diag(H_k,\ldots,H_k)\|_\infty \leq \|\Pi_k\|_\infty \|diag(H_k,\ldots,H_k)\|_\infty.$$

Therefore,

$$\begin{aligned}
tr(\Sigma_{\mathcal{B}}(D_o)) &= \mathbb{E}[\|\mathcal{B}(D_o) - \mathcal{I}_k(D_o)\|^2] \\
&= \mathbb{E}[\|\Pi_k \cdot diag(H_k, \ldots, H_k) \cdot \mathcal{A}(D_z) - \Pi_k \cdot diag(H_k, \ldots, H_k) \cdot \mathcal{C}_k(D_z)\|^2] \\
&\leq \mathbb{E}[\|\Pi_k \cdot diag(H_k, \ldots, H_k)\|_\infty^2 \|\mathcal{A}(D_z) - \mathcal{C}_k(D_z)\|^2] \\
&= \|\Pi_k \cdot diag(H_k, \ldots, H_k)\|_\infty^2 \| \mathbb{E}[\|\mathcal{A}(D_z) - \mathcal{C}_k(D_z)\|^2] = 2^k tr(\Sigma_{\mathcal{A}}(D_z)).
\end{aligned}$$

Therefore, if $tr(\Sigma_{\mathcal{A}}(D_z)) \leq T$, then $tr(\Sigma_{\mathcal{B}}(D_o)) \leq 2^k T$. □

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

D Tightness of Lemma 4.11

We present an example of an (ϵ, δ)-differentially private algorithm $\mathcal{A}$ and neighboring databases D, D' where $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \Delta \rangle^2]$ is small, whereas, $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_k(D'), \Delta \rangle^2]$ is big. For simplicity, we set $n = 1$. Let k be an odd number. Let $D = (-1)^d$ and $D' = (1)^d$. D and D' are neighbors, and $\Delta = \mathcal{I}_k(D') - \mathcal{I}_k(D) = (2)^{m_k}$.

Let $\mathcal{A}$ be an algorithm for $\mathcal{I}_k$ whose output for all inputs is $(-1)^{m_k}$. $\mathcal{A}$ is differentially private (in fact, for the best possible parameters of $\epsilon = 0$ and $\delta = 0$). Now, $\Sigma_{\mathcal{A}}(D)$ is an $m_k \times m_k$ matrix all whose entries are 0, and $\Sigma_{\mathcal{A}}(D')$ is an $m_k \times m_k$ matrix all whose entries are 2. Therefore, $\mathbb{E}[\langle \mathcal{A}(D) - \mathcal{I}_k(D), \Delta \rangle^2] = 0$, whereas, $\mathbb{E}[\langle \mathcal{A}(D') - \mathcal{I}_k(D'), \Delta \rangle^2] = \Omega(m_k^2)$.