The Power of Linear Reconstruction Attacks

Shiva Prasad Kasiviswanathan* Mark Rudelson† Adam Smith‡

Abstract

We consider the power of "linear reconstruction attacks" in statistical data privacy, showing that they can be applied to a much wider range of settings than previously understood. Linear attacks have been studied before [3, 6, 11, 1, 14] but have so far been applied only in settings with releases that are "obviously" linear.

Consider a database curator who manages a database of sensitive information but wants to release statistics about how a *sensitive* attribute (say, disease) in the database relates to some *nonsensitive* attributes (e.g., postal code, age, gender, etc). This setting is widely considered in the literature, partly since it arises with medical data. Specifically, we show one can mount linear reconstruction attacks based on any release that gives:

1. the fraction of records that satisfy a given non-degenerate boolean function. Such releases include contingency tables (previously studied by Kasiviswanathan *et al.* [11]) as well as more complex outputs like the error rate of classifiers such as decision trees;
2. any one of a large class of M-estimators (that is, the output of empirical risk minimization algorithms), including the standard estimators for linear and logistic regression.

We make two contributions: first, we show how these types of releases can be transformed into a *linear* format, making them amenable to existing polynomial-time reconstruction algorithms. This is already perhaps surprising, since many of the above releases (like M-estimators) are obtained by solving highly nonlinear formulations.

Second, we show how to analyze the resulting attacks under various distributional assumptions on the data. Specifically, we consider a setting in which the same statistic (either 1 or 2 above) is released about how the sensitive attribute relates to all subsets of size k (out of a total of d) nonsensitive boolean attributes.

1 Introduction

The goal of *private data analysis* is to provide global, statistical properties of a database of sensitive information while protecting the privacy of the individuals whose records the database contains. There is a vast body of work on this problem in statistics and computer science.

Until a few years ago, 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 [3] and Dinur, Dwork and Nissim [2] initiated a rigorous study of the tradeoff between privacy and utility. The notion of *differential privacy* (Dwork, McSherry, Nissim and Smith [5],

*GE Global Research Center, San Ramon. Most of the work was done when the author was a postdoc at the IBM T. J. Watson Research Center, kasivisw@gmail.com.

†University of Michigan, rudelson@umich.edu.

‡Pennsylvania State University, asmith@cse.psu.edu.

Dwork [4]) 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. A typical objective is to release as accurate an approximation as possible to some high-dimensional function f evaluated on the database D.

A complementary line of work seeks to establish lower bounds on how much distortion is necessary for particular functions f. Some of these bounds apply only to specific notions of privacy (*e.g.,* lower bounds for differential privacy [5, 8, 9, 13, 1]). A second class of bounds rules out *any* reasonable notion of privacy by giving algorithms to reconstruct almost all of the data D given sufficiently accurate approximations to $f(D)$ [3, 6, 7, 11, 1]. We refer to the latter results as *reconstruction attacks*.

We consider reconstruction attacks against *attribute privacy*: consider a curator who manages a database of sensitive information but wants to release statistics about how a *sensitive* attribute (say, disease) in the database relates to some *nonsensitive* attributes (e.g., postal code, age, gender, etc). This setting is widely considered in the applied data privacy literature, partly since it arises with medical and retail data.

For concreteness, consider a database D that contains, for each individual i, a sensitive attribute $s_i \in \{0,1\}$ as well as some other information $U_i \in \mathbb{R}^d$ which is assumed to be known to the attacker. The ith record is thus (U_i, s_i). We denote the entire database $D = (U|\mathbf{s})$ where $U \in \mathbb{R}^{n\times d}$, $\mathbf{s} \in \{0,1\}^n$, and $|$ denote concatenation. Given some released information $\mathbf{y}$, the attacker constructs an estimate $\hat{\mathbf{s}}$ that she hopes is close to $\mathbf{s}$. We measure the attack's success in terms of the Hamming distance $d_H(\mathbf{s}, \hat{\mathbf{s}})$. A scheme is *not* attribute private if an attacker can consistently get an estimate that is within distance $o(n)$. Formally:

Definition 1 (Failure of Attribute Privacy)**.** [1] *A (randomized) mechanism $\mathcal{M} : \mathbb{R}^{n\times d+1} \to \mathbb{R}^a$ is said to allow (α, β) attribute reconstruction if there exists a setting of the nonsensitive attributes $U \in \mathbb{R}^{n\times d}$ and an algorithm (adversary) $\mathcal{A} : \mathbb{R}^{n\times d} \times \mathbb{R}^a \to \mathbb{R}^n$ such that for every $\mathbf{s} \in \{0,1\}^n$,*

$$\Pr_{\mathbf{y}\leftarrow\mathcal{M}((U|\mathbf{s}))}[\mathcal{A}(U, \mathbf{y}) = \hat{\mathbf{s}} : d_H(\mathbf{s}, \hat{\mathbf{s}}) \leq \alpha] \geq 1 - \beta.$$

Asymptotically, we say that a mechanism is *attribute nonprivate* if there is an infinite sequence of n for which $\mathcal{M}$ allows $(o(1), o(1))$-reconstruction. Here $d = d(n)$ is a function of n. We say the attack $\mathcal{A}$ is *efficient* if it runs in time $\text{poly}(n, d)$.

Instead of simply showing that a setting of U exists, we will normally aim to show that reconstruction is possible with high probability when U is chosen from one of a class of natural distributions.

Linear Reconstruction Attacks. In this paper, we consider the power of *linear* reconstruction attacks. Given the released information $\mathbf{y}$, the attacker constructs a system of approximate linear equalities, namely a matrix A and vector $\mathbf{z}$ such that $A\mathbf{s} \approx \mathbf{z}$ and attempts to solve for $\mathbf{s}$. A typical algorithmic approach is to find $\hat{\mathbf{s}}$ which minimizes some norm (ℓ_2 or ℓ_1) of the error $(A\hat{\mathbf{s}} - \mathbf{z})$. Minimizing the ℓ_2 error is known as *least squares decoding* and minimizing the ℓ_1 error is known as *LP decoding*. One sometimes also considers algorithms that exhaustively search over all 2^n possible choices for $\mathbf{s}$ (as in [3, 14]).

Such attacks were first considered in the context of data privacy by Dinur and Nissim [3]. They showed that any mechanism which answers (or allows the user to compute) $\Omega(n \log n)$ *random* inner product queries with $\{0,1\}$ vectors on a database $\mathbf{s} \in \{0,1\}^n$ with $o(\sqrt{n})$ noise per query is not private. That is, they assume that the mechanism releases $\mathbf{y} = A\mathbf{s} + \mathbf{e}$, where A is a random matrix in $\{0,1\}^{\Omega(n\log n)\times n}$ and $\mathbf{e}$ is a noise

[1] This definition generalizes *blatant non-privacy* (Dinur and Nissim [3]) and first appeared in [11]. The order of the qualifiers here has been changed, correcting an error pointed out by Graham Cormode.

vector with $\|\mathbf{e}\|_\infty = o(\sqrt{n})$. Their attack was subsequently extended to use a linear number of queries [6], allow a small fraction of answers to be arbitrarily distorted [6], and run significantly more quickly [7].

In their simplest form, such inner product queries require the adversary to be able to "name rows", that is, specify a coefficient for each component of the vector $\mathbf{s}$. Thus, the lower bound does not seem to apply to any functionality that is symmetric in the rows of the database (such as, for example, "counting queries").[2]

Natural Queries. This paper focuses on linear attacks mounted based on the release of natural, symmetric statistics. A first attack along these lines appeared in a previous work of ours (together with J. Ullman) [11] in which we analyzed the release of marginal tables (also called contingency tables). Specifically, in [11], we showed that any mechanism which releases the marginal distributions of all subsets of $k+1$[3] attributes with $o(\sqrt{n})$ noise per entry is attribute non-private if $d = \tilde{\Omega}(n^{1/k})$.[4] These noise bounds were improved in [1], which presented an attack that can tolerate a constant fraction of entries with arbitrarily high noise, as long as the remaining positions have $o(\sqrt{n})$ noise. We generalize both these results in this paper.

Recently, linear attacks were also considered based on range query releases [14] (which, again, are natural, linear queries).

Our Results. We greatly expand the applicability of linear attacks in "natural" settings. Specifically, we show one can mount linear reconstruction attacks based on any release that gives:

1. the fraction of records that satisfy a given *non-degenerate* boolean function (a boolean function over p variables is non-degenerate if its multilinear representation has degree exactly p). Such releases include contingency tables as well as more complex outputs like the error rate of certain classifiers such as decision trees; or

2. the M-estimator associated with a differentiable loss function. M-estimators are a broad class of estimators which are obtained by minimizing sums of functions of the records (they are also called *empirical risk minimization* estimators). M-estimators include the standard estimators for linear and logistic regression (both these estimators are associated with differentiable loss functions). See Section 4 for definitions.

Our contributions are two-fold. First, we show how these types of releases can be transformed into a (noisy) linear release problem, making them amenable to linear reconstruction attacks. This is already perhaps surprising, since many of the above statistics (like M-estimators) are obtained by solving highly nonlinear formulations. After performing this transformation, we can apply polynomial-time methods (like least squares or LP decoding) on this linear release problem to estimate the sensitive data.

Second, we show how to analyze these attacks under various distributional assumptions on the data. This gives lower bounds on the noise needed to release these statistics attribute privately. Specifically, we consider a setting in which the same statistic (either 1 or 2 above) is released about how the sensitive attribute relates to all subsets of (constant) size k (out of a total of d) nonsensitive boolean attributes. For a subset $J \subseteq [d]$ of size k, let $U|_J$ denote the submatrix of U consisting of the columns in J.

[2] It was pointed out in [2] that in databases with more than one entry per row, random inner product queries on the sensitive attribute vector $\mathbf{s}$ can be simulated via hashing: for example, the adversary could ask for the sum the function $H(U_i) \cdot s_i$ over the whole database, where $H : \{0,1\}^{d-1} \to \{0,1\}$ is an appropriate hash function. This is a symmetric statistic, but it is unlikely to come up in a typical statistical publication.

[3] For asymptotic statements, k is considered constant in this paper, as in previous works [11, 1].

[4] The $\tilde{\Omega}$ notation hides polylogarithmic factors.

The $\binom{d}{k}$ entries for a statistic are obtained by evaluating the same statistic on $(U|_J|\mathbf{s})$ for all sets J of size k. Specifically:

- Consider a mechanism which releases, for every set J of size k, the fraction of records (rows) in $(U|_J|\mathbf{s})$ satisfying a non-degenerate boolean function over $k+1$ variables. We show that if the mechanism adds $o(1/\sqrt{n})$ noise per entry and if $d = \tilde{\Omega}(n^{1/k})$, then it is attribute non-private.
- Consider a mechanism which releases, for every set J of size k, a particular M-estimator evaluated over $(U|_J|\mathbf{s})$. We show that if the mechanism adds $o(1/(\lambda\sqrt{n}))$ noise per entry and if $d = \Omega(n)$, then it is attribute non-private. Here, λ is the Lipschitz constant of the loss function gradient. The loss function also needs to satisfy a mild variance condition. For the case of linear and logistic regression estimators, $\lambda = \Theta(1)$ for bounded data, and so the noise bound is $o(1/\sqrt{n})$.

The statements above are based on the least squares attack. For most settings, we show the LP decoding attack can also handle a constant fraction of entries with arbitrarily high noise (the exception is the setting of general M-estimators).

Techniques for Deriving the Attacks. Casting the releases as a system of linear equations requires two simple insights which we hope will be broadly useful. First, we note that when $\mathbf{s} = (s_1, \ldots, s_n)$ is boolean, then any release which allows us to derive an equation which involves a sum over database records can in fact be made linear in $\mathbf{s}$. Specifically, suppose we know that $\sum_i g_i(s_i) = t$, where t is a real number and g_i is an arbitrary real-valued function that could depend on the index i, the public record U_i, and any released information. We can rewrite $g_i(s_i)$ as $g_i(0) + s_i(g_i(1) - g_i(0))$; the constraint $\sum_i g_i(s_i) = t$ can then be written as $\sum_i s_i \cdot (g_i(1) - g_i(0)) = t - \sum_i g_i(0)$, which is affine in $\mathbf{s}$. This allows us to derive linear constraints from a variety of not-obviously-linear releases; for example, it allows us to get linear attacks based on the error rate of a given binary classifier (see Section 3).

The second observation is that for many *nonlinear* optimization problems, every optimal solution must satisfy constraints that are, in fact, sums over data records. For example, for M-estimators associated with differentiable loss functions, the gradient at the solution $\hat{\theta}$ must equal 0, leading to an equation of the form $\sum_i \partial\, \ell(\hat{\theta}; (U_i, s_i)) = 0$. This can be made linear in $\mathbf{s}$ using the first technique. We bound the effect of any noise added to the entries of M-estimator ($\hat{\theta}$) via the Lipschitz properties of the gradient of the loss function ℓ.

Techniques for Analyzing the Attacks. The techniques just mentioned give rise to a variety of linear reconstruction attacks, depending on the type of released statistics. We can provide theoretical guarantees on the performance of these attacks in some settings, for example when the same statistic is released about many subsets of the data (*e.g.,* all sets of a given size k) and when the data records themselves are drawn i.i.d. from some underlying distribution. The main technique here is to analyze the geometry of the constraint matrix A that arises in the attack. For the case of non-degenerate boolean functions, we do so by relating the constraint matrix to a row-wise product of a matrix with i.i.d. entries (referred to as a random row product matrix, see Section 3.2.1), which was recently analyzed by Rudelson [17] (see also [11]). The results of [17] showed that the least singular value of a random row product matrix is asymptotically the same as that of a matrix of same dimensions with i.i.d. entries, and a random row product matrix is a Euclidean section. Our results show that a much broader class of matrices with correlated rows satisfy these properties.

Organization. In Section 2, we introduce some notation and review the least squares and LP decoding techniques for solving noisy linear systems. In Section 3, we present our results on evaluating non-

degenerate boolean functions. As mentioned earlier, we first reduce the release problem to a linear reconstruction problem (Section 3.1), and then the attacks works by using either least squares or LP decoding techniques. The analysis requires analyzing spectral and geometric properties of the constraint matrix that arises in these attack which we do in Section 3.2. In Section 4, we present our results on releasing M-estimators associated with differentiable loss functions. For clarity, we first discuss the attacks for the special cases of linear and logistic regression estimators (in Section 4.1), and then discuss the attacks for the general case (in Section 4.2).

2 Preliminaries

Notation. We use $[n]$ to denote the set $\{1, \ldots, n\}$. $d_H(\cdot, \cdot)$ measures the Hamming distance. Vectors used in the paper are by default column vectors and are denoted by boldface letters. For a vector $\mathbf{v}$, $\mathbf{v}^\top$ denotes its transpose, $\|\mathbf{v}\|$ denotes its Euclidean norm, $\|\mathbf{v}\|_1$ denotes its ℓ_1 norm, and $\|\mathbf{v}\|_\infty$ denotes its ℓ_∞ norm. For two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ denotes the inner product of $\mathbf{v}_1$ and $\mathbf{v}_2$. We use $(a)_n$ to denote a vector of length of n with all entries equal to a. For a matrix M, $\|M\|$ denotes the operator norm and M_i denotes the ith row of M. Random matrices are denoted by boldface capitalized letters. We use $diag(a_1, \ldots, a_n)$ to denote an $n \times n$ diagonal matrix with entries $a_1, \ldots, a_n$ along the main diagonal. The notation vert$(\cdot, \ldots, \cdot)$ denotes vertical concatenation of the argument matrices.

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 least singular value $\sigma_n(M) = \inf_{\mathbf{z}:\|\mathbf{z}\|=1} \|M\mathbf{z}\|$. The unit sphere in n dimensions centered at origin is denoted by $S^{n-1} = \{\mathbf{z} : \|\mathbf{z}\| = 1\}$.

Our analysis uses random matrices, and we add a subscript of r to differentiate a random matrix from a non-random matrix. As mentioned earlier, k is a constant in this paper and we often omit dependence on k in our results.

2.1 Background on Noisy Linear Systems.

Noisy linear systems arise in a wide variety of statistical and signal-processing contexts. Suppose we are given a matrix A and vector $\mathbf{z}$ such that $\mathbf{z} = A\mathbf{s} + \mathbf{e}$, where $\mathbf{e}$ is assumed to be "small" (in a sense defined below). A natural approach to estimating $\mathbf{s}$ is to output $\hat{\mathbf{s}} = \text{argmin}_{\mathbf{s}} \|A\mathbf{s} - \mathbf{z}\|_p$ for some $p \geq 1$. We will consider $p = 1$ and 2; we summarize the assumptions and guarantees for each method below. When it is known that $\mathbf{s} \in \{0, 1\}^n$, the attacker can then round the entries of $\hat{\mathbf{s}}$ to the nearer of $\{0, 1\}$ to improve the estimate.

In the sequel, we call a vector $\mathbf{z} \in \mathbb{R}^m$ (a, b)-*small* if at least $1 - a$ fraction of its entries have magnitude less than b. In other words, for some set S, $|S| \geq (1 - a) \cdot m$, it is the case that $|z_i| \leq b$ for all $i \in S$.

ℓ_2 error minimization ("least squares"). Widely used in regression problems, the least squares method guarantees a good approximation to $\mathbf{s}$ when the Euclidean norm $\|\mathbf{e}\|$ is small and A has no small eigenvalues. It was first used in data privacy by Dwork and Yekhanin [7]. For completeness, we present the entire analysis of this attack (in a general setting) here.

Let $A = P\Sigma Q^\top$ be the singular value decomposition of A. Here, P is an orthogonal $m \times m$ matrix, Σ is a diagonal $m \times n$ matrix, and Q is an orthogonal $n \times n$ matrix. Let $\mathbf{0}_{n\times(m-n)}$ be an $n \times (m - n)$ matrix with all entries zero. Define

$$\Gamma_{\text{inv}} = (diag(\sigma_n(A)^{-1}, \ldots, \sigma_1(A)^{-1}) | \mathbf{0}_{n\times(m-n)}).$$

The dimension of Γ_{inv} is $n \times m$. Define $A_{\text{inv}} = Q\Gamma_{\text{inv}}P^\top$.

Given $\mathbf{y}$, the adversary uses A_{inv} to construct $\hat{\mathbf{s}} = (\hat{s}_1, \ldots, \hat{s}_n)$ as follows:

$$\hat{s}_i = \begin{cases} 0 & \text{if } i\text{th entry in } A_{\text{inv}}\mathbf{z} < 1/2, \\ 1 & \text{otherwise.} \end{cases}$$

In other words, $\hat{\mathbf{s}}$ is obtained by rounding $A_{\text{inv}}\mathbf{z}$ to closest of $0, 1$.

Now the claim is that $\hat{\mathbf{s}}$ is a good reconstruction of $\mathbf{s}$. The idea behind the analysis is that $A_{\text{inv}}\mathbf{z} = \mathbf{s} + A_{\text{inv}}\mathbf{e}$. Now (as P and Q are orthogonal matrices, they don't affect the norms),

$$\|A_{\text{inv}}\mathbf{e}\| = \|Q\Gamma_{\text{inv}}P^\top\mathbf{e}\| \leq \|\Gamma_{\text{inv}}\|\|P^\top\mathbf{e}\| = \frac{\|\mathbf{e}\|}{\sigma_n(A)}.$$

Let us assume that $\sigma_n(A) = \sigma$. If (the absolute value of) all the entries in $\mathbf{e}$ are less than β then $\|\mathbf{e}\| = \beta\sqrt{m}$, and therefore $\|A_{\text{inv}}\mathbf{e}\| = (\beta\sqrt{m})/\sigma$. In particular, this implies that $A_{\text{inv}}\mathbf{e}$ cannot have $(4m\beta^2)/\sigma^2$ entries with absolute value above $1/2$, and therefore the Hamming distance between $\hat{\mathbf{s}}$ and $\mathbf{s}$ is $O(4m\beta^2)/\sigma^2$ (as the adversary only fails to recover those entries of $\mathbf{s}$ whose corresponding $A_{\text{inv}}\mathbf{e}$ entries are greater than $1/2$). The time complexity of the attack is dominated by the cost of computing the singular value decomposition of A which takes $O(mn^2)$ time.[5]

Theorem 2.1. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a full rank linear map with $(m > n)$ such that the least singular value of A is σ. Then if $\mathbf{e}$ is $(0, \beta)$ small (that is, if $\|\mathbf{e}\|_\infty \leq \beta$), the vector $\text{argmin}_{\mathbf{s}} \|A\mathbf{s} - \mathbf{z}\|$, rounded to $\{0,1\}^n$, satisfies $d_H(\mathbf{s}, \hat{\mathbf{s}}) \leq (4m\beta^2)/\sigma^2$. In particular, if $\sigma = \Omega(\sqrt{m})$ and $\beta = o(\sqrt{n})$, then $\hat{\mathbf{s}}$ agrees with a $1 - o(1)$ fraction of $\mathbf{s}$. The attack runs in $O(mn^2)$ time.*

ℓ_1 error minimization ("LP decoding"). In the context of privacy, the "LP decoding" approach was first used by Dwork *et al.* [6]. (The name stems from the fact that the minimization problem can be cast as a linear program.) The LP attack is slower than the least squares attack but can handle considerably more complex error patterns at the cost of a stronger assumption on A. Recently, De [1] gave a simple analysis of this attack based on the geometry of the operator A. We need the following definition of Euclidean section.

Definition 2. *A linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a α-Euclidean section if for all $\mathbf{s}$ in $\mathbb{R}^n$,*

$$\sqrt{m}\|A\mathbf{s}\| \geq \|A\mathbf{s}\|_1 \geq \alpha\sqrt{m}\|A\mathbf{s}\|.$$

Note that by Cauchy-Schwarz, the first inequality, $\sqrt{m}\|A\mathbf{s}\| \geq \|A\mathbf{s}\|_1$, always holds. We remark that when we say A is Euclidean, we simply mean that there is some constant $\alpha > 0$ such that A is α-Euclidean.

The following theorem gives a sufficient condition under which LP decoding gives a good approximation to $\mathbf{s}$. The time bound here was derived from the LP algorithm of Vaidya [21], which uses $O(((N_1+N_2)N_2^2 + (N_1 + N_2)^{1.5}N_2)N_3)$ arithmetic operations where N_1 is the number of constraints, N_2 is the number of variables, and N_3 is a bound on the number of bits used to describe the entries of A. In our setting, the LP has n variables, m constraints, and N_3 could be upper bounded by mn, and therefore the LP could be solved in $O(m^2n^3 + m^{2.5}n^2)$ time.

Theorem 2.2 (From [1])**.** *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a full rank linear map $(m > n)$ such that the least singular value of A is σ. Further, let A be a α-Euclidean section. Then there exists a $\gamma = \gamma(\alpha)$ such that if $\mathbf{e}$ is (γ, β) small, then any solution $\hat{\mathbf{s}} = \text{argmin}_{\mathbf{s}} \|A\mathbf{s} - \mathbf{z}\|_1$, rounded to $\{0,1\}^n$, satisfies $d_H(\mathbf{s}, \hat{\mathbf{s}}) \leq O(\beta\sqrt{mn}/\sigma)$ where the constant inside the $O(\cdot)$ notation depends on α. In particular, if $\sigma = \Omega(\sqrt{m})$ and $\beta = o(\sqrt{n})$, then this attack recovers $1 - o(1)$ fraction of $\mathbf{s}$. The attack runs in $O(m^2n^3 + m^{2.5}n^2)$ time.*

[5] SVD decomposition of $N_1 \times N_2$ sized matrix can be done in $O(N_1N_2^2)$ time.

3 Releasing Evaluations of Non-Degenerate Boolean Functions

In this section, we analyze privacy lower bounds for releasing evaluations of non-degenerate boolean functions. We use the following standard definition of representing polynomial (see Appendix A for a background about representing boolean functions as multilinear polynomials).

Definition 3. *A polynomial $P^{(f)}$ over the reals represents a function f over $\{0,1\}^{k+1}$ if $f(x_1,\ldots,x_{k+1}) = P^{(f)}(x_1,\ldots,x_{k+1})$ for all $(x_1,\ldots,x_{k+1}) \in \{0,1\}^{k+1}$.*

Definition 4. *A function $f : \{0,1\}^{k+1} \to \{0,1\}$ is non-degenerate iff it can be written as a multilinear polynomial of degree $k+1$.*

If f is non-degenerate, then it depends on all of its $k+1$ variables. Note that non-degenerate functions constitute a large class of functions.[6] For example, it includes widely used boolean functions like AND, OR, XOR, MAJORITY, and depth $k+1$ decision trees [15].

Problem Statement. Let $f : \{0,1\}^{k+1} \to \{0,1\}$ represent the function that we want to evaluate on a database D. Let $D = (U|\mathbf{s}) \in (\{0,1\})_{n\times(d+1)}$, where $U \in (\{0,1\})_{n\times d}$ and $\mathbf{s} \in \{0,1\}^n$. Let $U = (\delta_{i,j})$, i.e., $\delta_{i,j}$ denotes the (i,j)th entry in U (with $1 \le i \le n$ and $1 \le j \le d$). Let

$$J = (j_1,\ldots,j_k) \in \{1,\ldots,d\}^k$$
$$(\text{where } \{1,\ldots,d\}^k = \underbrace{\{1,\ldots,d\} \times \cdots \times \{1,\ldots,d\}}_{k \text{ times}}).$$

Note that J allows repeated entries.[7] Let $D|_J$ be the submatrix of D restricted to columns indexed by J. For a fixed J, define $F(D|_J)$ as

$$F(D|_J) = \sum_{i=1}^{n} f(\delta_{i,j_1},\ldots,\delta_{i,j_k},s_i), \quad J = (j_1,\ldots,j_k).$$

Note that $F(D|_J)$ is an integer between 0 to n. Let $\Sigma_f(D)$ be the vector obtained by computing F on all different $D|_J$'s: $\Sigma_f(D) = (F(D|_J))$ where $J \in \{1,\ldots,d\}^k$. Note that $\Sigma_f(D)$ is a vector of length d^k. The goal is to understand how much noise is needed to attribute privately release $\Sigma_f(D)$ (or $\Sigma_f(D)/n$) when f is non-degenerate.

Our Results. We prove the following results using the ℓ_2 and ℓ_1 error minimization attacks outlined in Section 2.1.

Theorem 3.1 (Informal Statements)**.** *Let $f : \{0,1\}^{k+1} \to \{0,1\}$ be a non-degenerate boolean function. Then*

1. *any mechanism which for every database $D \in (\{0,1\})_{n\times(d+1)}$ with $n \ll d^k$ releases $\Sigma_f(D)$ by adding $o(\sqrt{n})$ (or releases $\Sigma_f(D)/n$ by adding $o(1/\sqrt{n})$) noise to each entry is attribute non-private. The attack that achieves this non-privacy violation runs in $O(d^k n^2)$ time.*

[6] A simple counting argument shows that among the $2^{2^{k+1}}$ boolean functions over $k+1$ variables, $2^{2^{k+1}} - \binom{2^{k+1}}{2^k}$ are non-degenerate.

[7] We allow repeated entries for convenience of notation. Our results also hold if we use the more natural $J \subseteq [d]$, $|J| = k$.

2. *there exists a constant $\gamma > 0$ such that any mechanism which for every database $D \in (\{0,1\})_{n\times(d+1)}$ with $n \ll d^k$ releases $\Sigma_f(D)$ by adding $o(\sqrt{n})$ (or releases $\Sigma_f(D)/n$ by adding $o(1/\sqrt{n})$) noise to at most $1-\gamma$ fraction of the entries is attribute non-private. The attack that achieves this non-privacy violation runs in $O(d^{2k}n^3 + d^{2.5k}n^2)$ time.*

For convenience of notation, in this section, we work mostly with the transpose of U. Let $T = U^\top$. So T is a $d \times n$ matrix.

3.1 Reducing to a Linear Reconstruction Problem.

In this section, we reduce the problem of releasing $\Sigma_f(D)$ for a database D into a linear reconstruction problem. First, we define a simple decomposition of boolean functions. Consider a non-degenerate boolean function $f : \{0,1\}^{k+1} \to \{0,1\}$. Now there exists two function $f_0 : \{0,1\}^k \to \{0,1\}$ and $f_1 : \{0,1\}^k \to \{0,1\}$ such that

$$f(\delta_1, \ldots, \delta_{k+1}) = f_0(\delta_1, \ldots, \delta_k)(1-\delta_{k+1}) + f_1(\delta_1, \ldots, \delta_k)\delta_{k+1} \quad \forall (\delta_1, \ldots, \delta_{k+1}) \in \{0,1\}^{k+1}.$$

This can be re-expressed as

$$f(\delta_1, \ldots, \delta_{k+1}) = f_0(\delta_1, \ldots, \delta_k) + (f_1(\delta_1, \ldots, \delta_k) - f_0(\delta_1, \ldots, \delta_k))\delta_{k+1}.$$

Define $f_2(\delta_1, \ldots, \delta_k) = f_1(\delta_1, \ldots, \delta_k) - f_0(\delta_1, \ldots, \delta_k)$. Therefore,

$$f(\delta_1, \ldots, \delta_{k+1}) = f_0(\delta_1, \ldots, \delta_k) + f_2(\delta_1, \ldots, \delta_k)\delta_{k+1}. \tag{1}$$

Note that f_2 is a function from $\{0,1\}^k \to \{-1,0,1\}$. Since both f_0 and f_1 are both boolean functions and can be represented as multilinear polynomials over the variables $\delta_1, \ldots, \delta_k$, therefore f_2 also could be represented as a multilinear polynomial over the variables $\delta_1, \ldots, \delta_k$. Since f is represented by a multilinear polynomial of degree $k+1$, therefore, the multilinear polynomial representing f_2 has degree k (if it has any lower degree, then f could be represented as multilinear polynomial of degree strictly less than $k+1$, which is a contradiction). To aid our construction, we need to define a particular function of matrices.

Definition 5 (Row Function Matrix)**.** *Let h be a function from $\{0,1\}^k \to \{-1,0,1\}$. Let $T_{(1)} = (\delta^{(1)}_{i,j}), T_{(2)} = (\delta^{(2)}_{i,j}), \ldots, T_{(k)} = (\delta^{(k)}_{i,j})$ be k matrices with $\{0,1\}$ entries and dimensions $d \times n$. Define a row function matrix (of dimension $d^k \times n$) $\Pi_h(T_{(1)}, \ldots, T_{(k)})$ as follows. Any row of this matrix will correspond to a sequence*

$$J = (j_1, j_2, \ldots, j_k) \in \{1, \ldots, d\}^k$$

of k numbers, so the entries of $\Pi_h(T_{(1)}, \ldots, T_{(k)})$ will be denoted[8] by $\pi_{J,a}$, where $a \in \{1, \ldots, n\}$. For $J = (j_1, j_2, \ldots, j_k)$ the entries of the matrix $\Pi_h(T_{(1)}, \ldots, T_{(k)})$ will be defined by the relation

$$\pi_{J,a} = h(\delta^{(1)}_{j_1,a}, \delta^{(2)}_{j_2,a}, \ldots, \delta^{(k)}_{j_k,a}).$$

The row product matrices from [11] (see Definition 6) is a particular example of this construction where the function $h(\delta_1, \delta_2, \ldots, \delta_k) = \delta_1 \cdot \delta_2 \cdot \ldots \cdot \delta_k$,, which implies that $\Pi_h(T_{(1)}, \ldots, T_{(k)}) = T_{(1)} \odot \cdots \odot T_{(k)}$, where $\odot$ is the row-product operator from Definition 6.

[8]The definition assumes a certain order of the rows of the matrix $\Pi_h(T_{(1)}, \ldots, T_{(k)})$. This order, however, is not important, for our analysis. Note that changing the relative positions of rows of a matrix doesn't affect its eigenvalues and singular values.

Let $D = (U|\mathbf{s})$ be a database, and let $T = U^\top$. Let $U = (\delta_{i,j})$. Consider any fixed $J = (j_1, \ldots, j_k) \in \{1, \ldots, d\}^k$. Now for this J, there exists an entry in $\Sigma_f(D)$ equaling $\sum_{i=1}^n f(\delta_{i,j_1}, \ldots, \delta_{i,j_k}, s_i)$. Now consider the matrices $\Pi_{f_2}(T, \ldots, T)$ and $\Pi_{f_0}(T, \ldots, T)$. Consider the rows in $\Pi_{f_0}(T, \ldots, T)$ and $\Pi_{f_2}(T, \ldots, T)$ corresponding to this above J. Let this be the lth row in these matrices. Then the lth row of the matrix $\Pi_{f_0}(T, \ldots, T)$ has n entries equaling $f_0(\delta_{i,j_1}, \ldots, \delta_{i,j_k})$ and the lth row of the matrix $\Pi_{f_2}(T, \ldots, T)$ has n entries equaling $f_2(\delta_{i,j_1}, \ldots, \delta_{i,j_k})$ for $i = 1, \ldots, n$. Since

$$f(\delta_{i,j_1}, \ldots, \delta_{i,j_k}, s_i) = f_0(\delta_{i,j_1}, \ldots, \delta_{i,j_k}) + f_2(\delta_{i,j_1}, \ldots, \delta_{i,j_k}) s_i,$$

it follows that

$$\begin{aligned}\sum_{i=1}^n f(\delta_{i,j_1}, \ldots, \delta_{i,j_k}, s_i) &= \sum_{i=1}^n f_0(\delta_{i,j_1}, \ldots, \delta_{i,j_k}) + f_2(\delta_{i,j_1}, \ldots, \delta_{i,j_k}) s_i \\ &= \langle \Pi_{f_0}(T, \ldots, T)_l, \mathbf{1}_n \rangle + \langle \Pi_{f_2}(T, \ldots, T)_l, \mathbf{s} \rangle,\end{aligned}$$

where $\Pi_{f_0}(T, \ldots, T)_l$ and $\Pi_{f_2}(T, \ldots, T)_l$ denote the lth row of matrices $\Pi_{f_0}(T, \ldots, T)$ and $\Pi_{f_2}(T, \ldots, T)$ respectively and $\mathbf{1}_n$ denotes the vector $(1)^n$. Now define a vector $H_f(D)$ whose lth element ($1 \leq l \leq d^k$) is

$$H_f(D)_l = \langle \Pi_{f_0}(T, \ldots, T)_l, \mathbf{1}_n \rangle + \langle \Pi_{f_2}(T, \ldots, T)_l, \mathbf{s} \rangle. \tag{2}$$

The length of vector $H_f(D)$ is d^k. The above arguments show that all the entries of $H_f(D)$ are contained in the vector $\Sigma_f(D)$. Since every row in these $\Sigma_f(D)$ correspond to some J, it also follows that all the entries of $\Sigma_f(D)$ are contained in the vector $H_f(D)$, implying the following claim.

Claim 1. $\Sigma_f(D) = H_f(D)$.[9]

Setting up the Least Squares Attack. The privacy mechanism releases a noisy approximation to $\Sigma_f(D)$. Let $\mathbf{y} = (y_1, \ldots, y_{d^k})$ be this noisy approximation. The adversary tries to reconstruct an approximation of $\mathbf{s}$ from $\mathbf{y}$. Let $b_{f_i} = \langle \Pi_{f_0}(T, \ldots, T)_i, \mathbf{1}_n \rangle$, and $\mathbf{b}_f = (b_{f_1}, \ldots, b_{f_{d^k}})$. Given $\mathbf{y}$, the adversary solves the following linear reconstruction problem:

$$\mathbf{y} = \mathbf{b}_f + \Pi_{f_2}(T, \ldots, T)\mathbf{s}. \tag{3}$$

In the setting of attribute non-privacy the adversary knows T, and therefore can compute $\Pi_{f_2}(T, \ldots, T)$ and $\Pi_{f_0}(T, \ldots, T)$ (hence, $\mathbf{b}_f$). The goal of the adversary is to compute a large fraction of $\mathbf{s}$ given $\mathbf{y}$. The definition of iterated logarithm ($\log_{(q)}$) is given in Definition 8. In the below analysis, we use a random matrix $\mathbf{T}$ and the least singular value lower bound on a random row function matrix from Theorem 3.9. We use boldface letters to denote random matrices.

Theorem 3.2 (Part 1, Theorem 3.1)**.** *Let $f : \{0,1\}^{k+1} \to \{0,1\}$ be a non-degenerate boolean function and $n \leq cd^k / \log_{(q)} d$ for a constant c depending only on k, q. Any privacy mechanism which for every database $D \in (\{0,1\})_{n \times (d+1)}$ releases $\Sigma_f(D)$ by adding $o(\sqrt{n})$ (or releases $\Sigma_f(D)/n$ by adding $o(1/\sqrt{n})$) noise to each entry is attribute non-private. The attack that achieves this non-privacy violation runs in $O(d^k n^2)$ time.*

Proof. Consider the least squares attack outlined in Theorem 2.1 on Equation (3). Let $\mathbf{T}$ be a random matrix of dimension $d \times n$ with independent Bernoulli entries taking values 0 and 1 with probability $1/2$, and let database $D = (\mathbf{T}^\top|\mathbf{s})$ for some $\mathbf{s} \in \{0,1\}^n$. For analyzing the attack in Theorem 2.1, we need a lower bound on the least singular value of $\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})$. The following claim follows from Theorem 3.9. Note that f_2 is a function over k variables.

[9] Under proper ordering of both the vectors.

Claim 2. *For function f_2 defined above and $n \leq cd^k/\log_{(q)} d$ (where c is the constant from Theorem 3.8), the matrix $\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})$ satisfies*

$$\Pr\left[\sigma_n(\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})) \leq C'\sqrt{d^k}\right] \leq c_1 \exp\left(-c_2 d\right).$$

Proof. Apply Theorem 3.9 with function $h = f_2$. □

Claim 2 shows that with exponentially high probability $\sigma_n(\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})) = \Omega(\sqrt{d^k})$. Invoking Theorem 3.9 with $m = d^k$ and $\beta = o(\sqrt{n})$ shows that with exponentially high probability the adversary fails to recover only $o(n)$ entries of $\mathbf{s}$. Noise bound of $o(\sqrt{n})$ for releasing $\Sigma_f(D)$ translates into a noise bound of $o(1/\sqrt{n})$ for releasing $\Sigma_f(D)/n$. □

Setting up the LP Decoding Attack. The LP decoding attack solves a slightly different reconstruction problem than Equation (3). The reason is because $\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})$ is not a Euclidean section[10] (a property needed for applying Theorem 2.2). However, we show that a related reconstruction problem has all the properties needed for the LP decoding attack. The analysis goes via matrices with $\{1, -1\}$ entries which have the desired properties. We establish the Euclidean section property in Appendix B.

Let $D = (U|\mathbf{s})$ be a database, and let $T = U^\top$. Let $V = 2T - \mathbf{1}_{d\times n}$ where $\mathbf{1}_{d\times n}$ is a $d \times n$ matrix of all 1's. Define $g : \{-1, 1\}^{k+1} \to \{-1, 1\}$ as

$$g(\phi_1, \ldots, \phi_{k+1}) = 2f\left(\frac{1+\phi_1}{2}, \ldots, \frac{1+\phi_{k+1}}{2}\right) - 1. \tag{4}$$

We can decompose g as

$$g(\phi_1, \ldots, \phi_{k+1}) = \left(\frac{1+\phi_{k+1}}{2}\right) g_1(\phi_1, \ldots, \phi_k) + \left(\frac{1-\phi_{k+1}}{2}\right) g_{-1}(\phi_1, \ldots, \phi_k),$$

where $g_1 : \{-1, 1\}^k \to \{-1, 1\}$ and $g_{-1} : \{-1, 1\}^k \to \{-1, 1\}$. Using the notation, $g_1 = g_1(\phi_1, \ldots, \phi_k)$ and $g_{-1} = g_{-1}(\phi_1, \ldots, \phi_k)$, we get

$$g(\phi_1, \ldots, \phi_{k+1}) = (g_1 + g_{-1})/2 + \phi_{k+1}(g_1 - g_{-1})/2.$$

Define $g_2 = g_2(\phi_1, \ldots, \phi_k) = (g_1 - g_{-1})/2$ and $g_3 = g_3(\phi_1, \ldots, \phi_k) = (g_1 + g_{-1})/2$. Let us denote $\delta_i = (1 + \phi_i)/2$. Using the decomposition of f from Equation (1),

$$f\left(\frac{1+\phi_1}{2}, \ldots, \frac{1+\phi_{k+1}}{2}\right) = f(\delta_1, \ldots, \delta_{k+1}) = f_0(\delta_1, \ldots, \delta_k) + f_2(\delta_1, \ldots, \delta_k)\delta_{k+1}.$$

Using the notation, $f_0 = f_0(\delta_1, \ldots, \delta_k)$ and $f_2 = f_2(\delta_1, \ldots, \delta_k)$, and substituting the decomposition of f and g into Equation (4) gives:

$$g_3 + g_2\phi_{k+1} = 2(f_0 + f_2\delta_{k+1}) - 1 \equiv f_0 + f_2\delta_{k+1} = (1/2)(g_3 + g_2\phi_{k+1} + 1). \tag{5}$$

[10] If we take a matrix $\mathbf{T}$ of dimension $d \times n$ with independent Bernoulli entries taking values 0 and 1 with probability $1/2$, the resulting matrix $\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})$ is not a Euclidean section. This is because the matrix $\mathbf{T}$ is non-centered (expectation of each entry in the matrix is $1/2$) which makes the $\|\Pi_{f_2}(\mathbf{T}, \ldots, \mathbf{T})\|$ to be $\approx d^k$ instead of $\sqrt{d^k}$.

Claim 3. *Let $D = (U|\mathbf{s})$, $T = U^\top$, and $V = 2T - \mathbf{1}_{d\times n}$. Then*

$$\Sigma_f(D)_i = \frac{1}{2}\left(\langle \Pi_{g_3}(V,\ldots,V)_i, \mathbf{1}_n\rangle + \langle \Pi_{g_2}(V,\ldots,V)_i, 2\mathbf{s} - \mathbf{1}_n\rangle + n\right).$$

Proof. Note that from the earlier established decomposition

$$\Sigma_f(D)_i = \langle \Pi_{f_0}(T,\ldots,T)_i, \mathbf{1}_n\rangle + \langle \Pi_{f_2}(T,\ldots,T)_i, \mathbf{s}\rangle.$$

The reminder of the proof follows by using Equation (5) along with the definition of row-function matrices (Definition 5). □

Define

$$q_{g_i} = \frac{\langle \Pi_{g_3}(V,\ldots,V)_i, \mathbf{1}_n\rangle}{2} + \frac{n}{2} - \frac{\langle \Pi_{g_2}(V,\ldots,V)_i, \mathbf{1}_n\rangle}{2}.$$

Then $\Sigma_f(D)_i = q_{g_i} + \langle \Pi_{g_2}(V,\ldots,V)_i, \mathbf{s}\rangle$. Let $\mathbf{q}_g = (q_{g_1},\ldots,q_{g_{d^k}})$. We get

$$\Sigma_f(D) = \mathbf{q}_g + \Pi_{g_2}(V,\ldots,V)\mathbf{s}. \tag{6}$$

Let $\mathbf{y}$ be the noisy approximation to $\Sigma_f(D)$ released by the privacy mechanism. Given $\mathbf{y}$, the linear program that the adversary solves is:

$$\text{argmin}_{\mathbf{s}} \left\|\mathbf{y} - \mathbf{q}_g - \Pi_{g_2}(V,\ldots,V)\mathbf{s}\right\|_1. \tag{7}$$

The following theorem analyzes this attack using a random matrix $\mathbf{V}$.

Theorem 3.3 (Part 2, Theorem 3.1)**.** *Let $f : \{0,1\}^{k+1} \to \{0,1\}$ be a non-degenerate boolean function and let $n \le cd^k/\log_{(q)} d$ for a constant c depending only on k, q. Then there exists a constant $\gamma = \gamma(k,q) > 0$ such that any mechanism which for every database $D \in (\{0,1\})_{n\times(d+1)}$ releases $\Sigma_f(D)$ by adding $o(\sqrt{n})$ (or releases $\Sigma_f(D)/n$ by adding $o(1/\sqrt{n})$) noise to at most $1-\gamma$ fraction of the entries is attribute non-private. The attack that achieves this non-privacy violation runs in $O(d^{2k}n^3 + d^{2.5k}n^2)$ time.*

Proof. The proof uses the LP decoding attack outlined in Theorem 2.2 on Equation (7). Let $\mathbf{T}$ be a random matrix of dimension $d \times n$ with independent Bernoulli entries taking values 0 and 1 with probability $1/2$, and let database $D = (\mathbf{T}^\top|\mathbf{s})$ for some $\mathbf{s} \in \{0,1\}^n$. Let $\mathbf{V} = 2\mathbf{T} - \mathbf{1}_{d\times n}$. To use Theorem 2.2, we need to (i) establish that $\Pi_{g_2}(\mathbf{V},\ldots,\mathbf{V})$ is a Euclidean section and (ii) establish a lower bound on its least singular value. Since $g_2 : \{-1,1\}^k \to \{-1,0,1\}$ has a representation as a multilinear polynomial of degree k, Theorem B.4 shows that $\Pi_{g_2}(\mathbf{V},\ldots,\mathbf{V})$ is with exponentially high probability a Euclidean section. Repeating an analysis similar to Theorem 3.9 shows that the least singular value of $\Pi_{g_2}(\mathbf{V},\ldots,\mathbf{V})$ is with exponentially high probability at least $\Omega(\sqrt{d^k})$. Hence, with exponentially high probability both of the following statements hold simultaneously: (i) $\Pi_{g_2}(\mathbf{V},\ldots,\mathbf{V})$ is a Euclidean section and (ii) $\sigma_n(\Pi_{g_2}(\mathbf{V},\ldots,\mathbf{V})) = \Omega(\sqrt{d^k})$.

Invoking Theorem 2.2 with $\beta = o(\sqrt{n})$, $a = \sqrt{n}$, $b = d^k$, and $\sigma = \Omega(\sqrt{d^k})$ shows that with exponentially high probability the adversary fails to recover only $o(n)$ entries of $\mathbf{s}$. This shows that the mechanism is attribute non-private.

In the running time analysis of Theorem 2.2, m gets replaced by d^k, and N_3 by $d^k n$ (as the input matrix can be represented using $O(d^k n)$ bits). Noise bound of $o(\sqrt{n})$ for releasing $\Sigma_f(D)$ translates into a noise bound of $o(1/\sqrt{n})$ for releasing $\Sigma_f(D)/n$. □

3.2 Spectral and Geometric Properties of Random Row Function Matrices.

Analysis of our reconstruction attacks rely on spectral and geometric properties of random row function matrices that we discuss in this section. Rudelson [17] and Kasiviswanathan *et al.* [11] analyzed certain spectral and geometric properties of a certain class of correlated matrices that they referred to as row product matrices (or conjunction matrices). Our analysis builds upon these results. We first summarize some important definitions and useful results from [17, 11] in Section 3.2.1, and establish our least singular value bound in Section 3.2.2. The Euclidean section property is established in Appendix B.

3.2.1 Spectral and Geometric Properties of Random Row Product Matrices.

For two matrices with the same number of columns we define the row product as a matrix whose rows consist of entry-wise product of the rows of original matrices.

Definition 6 (Row Product Matrix)**.** *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 $T_{(1)}$ is an $N_1 \times n$ matrix, and $T_{(2)}$ is an $N_2 \times n$ matrix, denote by $T_{(1)} \odot T_{(2)}$ an $N_1 N_2 \times n$ matrix, whose rows are entry-wise products of the rows of $T_{(1)}$ and $T_{(2)}$: $(T_{(1)} \odot T_{(2)})_{j,k} = T_{1_j} \odot T_{2_k}$, where $(T_{(1)} \odot T_{(2)})_{j,k}, T_{1_j}, T_{2_k}$ denote rows of the corresponding matrices.*[11]

Rudelson [17] showed that if we take entry-wise product of k independent random matrices of dimension $d \times n$, then the largest and the least singular values of the resulting row product matrix (which of dimension $d^k \times n$) is asymptotically the same order as that of a $d^k \times n$ matrix with i.i.d. entries. To formulate this result formally, we introduce a class of uniformly bounded random variables, whose variances are uniformly bounded below.

Definition 7 (τ-random variable and matrix)**.** *Let $\tau > 0$. We will call a random variable ξ a τ-random variable if $|\xi| \leq 1$ a.s., $\mathbb{E}[\xi] = 0$, and $\mathbb{E}[\xi^2] \geq \tau^2$. A matrix $\mathbf{M}$ is called a τ-random matrix if all its entries are independent τ-random.*

We would also need the notion of iterated logarithm ($\log_{(q)}$) that is defined as:

Definition 8. *For $q \in \mathbb{N}$, define the function $\log_{(q)} : (0, \infty) \to \mathbb{R}$ by induction.*

1. $\log_{(1)} t = \max(\log t, 1)$;
2. $\log_{(q+1)} t = \log_{(1)}(\log_{(q)} t)$.

We are now ready to state the main result from [17] that establishes a lower bound on the ℓ_1 norm of $(\mathbf{T}_{(1)} \odot \cdots \odot \mathbf{T}_{(k)})\mathbf{x}$ where each $\mathbf{T}_{(i)}$ is an independent τ-random matrix and $\mathbf{x}$ is a unit vector.

Theorem 3.4 ([17])**.** *Let k, q, n, d be natural numbers. Assume that $n \leq cd^k / \log_{(q)} d$. Let $\mathbf{T}_{(1)}, \ldots, \mathbf{T}_{(k)}$ be k matrices with independent τ-random entries and dimensions $d \times n$. Then the k-times entry-wise product $\mathbf{T}_{(1)} \odot \mathbf{T}_{(2)} \odot \cdots \odot \mathbf{T}_{(k)}$ is a $d^k \times n$ matrix satisfying*

$$\Pr\left[\exists \mathbf{x} \in S^{n-1} \;\; \|(\mathbf{T}_{(1)} \odot \cdots \odot \mathbf{T}_{(k)})\mathbf{x}\|_1 \leq C_1 d^k\right] \leq c_1 \exp(-C_2 d).$$

The constants c, C_1, c_1, C_2 depends only on k and q.

One of the main ingredients in proving the above theorem is following fact about the norm of row product matrices.

[11] When $N_1 = N_2$, the row product matrix is also called the Hadamard product matrix.

Theorem 3.5 ([17]). *Let* $\mathbf{T}$ *be a matrix with independent* τ*-random entries and dimension* $d \times n$*. Then the* k*-times entry-wise product* $\mathbf{T} \odot \cdots \odot \mathbf{T}$ *is a* $d^k \times n$ *matrix satisfying*

$$\Pr\left[\|\mathbf{T} \odot \cdots \odot \mathbf{T}\| \geq c_3\left(\sqrt{d^k} + \sqrt{n}\right)\right] \leq \exp\left(-c_4 n^{\frac{1}{12k}}\right).$$

The constants c_3, c_4 *depends only on* k*.*

The bound on the norm appearing in the above theorem (asymptotically) matches that for a $d^k \times n$ matrix with independent τ-random entries (refer [22] for more details).

3.2.2 Least Singular Value of Random Row Function Matrices.

We start by proving a simple proposition about functions that can be represented as multilinear polynomials. The main step behind the following proposition proof is the following simple fact about multilinear polynomials. Let $P^{(h)}$ be a multilinear polynomial representing function h, and let $(\delta_1, \ldots, \delta_k) \in \{0,1\}^k$ and $\delta_i' \in \{0,1\}$ then

$$\begin{aligned} P^{(h)}(\delta_1, \ldots, \delta_i, \ldots, \delta_k) - P^{(h)}(\delta_1, \ldots, \delta_i', \ldots, \delta_k) &= h(\delta_1, \ldots, \delta_i, \ldots, \delta_k) - h(\delta_1, \ldots, \delta_i', \ldots, \delta_k) \\ &= (\delta_i - \delta_i') \cdot \frac{\partial}{\partial \delta_i} P^{(h)}. \end{aligned}$$

Proposition 3.6. *Let* h *be a function from* $\{0,1\}^k \to \{-1,0,1\}$ *having a representation as a multilinear polynomial of degree* k*. Let* $P^{(h)}$ *denote this multilinear polynomial. Let* $(\delta_1, \ldots, \delta_k), (\delta_1', \ldots, \delta_k') \in \{0,1\}^k$*. For* $I \subseteq \{1, \ldots, k\}$ *let* $\delta(I) \in \{0,1\}^k$ *be the point with coordinates* $\delta_j(I) = \delta_j'$ *if* $j \in I$*; and* $\delta_j(I) = \delta_j$ *if* $j \notin I$*. Then*

$$(\delta_1 - \delta_1') \cdot \ldots (\delta_k - \delta_k') = c_h(k) \sum_{I \subseteq [k]} (-1)^{|I|} h(\delta_1(I), \ldots, \delta_k(I)),$$

where $1/c_h(k)$ *is the coefficient of the monomial corresponding to all* k *variables in the multilinear representation of* h*.*

Proof. By definition, we know that for all $(\delta_1, \ldots, \delta_k) \in \{0,1\}^k$, $h(\delta_1, \ldots, \delta_k) = P^{(h)}(\delta_1, \ldots, \delta_k)$. Since $P^{(h)}$ is a linear function of δ_1,

$$\begin{aligned} P^{(h)}(\delta_1, \delta_2, \ldots, \delta_k) - P^{(h)}(\delta_1', \delta_2, \ldots, \delta_k) &= h(\delta_1, \delta_2, \ldots, \delta_k) - h(\delta_1', \delta_2, \ldots, \delta_k) \\ &= (\delta_1 - \delta_1') \cdot \frac{\partial}{\partial \delta_1} P^{(h)}, \end{aligned}$$

where $\frac{\partial}{\partial \delta_1} P^{(h)}$ denotes the partial derivative of $P^{(h)}(\delta_1, \delta_2, \ldots, \delta_k)$ with respect to δ_1. Repeating this for the other coordinates, we get

$$\sum_{I \subseteq [k]} (-1)^{|I|} h(\delta_1(I), \ldots, \delta_k(I)) = (\delta_1 - \delta_1') \cdot \ldots \cdot (\delta_k - \delta_k') \cdot \left(\frac{\partial}{\partial \delta_1} \ldots \frac{\partial}{\partial \delta_k} P^{(h)}\right).$$

The last term in the right hand side is the coefficient of the polynomial $P^{(h)}(\delta_1, \ldots, \delta_k)$ corresponding to the monomial $\delta_1 \cdot \ldots \cdot \delta_k$, and we denote it by $1/c_h(k)$. □

Corollary 3.7 (Corollary to Proposition 3.6)**.** *Let $T_{(1)},\dots,,T_{(k)},T'_{(1)},\dots,T'_{(k)}$ be $2k$ matrices with $\{0,1\}$ entries and dimensions $d\times n$. For a set $I\subseteq[k]$ denote $T_{(j)}(I)=T_{(j)}$ if $j\in I$ and $T_{(j)}(I)=T'_{(j)}$ if $j\notin I$. Then the following holds,*

$$(T_{(1)}-T'_{(1)})\odot\ldots\odot(T_{(k)}-T'_{(k)})=c_h(k)\sum_{I\subseteq[k]}(-1)^{|I|}\Pi_h(T_{(1)}(I),\dots,T_{(k)}(I)).$$

Proof. Follows by a simple extension of Proposition 3.6 and using Definitions 5 and 6. □

For a random $\mathbf{T}$, the following theorem shows that the ℓ_1-norm of $\Pi_h(\mathbf{T},\dots,\mathbf{T})\mathbf{x}$ is "big" for all $\mathbf{x}\in S^{n-1}$.

Theorem 3.8. *Let k,q,n,d be natural numbers. Assume that $n\le cd^k/\log_{(q)}d$. Consider a $d\times n$ matrix $\mathbf{T}$ with independent Bernoulli entries taking values 0 and 1 with probability $1/2$. Let h be a function from $\{0,1\}^k\to\{-1,0,1\}$ having a representation as a multilinear polynomial of degree k. Then the matrix $\Pi_h(\mathbf{T},\dots,\mathbf{T})$ satisfies*

$$\Pr\left[\exists\mathbf{x}\in S^{n-1}\ \left\|\Pi_h(\mathbf{T},\dots,\mathbf{T})\mathbf{x}\right\|_1\le C'd^k\right]\le c_1\exp(-c_2d).$$

The constants c,C',c_1,c_2 depend only on k and q.

Proof. First notice that if an $a_1\times n$ matrix M' is formed from the $a_2\times n$ matrix M by taking a subset of rows, then for any $\mathbf{x}\in\mathbb{R}^n$, $\|M'\mathbf{x}\|_1\le\|M\mathbf{x}\|_1$.

Let $d=2kd'+l$, where $0\le l<2k$. For $j=1,\dots,k$ denote by $\mathbf{T}^1_j$ the submatrix of $\mathbf{T}$ consisting of rows $(2d'(j-1)+1),\dots,(2d'(j-1)+d')$, and by $\mathbf{T}^0_j$ the submatrix consisting or rows $(2d'(j-1)+d'+1),\dots,2jd'$. For a set $I\subseteq[k]$ denote

$$\mathbf{T}_{(j)}(I)=\begin{cases}\mathbf{T}^0_j & \text{if } j\in I,\\ \mathbf{T}^1_j & \text{if } j\notin I.\end{cases}$$

Then Corollary 3.7 implies

$$(\mathbf{T}^1_1-\mathbf{T}^0_1)\odot\ldots\odot(\mathbf{T}^1_k-\mathbf{T}^0_k)=c_h(k)\sum_{I\subseteq[k]}(-1)^{|I|}\Pi_h(\mathbf{T}_{(1)}(I),\dots,\mathbf{T}_{(k)}(I)).$$

Therefore, by triangle inequality,

$$\begin{aligned}\left\|(\mathbf{T}^1_1-\mathbf{T}^0_1)\odot\ldots\odot(\mathbf{T}^1_k-\mathbf{T}^0_k)\mathbf{x}\right\|_1&\le c_h(k)\sum_{I\subseteq[k]}\left\|\Pi_h(\mathbf{T}_{(1)}(I),\dots,\mathbf{T}_{(k)}(I))\mathbf{x}\right\|_1\\&\le c_h(k)2^k\left\|\Pi_h(\mathbf{T},\dots,\mathbf{T})\mathbf{x}\right\|_1.\end{aligned}$$

The last inequality follows since $\forall I\ \Pi_h(\mathbf{T}_{(1)}(I),\dots,\mathbf{T}_{(k)}(I))$ is a submatrix of $\Pi_h(\mathbf{T},\dots,\mathbf{T})$, which implies that

$$\left\|\Pi_h(\mathbf{T}_{(1)}(I),\dots,\mathbf{T}_{(k)}(I))\mathbf{x}\right\|_1\le\left\|\Pi_h(\mathbf{T},\dots,\mathbf{T})\mathbf{x}\right\|_1.$$

The entries of the matrices $(\mathbf{T}^1_1-\mathbf{T}^0_1),\dots,(\mathbf{T}^1_k-\mathbf{T}^0_k)$ are independent τ-random variables. Thus, Theorem 3.4 (note that $d'=O(d)$) yields

$$\Pr[\exists\mathbf{x}\in S^{n-1}\left\|(\mathbf{T}^1_1-\mathbf{T}^0_1)\odot\ldots\odot(\mathbf{T}^1_k-\mathbf{T}^0_k)\mathbf{x}\right\|_1\le Cd^k]\le c_1\exp\left(-c_2d\right),$$

which along with Equation (**??**) proves the theorem. Note that $c_h(k)$ can be bounded as a function of k alone. □

Combining Theorem 3.8 with Cauchy-Schwarz's inequality, we obtain a lower bound on the least singular value of $\Pi_h(\mathbf{T}, \ldots, \mathbf{T})$. It is well known [19, 22] that the least singular value of a $d^k \times n$ matrix with independent τ-random entries is $\approx d^k$. Therefore, we get that in spite of all the correlations that exist in $\Pi_h(\mathbf{T}, \ldots, \mathbf{T})$ its least singular value is asymptotically the same order as that of an i.i.d. matrix.

Theorem 3.9. *Under the assumptions of Theorem 3.8*

$$\Pr\left[\sigma_n(\Pi_h(\mathbf{T}, \ldots, \mathbf{T})) \leq C'\sqrt{d^k}\right] \leq c_1 \exp\left(-c_2 d\right).$$

Proof. By Cauchy-Schwarz inequality,

$$\|\Pi_h(\mathbf{T}, \ldots, \mathbf{T})\mathbf{x}\|_1 \leq \sqrt{d^k}\, \|\Pi_h(\mathbf{T}, \ldots, \mathbf{T})\mathbf{x}\|.$$

Therefore,

$$\Pr\left[\exists \mathbf{x} \in S^{n-1}\ \ \|\Pi_h(\mathbf{T}, \ldots, \mathbf{T})\mathbf{x}\| \leq C'\sqrt{d^k}\right] \leq \Pr\left[\exists \mathbf{x} \in S^{n-1}\ \ \|\Pi_h(\mathbf{T}, \ldots, \mathbf{T})\mathbf{x}\|_1 \leq C' d^k\right].$$

The right-hand side probability could be bounded using Theorem 3.8 and the left-hand probability is exactly $\Pr\left[\sigma_n(\Pi_h(\mathbf{T}, \ldots, \mathbf{T})) \leq C'\sqrt{d^k}\right]$. □

4 Releasing M-estimators

In this section, we analyze privacy lower bounds for releasing M-estimators. Assume we have n samples $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^{k+1}$, consider the following optimization problem:

$$\mathcal{L}(\theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta; \mathbf{x}_i), \tag{8}$$

where $\theta \in \Theta \subset \mathbb{R}^{k+1}$, the *separable* loss function $\mathcal{L} : \Theta \times (\mathbb{R}^{k+1})^n \to \mathbb{R}$ measures the "fit" of $\theta \in \Theta$ to any given data $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and $\ell : \Theta \times \mathbb{R}^{k+1} \to \mathbb{R}$ is the loss function associated with a single data point. It is common to assume that the loss function has certain properties, e.g., for any given sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the loss function assigns a cost $\mathcal{L}(\theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) \geq 0$ to the estimator θ. The M-estimator ($\hat{\theta}$) associated with a given a function $\mathcal{L}(\theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) \geq 0$ is

$$\hat{\theta} = \text{argmin}_{\theta \in \Theta}\, \mathcal{L}(\theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \text{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} \ell(\theta; \mathbf{x}_i).$$

For a differentiable loss function ℓ, the estimator $\hat{\theta}$ could be found by setting $\partial\, \mathcal{L}(\theta; \mathbf{x}_1, \ldots, \mathbf{x}_n)$ to zero.

M-estimators are natural extensions of the Maximum Likelihood Estimators (MLE) [16]. They enjoy similar consistency and are asymptotically normal. natural extensions of the MLE. There are several reasons for studying these estimators: (i) they may be more computationally efficient than the MLE, and (ii) they may be more robust (resistant to deviations) than MLE. The linear regression MLE is captured by setting $\ell(\theta; \mathbf{x}) = (y - \langle \mathbf{z}, \theta \rangle)^2$ in Equation (8) where $\mathbf{x} = (\mathbf{z}, y)$, and MLE for logistic regression is captured by setting $\ell(\theta; \mathbf{x}) = y\langle \mathbf{z}, \theta \rangle - \ell(1 + \exp(\mathbf{z}, \theta))$ [12].

Problem Statement. Let $D = (U|\mathbf{s})$ be a database dimension $n \times d+1$, and let U be a real-valued matrix of dimension $n \times d$ and $\mathbf{s} \in \{0,1\}^n$ be a (secret) vector. Let $D = (\delta_{i,j})$. Consider the submatrix $D|_J$ of D, where $J \in \{1,\ldots,d\}^k$. Let $\hat{\theta}_J$ be the M-estimator for $D|_J$ defined as

$$\hat{\theta}_J = \text{argmin}_{\theta\in\Theta} \sum_{i=1}^{n} \ell(\theta; (\delta_{i,j_1}, \ldots, \delta_{i,j_k}, s_i)) \text{ where } J = (j_1, \ldots, j_k).$$

The goal is to understand how much noise is needed to attribute privately release $\hat{\theta}_J$ for all J's when the loss function (ℓ) is differentiable.

Basic Scheme. Consider a differentiable loss function ℓ. Let $D = (U|\mathbf{s})$. Let $U = (U_{(1)}|, \ldots, |U_{(d/k)})$, where each $U_{(i)}$ is an $n \times k$ matrix (assume d is a multiple of k for simplicity). Consider any $U_{(i)}$. We have

$$\partial\, \mathcal{L}(\theta; (U_{(i)}, \mathbf{s})) = \frac{1}{n} \sum_{j=1}^{n} \partial\, \ell(\theta_i; (U_{(i)_j}, s_j)),$$

where $U_{(i)_j}$ is the jth row in $U_{(i)}$. Then M-estimators $\hat{\theta}_i$ for $i \in [d/k]$ is obtained by solving

$$\frac{1}{n} \sum_{j=1}^{n} \partial\, \ell(\theta_i; (U_{(i)_j}, s_j)) = 0. \tag{9}$$

This gives a set of constraints over $\mathbf{s}$ which the adversary could use to construct $\hat{\mathbf{s}}$. For the case of linear and logistic regression, Equation (9) reduces to a form $U_{(i)}^{\top}\mathbf{s} - \mathbf{r} = 0$, where $\mathbf{r}$ is a vector independent of $\mathbf{s}$. For general loss function, we would use the fact that $\mathbf{s}$ is binary and use a decomposition similar to Equation (1). The other issue is that the adversary gets only a noisy approximation of $\hat{\theta}_1, \ldots, \hat{\theta}_{d/k}$ and we overcome this problem by using the Lipschitz properties of the gradient of the loss function.

In the next subsection, we focus on the standard MLE's for linear and logistic regression. In Section 4.2, we consider general M-estimators. Here, we would require an additional variance condition on the loss function. We would use this following standard definition of Lipschitz continuous gradient.

Definition 9 (Lipschitz Continuous Gradient)**.** *The gradient of a function $G : \mathbb{R}^k \to \mathbb{R}$ is Lipschitz continuous with parameter $\lambda > 0$ if, $\|\partial\, G(\mathbf{x}) - \partial\, G(\mathbf{y})\| \leq \lambda\|\mathbf{x} - \mathbf{y}\|$.*

Remark: Also, for any twice differentiable function G, $\lambda I \succeq \partial^2 G(\mathbf{x})$ for all $\mathbf{x}$ (where $\partial^2 G(\mathbf{x})$ denotes the Hessian matrix) [20].

4.1 Releasing Linear and Logistic Regression Estimators.

In this section, we establish distortion lower bounds for attribute privately releasing linear and logistic regression estimators.

Linear Regression Background. A general linear regression problem could be represented as $\mathbf{s} = X\theta + \varepsilon$, where $\mathbf{s} = (s_1, \ldots, s_n) \in \mathbb{R}^n$ is a vector of observed responses, $X \in \mathbb{R}^{n\times k}$ is a matrix of regressors, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ is an unobserved error vector where each ε_i accounts for the discrepancy between the actual responses (s_i) and the predicted outcomes ($\langle X_i, \theta\rangle$), and $\theta \in \mathbb{R}^k$ is a vector of unknown estimators.

This above optimization problem has a closed-form solution given as $\hat{\theta} = (X^\top X)^{-1} X^\top \mathbf{s}$ (also known as the ordinary least squares estimator). Let $\mathcal{L}_{\text{lin}}(\theta; \mathbf{s}, X, \sigma^2)$ denote the log-likelihood of linear regression (the likelihood expression is given in Appendix C). The gradient (w.r.t. to θ) of the log-likelihood is given by [12]

$$\partial \mathcal{L}_{\text{lin}}(\theta; \mathbf{s}, X, \sigma^2) = \frac{1}{\sigma^2} \left(X^\top \mathbf{s} - X^\top X \theta \right).$$

Logistic Regression Background. Logistic regression models estimate probabilities of events as functions of independent variables. Let s_i be binary variable representing the value on the dependent variable for ith input, and the values of k independent variables for this same case be represented as $x_{i,j}$ ($j = 1, \ldots, k$). Let n denote the total sample size, and let ζ_i denote the probability of success ($\Pr[s_i = 1]$). The design matrix of independent variables, $X = (x_{i,j})$, is composed of n rows and k columns.

The logistic regression model equates the logit transform, the log-odds of the probability of a success, to the linear component:

$$\ln\left(\frac{\zeta_i}{1-\zeta_i}\right) = \sum_{i=1}^{k} x_{i,j}\theta \equiv \zeta_i = \frac{1 + \exp(\langle \theta, X_i \rangle)}{\exp(\langle \theta, X_i \rangle)} \text{ where } X_i \text{ is the } i\text{th row in } X.$$

To emphasize the dependence of ζ_i on θ, we use the notation $\zeta_i = \zeta_i(\theta)$. Let $\mathcal{L}_{\log}(\theta; \mathbf{s}, X)$ denote the log-likelihood of logistic regression (see Appendix C for the likelihood expression). The gradient (w.r.t. to θ) of the log-likelihood is given by [12]

$$\partial \mathcal{L}_{\log}(\theta; \mathbf{s}, X) = X^\top \mathbf{s} - X^\top \text{vert}(\zeta_1(\theta), \ldots, \zeta_n(\theta)),$$

where the notation $\text{vert}(\cdot, \ldots, \cdot)$ denotes vertical concatenation of the argument matrices/vectors. Our analysis will require a bound on the Lipschitz constant of the gradient of the log-likelihood function, which we bound using the following claim.

Claim 4. *The Lipschitz constant Λ_{log} of the gradient of the log-likelihood $\partial \mathcal{L}_{\log}(\theta; \mathbf{s}, X)$ can be bounded by the operator norm of $X^\top X$.*

Proof. From Definition 9, we know that the Lipschitz constant of the gradient of the log-likelihood ($\partial \mathcal{L}_{\log}(\theta; \mathbf{s}, X)$), can be bounded by maximum eigenvalue of Hessian of $\mathcal{L}_{\log}(\theta; \mathbf{s}, X)$. The (i, j)th entry of Hessian of $\mathcal{L}_{\log}(\theta; \mathbf{s}, X)$ is

$$\frac{\partial^2 \mathcal{L}_{\log}(\theta; \mathbf{s}, X)}{\partial \theta_i \partial \theta_j} = -\sum_{l=1}^{n} \zeta_\ell(\theta)(1 - \zeta_\ell(\theta)) X_{l,i} X_{l,j},$$

where $X_{a,b}$ denote the (a, b)th entry in X. Note that the Hessian is a $k \times k$ matrix. Since $0 \leq \zeta_\ell(\theta) \leq 1$ (as $\zeta_\ell(\theta)$ represents a probability), we have

$$\frac{\partial^2 \mathcal{L}_{\log}(\theta; \mathbf{s}, X)}{\partial \theta_i \partial \theta_j} \leq -\sum_{l=1}^{n} X_{l,i} X_{l,j}.$$

The Hessian matrix is therefore $-X^\top X$. Hence, the Lipschitz constant ($\Lambda_{\log}$) of $\partial \mathcal{L}_{\log}(\theta; \mathbf{s}, X)$ can be bounded by the operator norm of $X^\top X$ (see the remark after Definition 9). □

Creating a Linear Reconstruction Problem for Linear Regression. Let $U = (U_{(1)}|, \ldots, |U_{(d/k)})$, where each $U_{(i)}$ is an $n \times k$ matrix (assume d is a multiple of k). Let $\hat{\theta}_i$ be the solution to the MLE equation

$$U_{(i)}^\top \mathbf{s} - U_{(i)}^\top U_{(i)} \theta = 0.$$

The adversary gets $\tilde{\theta}_i$'s which are a noisy approximation to $\hat{\theta}_i$'s. Given $\tilde{\theta}_1, \ldots, \tilde{\theta}_{d/k}$, the adversary solves the following set of linear constraints[12] to construct $\hat{\mathbf{s}}$:

$$\partial \mathcal{L}_{\text{lin}}(\tilde{\theta}_1; \mathbf{s}, U_{(1)}, \sigma^2) =, \ldots, = \partial \mathcal{L}_{\text{lin}}(\tilde{\theta}_{d/k}; \mathbf{s}, U_{(d/k)}, \sigma^2) = 0.$$

This can be re-written as:

$$U_{(1)}^\top \mathbf{s} - U_{(1)}^\top U_{(1)} \tilde{\theta}_1 =, \ldots, = U_{(d/k)}^\top \mathbf{s} - U_{(d/k)}^\top U_{(d/k)} \tilde{\theta}_{d/k} = 0. \tag{10}$$

Creating a Linear Reconstruction Problem for Logistic Regression. This reduction is very similar to the linear regression case. As before let $U = (U_{(1)}|, \ldots, |U_{(d/k)})$. Let $\hat{\theta}_i$ be the solution to the MLE equation

$$U_{(i)}^\top \mathbf{s} - U_{(i)}^\top \text{vert}(\zeta_{(i)_1}(\theta), \ldots, \zeta_{(i)_n}(\theta)) = 0.$$

Let $\zeta_{(i)}(\theta) = \text{vert}(\zeta_{(i)_1}(\theta), \ldots, \zeta_{(i)_n}(\theta))$. The adversary gets $\tilde{\theta}_i$'s which are a noisy approximation to $\hat{\theta}_i$'s. Using $\tilde{\theta}_1, \ldots, \tilde{\theta}_{d/k}$ and U, the adversary can construct $\zeta_{(1)}(\tilde{\theta}_1), \ldots, \zeta_{(d/k)}(\tilde{\theta}_{d/k})$. The adversary then solves the following set of linear constraints to construct $\hat{\mathbf{s}}$:

$$\partial \mathcal{L}_{\log}(\tilde{\theta}_1; \mathbf{s}, U_{(1)}) =, \ldots, = \partial \mathcal{L}_{\log}(\tilde{\theta}_{d/k}; \mathbf{s}, U_{(d/k)}) = 0.$$

This can be re-written as:

$$U_{(1)}^\top \mathbf{s} - U_{(1)}^\top \zeta_{(1)}(\tilde{\theta}_1) =, \ldots, = U_{(d/k)}^\top \mathbf{s} - U_{(d/k)}^\top \zeta_{(d/k)}(\tilde{\theta}_{d/k}) = 0. \tag{11}$$

Setting up the Least Squares and LP Decoding Attacks. Consider linear regression. The attacks operate on the linear reconstruction problem of Equation (10). The least squares attack constructs $\hat{\mathbf{s}}$ (an approximation of $\mathbf{s}$) by minimizing the ℓ_2 norm of the left hand side of Equation (10), whereas LP decoding attack works by minimizing ℓ_1 norm. The attacks are similar for logistic regression except that they now operate on Equation (11).

In the following analysis, we use a random matrix for U where each entry of the random matrix is an independent τ-random variable (Definition 7).

Theorem 4.1. *Let $d \geq 2n$ and set $k = 1$. Then*

- *Any privacy mechanism which for every database $D = (U|\mathbf{s})$ where $U \in \mathbb{R}^{n \times d}$ and $\mathbf{s} \in \{0,1\}^n$ releases the estimators of the linear/logistic regression model between every column of U and $\mathbf{s}$ by adding $o(1/\sqrt{n})$ noise to each estimator is attribute non-private. The attack that achieves this attribute non-privacy violation runs in $O(dn^2)$ time.*

[12] Equivalently, we could write the below equation as a single constraint:

$$\text{vert}(\partial \mathcal{L}_{\text{lin}}(\tilde{\theta}_1; \mathbf{s}, U_{(1)}, \sigma^2), \ldots, \partial \mathcal{L}_{\text{lin}}(\tilde{\theta}_{d/k}; \mathbf{s}, U_{(d/k)}, \sigma^2)) = 0.$$

- *There exists a constant $\gamma > 0$ such that any mechanism which for every database $D = (U|\mathbf{s})$ where $U \in \mathbb{R}^{n\times d}$ and $\mathbf{s} \in \{0,1\}^n$ releases the estimators of the linear/logistic regression model between every column of U and $\mathbf{s}$ by adding $o(1/\sqrt{n})$ noise to at most $1-\gamma$ fraction of the estimators is attribute non-private. The attack that achieves this attribute non-privacy violation runs in $O(d^2n^3 + d^{2.5}n^2)$ time.*

Proof. We first do the analysis for linear regression. Let $D = (\mathbf{U}|\mathbf{s})$, where $\mathbf{U}$ is a τ-random matrix of dimension $n \times d$.

Analysis for Linear Regression. Now $\hat{\theta}_i$ is the solution to the MLE equation: $\mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \mathbf{U}_{(i)}\theta = 0$ where $\mathbf{U}_{(i)}$ is the ith column of $\mathbf{U}$. Since $k = 1$, we have $\hat{\theta}_i \in \mathbb{R}$. The adversary gets noisy approximations of $\hat{\theta}_1, \ldots, \hat{\theta}_d$. Let $\tilde{\theta}_1, \ldots, \tilde{\theta}_d$ be these noisy approximations with $\tilde{\theta}_i = \hat{\theta}_i + e_i$ (for some unknown e_i). The adversary solves the following set of linear constraints:

$$\mathbf{U}_{(1)}^\top \mathbf{s} - \mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\tilde{\theta}_1 =, \ldots, = \mathbf{U}_{(d)}^\top \mathbf{s} - \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\tilde{\theta}_d = 0.$$

This could be re-written as:

$$\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\tilde{\theta}_1, \ldots, \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\tilde{\theta}_d) = 0.$$

Let us first look at the least squares attack. Let $\mathbf{e} = (e_1, \ldots, e_d)$ (i.e., $\mathbf{e}$ is the error vector). We have $\mathbf{U}_{(i)}^\top \mathbf{U}_{(i)} \leq n$ for all $i \in [d]$ (as $\mathbf{U}$ is a τ-random matrix, all entries in the matrix are at most 1, and $\mathbf{U}_{(i)}$ is the ith column in $\mathbf{U}$). The least squares attack produces an estimate $\hat{\mathbf{s}}$ by solving:

$$\text{argmin}_{\mathbf{s}} \|\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\tilde{\theta}_1, \ldots, \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\tilde{\theta}_d)\|.$$

Since $\tilde{\theta}_i = \hat{\theta}_i + e_i$, we get for all $i \in [d]$

$$\mathbf{U}_{(i)}^\top \mathbf{U}_{(i)}\tilde{\theta}_i = \mathbf{U}_{(i)}^\top \mathbf{U}_{(i)}\hat{\theta}_i + \mathbf{U}_{(i)}^\top \mathbf{U}_{(i)}e_i \leq \mathbf{U}_{(i)}^\top \mathbf{U}_{(i)}\hat{\theta}_i + ne_i.$$

This implies

$$\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\tilde{\theta}_1, \ldots, \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\tilde{\theta}_d) \leq \mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\hat{\theta}_1, \ldots, \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\hat{\theta}_d) + n \cdot \mathbf{e}.$$

The remaining analysis is similar to Theorem 2.1, except that again each error term e_i is scaled up by a factor (at most) n. Since $\mathbf{U}$ is a τ-random matrix, it is well-known that if $d \geq 2n$, then the least singular value of $\mathbf{U}$ is with exponentially high probability $\Omega(\sqrt{d})$ (see, e.g., [19]). Therefore, if a privacy mechanism adds $o(\sqrt{n})/n$[13] noise to each $\hat{\theta}_i$, then the least squares attack with exponentially high probability recovers $1 - o(1)$ fraction of the entries in $\mathbf{s}$. The time for executing this attack is $O(dn^2)$ as the attack requires computing the SVD of a $d \times n$ matrix.

The LP decoding attack produces an estimate $\hat{\mathbf{s}}$ by solving:

$$\text{argmin}_{\mathbf{s}} \|\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \mathbf{U}_{(1)}\tilde{\theta}_1, \ldots, \mathbf{U}_{(d)}^\top \mathbf{U}_{(d)}\tilde{\theta}_d)\|_1.$$

The analysis follows from Theorem 2.2, except that each error is scaled up by a factor (at most) n. Since $\mathbf{U}$ is a τ-random matrix, it also holds that with exponentially high probability $\mathbf{U}$ is a Euclidean section [10]. Therefore, if a privacy mechanism adds $o(\sqrt{n})/n = o(1/\sqrt{n})$ noise to at most $1-\gamma$ fraction of the $\hat{\theta}_i$'s, then the LP decoding attack with exponentially high probability recovers $1 - o(1)$ fraction of the entries in $\mathbf{s}$. The time for executing this attack is $O(d^2n^3 + d^{2.5}n^2)$ (as there are d constraints, n variables, and the

[13] The n in the denominator is because of the scaling of the noise $\mathbf{e}$.

number of the bits in the input is bounded by dn).

Analysis for Logistic Regression. Consider the estimator $\hat{\theta}_i$ of the logistic regression model between the ith column of $\mathbf{U}$ and $\mathbf{s}$. $\hat{\theta}_i$ is the solution to the MLE equation: $\mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \zeta_{(i)}(\theta) = 0$ where $\mathbf{U}_{(i)}$ is the ith column of $\mathbf{U}$. Since $k = 1$, we have $\hat{\theta}_i \in \mathbb{R}$. The adversary gets noisy approximations of $\hat{\theta}_1, \ldots, \hat{\theta}_d$. Let $\tilde{\theta}_1, \ldots, \tilde{\theta}_d$ be these noisy approximations with $\tilde{\theta}_i = \hat{\theta}_i + e_i$.

Using $\tilde{\theta}_1, \ldots, \tilde{\theta}_d$ and $\mathbf{U}$, the adversary can construct $\zeta_{(1)}(\tilde{\theta}_1), \ldots, \zeta_{(d)}(\tilde{\theta}_d)$. The adversary then solves the following set of linear constraints:

$$\mathbf{U}_{(1)}^\top \mathbf{s} - \mathbf{U}_{(1)}^\top \zeta_{(1)}(\tilde{\theta}_1) =, \ldots, = \mathbf{U}_{(d)}^\top \mathbf{s} - \mathbf{U}_{(d)}^\top \zeta_{(d)}(\tilde{\theta}_d) = 0.$$

Let us now apply Lipschitz condition on the gradient of the log-likelihood function. Note that since $k = 1$, $\partial \mathcal{L}_{\log}(\tilde{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)})$ is a scalar variable (for every $i \in [d]$). By Lipschitz condition,

$$|\partial \mathcal{L}_{\log}(\tilde{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)}) - \partial \mathcal{L}_{\log}(\hat{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)})| \leq \Lambda_{\log} |\tilde{\theta}_i - \hat{\theta}_i| = \Lambda_{\log} |e_i|,$$

where $\Lambda_{\log} = \mathbf{U}_{(i)}^\top \mathbf{U}_{(i)} \leq n$ (using Claim 4). Therefore,

$$\partial \mathcal{L}_{\log}(\tilde{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)}) \leq n|e_i| + \partial \mathcal{L}_{\log}(\hat{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)}).$$

Substituting $\partial \mathcal{L}_{\log}(\tilde{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)}) = \mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \zeta_{(i)}(\tilde{\theta}_i)$ and $\partial \mathcal{L}_{\log}(\hat{\theta}_i; \mathbf{s}, \mathbf{U}_{(i)}) = \mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \zeta_{(i)}(\hat{\theta}_i)$ in the above equation,

$$\mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \zeta_{(i)}(\tilde{\theta}_i) \leq \mathbf{U}_{(i)}^\top \mathbf{s} - \mathbf{U}_{(i)}^\top \zeta_{(i)}(\hat{\theta}_i) + n|e_i|.$$

This implies

$$\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \zeta_{(1)}(\tilde{\theta}_1), \ldots, \mathbf{U}_{(d)}^\top \zeta_{(1)}(\tilde{\theta}_d)) \leq \mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \zeta_{(1)}(\hat{\theta}_1), \ldots, \mathbf{U}_{(d)}^\top \zeta_{(d)}(\hat{\theta}_d)) + n \cdot |\mathbf{e}|.$$

The least squares attack produces an estimate $\hat{\mathbf{s}}$ by solving:

$$\text{argmin}_{\mathbf{s}} \|\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \zeta_{(1)}(\tilde{\theta}_1), \ldots, \mathbf{U}_{(d)}^\top \zeta_{(d)}(\tilde{\theta}_d))\|.$$

The remaining analysis follows as in the linear regression case (from Theorem 2.1) and again the errors are scaled by a factor (at most) n. Therefore, if a privacy mechanism adds $o(\sqrt{n})/n = o(1/\sqrt{n})$ noise to each $\hat{\theta}_i$, then the least squares attack with exponentially high probability recovers $1 - o(1)$ fraction of the entries in $\mathbf{s}$. The time for executing this attack is $O(dn^2)$.

The LP decoding attack produces an estimate $\hat{\mathbf{s}}$ by solving:

$$\text{argmin}_{\mathbf{s}} \|\mathbf{U}^\top \mathbf{s} - \text{vert}(\mathbf{U}_{(1)}^\top \zeta_{(1)}(\tilde{\theta}_1), \ldots, \mathbf{U}_{(d)}^\top \zeta_{(d)}(\tilde{\theta}_d))\|_1.$$

Again like in the linear regression case (using Theorem 2.2), we get that any mechanism that adds $o(\sqrt{n})/n = o(1/\sqrt{n})$ noise to at most $1-\gamma$ fraction of the $\hat{\theta}_i$'s is attribute non-private. The time for executing this attack is $O(d^2n^3 + d^{2.5}n^2)$. □

Compared to Theorems 3.2 and 3.3, this above theorem requires a much larger d (about $O(n)$). However, it is possible to reduce to dependence on $d \approx n^{O(1/k)}$ if the released statistic is a degree k polynomial kernel of these regression functions. We defer the details to the full version.

4.2 Releasing General M-estimators.

In this section, we establish distortion lower bounds for attribute privately releasing M-estimators associated with differentiable loss functions.

Creating a Linear Reconstruction Problem. Consider Equation (9). Since $\mathbf{s}$ is a binary vector, we can decompose $\partial\,\ell(\theta;(U_{(i)_j},s_j)) \in \mathbb{R}^k$ as follows:

$$\partial\,\ell(\theta;(U_{(i)_j},s_j)) = \ell_0(\theta;U_{(i)_j})(1-s_j) + \ell_1(\theta;U_{(i)_j})s_j = \ell_0(\theta;U_{(i)_j}) + (\ell_1(\theta;U_{(i)_j}) - \ell_0(\theta;U_{(i)_j}))s_j,$$

where $\ell_0(\theta;U_{(i)_j}) = \partial\,\ell(\theta;(U_{(i)_j},0)) \in \mathbb{R}^k$ and $\ell_1(\theta;U_{(i)_j}) = \partial\,\ell(\theta;(U_{(i)_j},1)) \in \mathbb{R}^k$. This is similar to the decomposition in Equation (1). Let

$$\ell_2(\theta;U_{(i)_j}) = \ell_1(\theta;U_{(i)_j}) - \ell_0(\theta;U_{(i)_j}).$$

Now the M-estimator ($\hat{\theta}_i$) between $U_{(i)}$ and $\mathbf{s}$ can be found by setting $\partial\,\mathcal{L}(\theta;(U_{(i)},\mathbf{s})) = 0$. Therefore, $\hat{\theta}_i$ is the solution to (ignoring the scaling multiplier $1/n$)

$$\begin{aligned}\partial\,\mathcal{L}(\theta;(U_{(i)},\mathbf{s})) &= \sum_{j=1}^{n} \partial\,\ell(\theta_i;(U_{(i)_j},s_j)) = 0\\ \equiv \sum_{j=1}^{n} \ell_0(\theta;U_{(i)_j}) &+ \ell_2(\theta;U_{(i)_j})s_j = 0.\end{aligned}$$

The adversary gets $\tilde{\theta}_i$'s which are a noisy approximation to $\hat{\theta}_i$'s. Given $\tilde{\theta}_1,\ldots,\tilde{\theta}_{d/k}$, the adversary solves the following set of linear constraints:

$$\begin{aligned}&\partial\,\mathcal{L}(\tilde{\theta}_1;(U_{(1)},\mathbf{s})) =,\ldots,= \partial\,\mathcal{L}(\tilde{\theta}_{d/k};(U_{(d/k)},\mathbf{s})) = 0 \qquad (12)\\ &\equiv \left(\sum_{j=1}^{n} \ell_0(\tilde{\theta}_1;U_{(1)_j}) + \ell_2(\tilde{\theta}_1;U_{(1)_j})s_j\right) =,\ldots,= \left(\sum_{j=1}^{n} \ell_0(\tilde{\theta}_{d/k};U_{(d/k)_j}) + \ell_2(\tilde{\theta}_{d/k};U_{(d/k)_j})s_j\right) = 0.\end{aligned}$$

This could also be represented in a matrix-form as we show below. For every $i \in [d/k]$, define $A_{(i)}$ as a $k \times n$ matrix whose jth column is $\ell_0(\tilde{\theta}_i;U_{(i)_j})$ and $B_{(i)}$ as a $k \times n$ matrix whose jth column is $\ell_2(\tilde{\theta}_i;U_{(i)_j})$. Then Equation (12) can be re-written as

$$A_{(1)}\mathbf{1}_n + B_{(1)}\mathbf{s} =,\ldots,= A_{(d/k)}\mathbf{1}_n + B_{(d/k)}\mathbf{s} = 0.$$

The adversary solves the above equation to obtain $\hat{\mathbf{s}}$. The analysis uses the following condition on ℓ.

Definition 10 (Variance Condition)**.** [14] *Consider the decomposition of the gradient of a loss function, $\ell : \Theta \times \mathbb{R}^{k+1} \to \mathbb{R}$ as $\partial\ell(\theta;(\mathbf{x},y)) = \ell_0(\theta;\mathbf{x}) + \ell_2(\theta;\mathbf{x})y$ where $\theta \in \mathbb{R}^k, \mathbf{x} \in \mathbb{R}^k, y \in \{0,1\}$. The loss function ℓ is said to be satisfy the variance condition if for every θ, $Var_{\mathbf{x}}[\ell_2(\theta;\mathbf{x})]$ is bounded away from zero.*

[14] For the LP decoding attack, we need a stricter condition to achieve the guarantees of Theorem 2.2. We defer this discussion to the full version.

Theorem 4.2. *Let ℓ be a differentiable loss function which satisfies the variance condition. Let λ denote the Lipschitz constant of the gradient of the loss function ℓ. Let $d \geq 2n$ and set $k = 1$. Then any privacy mechanism which for every database $D = (U|\mathbf{s})$ where $U \in \mathbb{R}^{n\times d}$ and $\mathbf{s} \in \{0,1\}^n$ releases the M-estimators associated with the loss function ℓ of the models fitted between every column of U and $\mathbf{s}$ by adding $o(1/(\sqrt{n}\lambda))$ noise to each M-estimator is attribute non-private. The attack that achieves this attribute non-privacy violation runs in $O(dn^2)$ time.*

Proof. Let $D = (\mathbf{U}|\mathbf{s})$, where $\mathbf{U}$ is a τ-random matrix of dimension $n \times d$.

The M-estimator $\hat{\theta}_i \in \mathbb{R}$ between the ith column $\mathbf{U}_{(i)}$ of $\mathbf{U}$ and $\mathbf{s}$ is given by the solution to the equation:

$$\sum_{j=1}^{n} \ell_0(\theta; \mathbf{U}_{(i)_j}) + \ell_2(\theta; \mathbf{U}_{(i)_j}) s_j = 0.$$

The adversary gets noisy approximations of $\hat{\theta}_1, \ldots, \hat{\theta}_d$. Let $\tilde{\theta}_1, \ldots, \tilde{\theta}_d$ be these noisy approximations with $\tilde{\theta}_i = \hat{\theta}_i + e_i$. Consider $\partial\,\mathcal{L}(\tilde{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s}))$ and $\partial\,\mathcal{L}(\hat{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s}))$. By Lipschitz condition[15],

$$|\partial\,\mathcal{L}(\tilde{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s})) - \partial\,\mathcal{L}(\hat{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s}))| \leq \Lambda |\tilde{\theta}_i - \hat{\theta}_i| \leq \lambda n |\tilde{\theta}_i - \hat{\theta}_i| = \lambda n |e_i|.$$

This implies that

$$\partial\,\mathcal{L}(\tilde{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s})) \leq \partial\,\mathcal{L}(\hat{\theta}_i; (\mathbf{U}_{(i)}, \mathbf{s})) + \lambda n |e_i|.$$

Substituting for the decomposition of ℓ in the above equation gives

$$\sum_{j=1}^{n} \ell_0(\tilde{\theta}_i; \mathbf{U}_{(i)_j}) + \ell_2(\tilde{\theta}_i; \mathbf{U}_{(i)_j}) s_j \leq \sum_{j=1}^{n} \ell_0(\hat{\theta}_i; \mathbf{U}_{(i)_j}) + \ell_2(\hat{\theta}_i; \mathbf{U}_{(i)_j}) s_j + \lambda n |e_i|. \tag{13}$$

Let $\mathbf{A}_1, \mathbf{B}_2, \mathbf{A}_2, \mathbf{B}_2$ be four matrices of dimension $d \times n$ defined as follows:

$$\begin{aligned}
\mathbf{A}_1 &: \ i\text{th row of } \mathbf{A}_1 \text{ is } \ell_0(\hat{\theta}_i; \mathbf{U}_{(i)_1}), \ldots, \ell_0(\hat{\theta}_i; \mathbf{U}_{(i)_n}),\\
\mathbf{B}_1 &: \ i\text{th row of } \mathbf{B}_1 \text{ is } \ell_2(\hat{\theta}_i; \mathbf{U}_{(i)_1}), \ldots, \ell_2(\hat{\theta}_i; \mathbf{U}_{(i)_n}),\\
\mathbf{A}_2 &: \ i\text{th row of } \mathbf{A}_2 \text{ is } \ell_0(\tilde{\theta}_i; \mathbf{U}_{(i)_1}), \ldots, \ell_0(\tilde{\theta}_i; \mathbf{U}_{(i)_n}),\\
\mathbf{B}_2 &: \ i\text{th row of } \mathbf{B}_2 \text{ is } \ell_2(\tilde{\theta}_i; \mathbf{U}_{(i)_1}), \ldots, \ell_2(\tilde{\theta}_i; \mathbf{U}_{(i)_n}).
\end{aligned}$$

The adversary solves the following reconstruction problem to compute $\hat{\mathbf{s}}$:

$$\mathbf{A}_2 \mathbf{1}_n + \mathbf{B}_2 \mathbf{s} = 0. \tag{14}$$

From Equation (13) it follows that

$$\mathbf{A}_2 \mathbf{1}_n + \mathbf{B}_2 \mathbf{s} \leq \mathbf{A}_1 \mathbf{1}_n + \mathbf{B}_1 \mathbf{s} + \lambda n |\mathbf{e}|.$$

$\mathbf{B_2}$: Least Singular Value. Since $\mathbf{U}$ is a τ-random matrix, $\mathbf{B}_2$ is another random matrix. However, $\mathbf{B}_2$ may not be centered (i.e., its entries might have non-zero means). We can re-express $\mathbf{B}_2$ as

$$\mathbf{B}_2 = \underbrace{\mathbf{B}_2 - \mathbb{E}[\mathbf{B}_2]}_{\mathbf{R}} + \mathbb{E}[\mathbf{B}_2].$$

[15]The Lipschitz constant Λ of $\mathcal{L}$ is at most n times the Lipschitz constant of ℓ, and therefore $\Lambda \leq n\lambda$.

Here, $\mathbf{R}$ is a τ'-random matrix[16] and $\mathbb{E}[\mathbf{B}_2]$ is a matrix of the form $\mathbf{u}\mathbf{v}^\top$ where $\mathbf{u}$ is a d-dimensional column vector with entries $\mathbb{E}_x[\ell_2(\tilde{\theta}_1; x)], \ldots, \mathbb{E}_x[\ell_2(\tilde{\theta}_d; x)]$ and $\mathbf{v}$ is a n-dimensional column vector of all ones.

Claim 5. $\Pr[\sigma_n(\mathbf{B}_2) \leq c_{19}\sqrt{d}] \leq \exp(-c_{20}d)$.

Proof. $\mathbf{B}_2 = \mathbf{R} + \mathbf{u}\mathbf{v}^\top$, where $\mathbf{R}$ is a τ'-random matrix. The rank of $\mathbf{u}\mathbf{v}^\top$ is 1, and its operator norm can be polynomially bounded in d. Applying Lemma D.2 implies the result. □

Least Squares Attack. The least squares attack produces an estimate $\hat{\mathbf{s}}$ by solving:

$$\hat{\mathbf{s}} = \text{argmin}_{\mathbf{s}} \|\mathbf{A}_2\mathbf{1}_n + \mathbf{B}_2\mathbf{s}\|.$$

The analysis is similar to Theorem 2.1, except that each error term e_i is scaled up by a factor (at most) λn. Therefore, if a privacy mechanism adds $o(1/(\sqrt{n}\lambda))$ noise to each $\hat{\theta}_i$, then the least squares attack with exponentially high probability recovers $1 - o(1)$ fraction of the entries in $\mathbf{s}$. The time for executing this attack is $O(dn^2)$.

□

References

[1] DE, A. Lower Bounds in Differential Privacy. In *TCC* (2012), pp. 321–338.

[2] DINUR, I., DWORK, C., AND NISSIM, K. Revealing information while preserving privacy, full version of [3], in preparation, 2009.

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

[4] DWORK, C. Differential privacy. In *ICALP* (2006), LNCS, pp. 1–12.

[5] DWORK, C., MCSHERRY, F., NISSIM, K., AND SMITH, A. Calibrating noise to sensitivity in private data analysis. In *TCC* (2006), vol. 3876 of *LNCS*, Springer, pp. 265–284.

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

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

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

[9] HARDT, M., AND TALWAR, K. On the Geometry of Differential Privacy. In *STOC* (2010), ACM, pp. 705–714.

[16] The expected value of each entry in $\mathbf{R}$ is zero and due to variance condition on the loss function (Definition 10), the second moment of each entry of $\mathbf{R}$ is bounded away from zero. We can re-scale Equation (14) to ensures that all the entries in $\mathbf{B}_2$ have absolute value less than 1 without changing the solution $\hat{\mathbf{s}}$. This implies that $\mathbf{R}$ is τ'-random matrix for an appropriate τ' depending on $\ell_2, \tilde{\theta}_1, \ldots, \tilde{\theta}_d$.

[10] KASHIN, B. S. Diameters of some finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat. 41* (1977), 334–351.

[11] KASIVISWANATHAN, S. P., RUDELSON, M., SMITH, A., AND ULLMAN, J. The Price of Privately Releasing Contingency Tables and the Spectra of Random Matrices with Correlated Rows. In *STOC* (2010), pp. 775–784.

[12] LEE, M.-J. *Micro-Econometrics: Methods of Moments and Limited Dependent Variables*. Springer, 2010.

[13] MCGREGOR, A., MIRONOV, I., PITASSI, T., REINGOLD, O., TALWAR, K., AND VADHAN, S. P. The limits of two-party differential privacy. In *FOCS* (2010), IEEE Computer Society, pp. 81–90.

[14] MUTHUKRISHNAN, S., AND NIKOLOV, A. Optimal private halfspace counting via discrepancy. In *STOC* (2012), H. J. Karloff and T. Pitassi, Eds., ACM, pp. 1285–1292.

[15] O'DONNELL, R. Analysis of boolean functions, 2012.

[16] PFANZAGL, J. *Parametric Statistical Theory*. WDeG Press, 1991.

[17] RUDELSON, M. Row Products of Random Matrices. *ArXiv e-prints* (2011).

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

[19] RUDELSON, M., AND VERSHYNIN, R. Smallest Singular Value of a Random Rectangular Matrix. *Communications on Pure and Applied Mathematics 62*, 12 (2009), 1707–1739.

[20] SMOLA, A., AND VISHWANATHAN, S. *Introduction to Machine Learning*. `http://alex.smola.org/drafts/thebook.pdf`, 2012.

[21] VAIDYA, P. M. An algorithm for linear programming which requires $o(((m+n)n^2+(m+n)^{1.5}n)l)$ arithmetic operations. *Math. Program. 47*, 2 (1990), 175–201.

[22] VERSHYNIN, R. Introduction to the Non-Asymptotic Analysis of Random Matrices, 2011.

A Preliminaries about Boolean Functions

We start with an alternate definition of non-degeneracy and show that it is equivalent to Definition 4.

Definition 11. *A boolean function $f : \{0,1\}^{k+1} \to \{0,1\}$ is called non-degenerate if*

$$\sum_{(\delta_1,\ldots,\delta_{k+1})\in\{0,1\}^{k+1}} (-1)^{f(\delta_1,\ldots,\delta_{k+1})-\sum_{j=1}^{k+1}\delta_j} \neq 0.$$

Any function on the discrete cube $\{-1,1\}^{k+1}$ can be decomposed into a linear combination of characters, which are Walsh functions. Such representation allows to extend the function f from the discrete cube to $\mathbb{R}^{k+1}$ as a multilinear (i.e., linear with respect to each variable separately) polynomial. In what follows, we will use this extension. The following lemma shows that the non-degeneracy condition is equivalent to the fact that this polynomial has the maximal degree.

Lemma A.1. *Definitions 11 and 4 are equivalent.*

Proof. Consider the function $g : \{-1,1\}^{k+1} \to \{-1,1\}$ defined by

$$g(\phi_1, \ldots, \phi_{k+1}) = 2f\left(\frac{1}{2}(1+\phi_1), \ldots, \frac{1}{2}(1+\phi_{k+1})\right) - 1.$$

Let $\phi = (\phi_1, \ldots, \phi_{k+1})$. For $S \subseteq \{1, \ldots, k+1\}$ let $\chi_S(\phi) = \prod_{j \in S} \phi_j$ be the corresponding Walsh function. Then the function g can be decomposed as

$$g(\phi) = \sum_{S \subseteq \{1,\ldots,k+1\}} \widehat{g}(S)\chi_S(\phi).$$

Note that $\deg(f) = \deg(g)$, and so $\deg(f) = k+1$ iff $\widehat{g}(1, \ldots, k+1) \neq 0$. We have

$$\widehat{g}(1, \ldots, k+1) = 2^{-k-1} \sum_{\phi \in \{-1,1\}^{k+1}} g(\phi) \prod_{j=1}^{k+1} \phi_j. \tag{15}$$

Since

$$g(\phi) = (-1)^{f(\delta_1,\ldots,\delta_{k+1})} \quad \text{and} \quad \prod_{j=1}^{k+1} \phi_j = (-1)^{\sum_{j=1}^{k+1} \delta_j},$$

where $\delta_j = (1-\phi_j)/2$, the lemma follows. □

Remark: There are 10 non-degenerate functions of two boolean variables:

AND: $\delta_1 \wedge \delta_2,\ \delta_1 \wedge (\neg\delta_2),\ (\neg\delta_1) \wedge \delta_2,\ (\neg\delta_1) \wedge (\neg\delta_2)$,
OR: $\delta_1 \vee \delta_2,\ \delta_1 \vee (\neg\delta_2),\ (\neg\delta_1) \vee \delta_2,\ (\neg\delta_1) \vee (\neg\delta_2)$,
XOR: $\delta_1 \oplus \delta_2,\ (\neg\delta_1) \oplus (\neg\delta_2)$.

The remaining 6 boolean functions of two variables are degenerate.

B Euclidean Section Property of Random Row Function Matrices

In this section, we establish the Euclidean section property needed for Theorem 3.3 (LP decoding attack). Let us consider a function $h : \{-1,1\}^k \to \{-1,0,1\}$ having a representation as a multilinear polynomial of degree k. Let $P^{(h)}$ denote this multilinear polynomial. We first prove a proposition analogous to Proposition 3.6 for h. For $I \subseteq [k]$, let us define

$$P_I^{(h)}(\phi_1, \ldots, \phi_k) = P^{(h)}(\phi_1', \ldots, \phi_k') \text{ where } \phi_i' = \phi_i \text{ if } i \in I, \text{ else } \phi_i' = 0.$$

That is $P_I^{(h)}(\phi_1, \ldots, \phi_k)$ is the multilinear polynomial $P_h(\phi_1, \ldots, \phi_k)$ restricted to variables only in I. Under this notation

$$P_{[k]}^{(h)}(\phi_1, \phi_2, \ldots, \phi_k) = P^{(h)}(\phi_1, \phi_2, \ldots, \phi_k).$$

Proposition B.1. *Let h be a function from $\{-1,1\}^k \to \{-1,0,1\}$ having a representation as a multilinear polynomial of degree k. Let $P^{(h)}$ be this multilinear polynomial. Let $(\phi_1,\ldots,\phi_k) \in \{-1,1\}^k$. Then*

$$\phi_1 \cdot \ldots \cdot \phi_k = c_h(k)(\sum_{I\subset[k]} (-1)^{k-|I|} P_I^{(h)}(\phi_1,\ldots,\phi_k) + h(\phi_1,\ldots,\phi_k)),$$

where $1/c_h(k)$ is the coefficient of the monomial corresponding to all k variables in $P^{(h)}$.

Proof. The proof is similar to Proposition 3.6. Since $P^{(h)}$ is a linear function in ϕ_1,

$$\begin{aligned} P^{(h)}(\phi_1,\phi_2,\ldots,\phi_k) - P^{(h)}(0,\phi_2,\ldots,\phi_k) &= P^{(h)}_{[k]}(\phi_1,\phi_2,\ldots,\phi_k) - P^{(h)}_{\{2,\ldots,k\}}(\phi_1,\phi_2,\ldots,\phi_k) \\ &= \phi_1 \frac{\partial}{\partial\phi_1} P^{(h)}, \end{aligned}$$

where $\frac{\partial}{\partial\phi_1}P^{(h)}$ is the partial derivative of $P^{(h)}(\phi_1,\ldots,\phi_k)$ with respect to ϕ_1. Repeating this for other coordinates, we get

$$\sum_{I\subseteq[k]} (-1)^{k-|I|} P_I^{(h)}(\phi_1,\ldots,\phi_k) = \phi_1 \cdot \ldots \cdot \phi_k \cdot \left(\frac{\partial}{\partial\phi_1} \ldots \frac{\partial}{\partial\phi_k} P^{(h)}\right).$$

When $I=[k]$,

$$h(\phi_1,\ldots,\phi_k) = P_I^{(h)}(\phi_1,\ldots,\phi_k),$$

so the above equation could be re-expressed as

$$\sum_{I\subset[k]} (-1)^{k-|I|} P_I^{(h)}(\phi_1,\ldots,\phi_k) + h(\phi_1,\ldots,\phi_k) = \phi_1 \cdot \ldots \cdot \phi_k \cdot \left(\frac{\partial}{\partial\phi_1} \ldots \frac{\partial}{\partial\phi_k} P^{(h)}\right).$$

The last term in the right hand side is the coefficient of the polynomial $P^{(h)}(\phi_1,\ldots,\phi_k)$ corresponding to the monomial $\phi_1 \cdot \ldots \cdot \phi_k$, and we denote it by $1/c_h(k)$. □

Let $V_{(1)} = (\phi^{(1)}_{i,j}),\ldots,V_{(k)} = (\phi^{(k)}_{i,j})$ be k matrices with $\{-1,1\}$ entries and dimensions $d \times n$. Let us define a $d^k \times n$ matrix $\Pi_{P_I^{(h)}}(V_{(1)},\ldots,V_{(k)})$ as in Definition 5 using the multilinear polynomial $P_I^{(h)}$. More concretely, for $J=(j_1,j_2,\ldots,j_k) \in \{1,\ldots,d\}^k$ the (J,a) entry of the matrix $\Pi_{P_I^{(h)}}(V_{(1)},\ldots,V_{(k)})$ will be defined by the relation

$$\pi_{J,a} = P_I^{(h)}(\phi^{(1)}_{j_1,a}, \phi^{(2)}_{j_2,a}, \ldots, \phi^{(k)}_{j_k,a}).$$

Using this definition, the following corollary follows immediately from Proposition B.1.

Corollary B.2. *Let $V_{(1)},\ldots,V_{(k)}$ be k matrices with $\{-1,1\}$ entries and dimensions $d \times n$. Then*

$$V_{(1)} \odot \ldots \odot V_{(k)} = c_h(k) \left(\sum_{I\subset[k]} (-1)^{k-|I|} \Pi_{P_I^{(h)}}(V_{(1)},\ldots,V_{(k)}) + \Pi_h(V_{(1)},\ldots,V_{(k)})\right).$$

Theorem B.3. *Let k, n, d be natural numbers. Consider a $d \times n$ matrix $\mathbf{V}$ with independent Bernoulli entries taking values -1 and 1 with probability $1/2$. Let h be a function from $\{-1,1\}^k \to \{-1,0,1\}$ having a representation as a multilinear polynomial of degree k. Then the matrix $\Pi_h(\mathbf{V},\dots,\mathbf{V})$ satisfies*

$$\Pr\left[\|\Pi_h(\mathbf{V},\dots,\mathbf{V})\| \geq c_6\left(\sqrt{d^k}+\sqrt{n}\right)\right] \leq c_7 \exp\left(-c_8\left(n^{\frac{1}{12k}}\right)\right).$$

The constants c_6, c_7, c_8 depend only on k.

Proof. From Corollary B.2, we know that

$$\Pi_h(\mathbf{V},\dots,\mathbf{V}) = \Pi_{P^{(h)}_{[k]}}(\mathbf{V},\dots,\mathbf{V}) = \frac{\mathbf{V}\odot\ldots\odot\mathbf{V}}{c_h(k)} - \sum_{I\subset[k]}(-1)^{k-|I|}\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V}). \tag{16}$$

To prove the theorem, we use induction over the size of I. Our inductive claim is that for every $I \subseteq [k]$,

$$\Pr\left[\left\|\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V})\right\| \geq c_6\left(d^{|I|/2}+\sqrt{n}\right)\right] \leq c_7\exp\left(-c_8 n^{\frac{1}{12|I|}}\right),$$

where constants c_6, c_7, c_8 depend only on $|I|$.
Step 1. Let $|I| = 1$. Then

$$\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V}) = c_9\mathbf{V}$$

for some constant c_9. Therefore,

$$\left\|\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V})\right\| = c_9\|\mathbf{V}\|.$$

Since $\mathbf{V}$ is a random $\{-1,1\}$ matrix, it is well known that (see e.g., [22]) there exists constant c_{10}, c_{11} such that

$$\Pr[\|\mathbf{V}\| \geq c_{10}(\sqrt{d}+\sqrt{n})] \leq \exp(-c_{11}n).$$

Therefore,

$$\Pr\left[\left\|\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V})\right\| \geq c_9c_{10}(\sqrt{d}+\sqrt{n})\right] \leq \exp\left(-c_{11}n\right).$$

Therefore, there exists constants c_6, c_7, c_8 such that

$$\Pr\left[\left\|\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V})\right\| \geq c_6\left(\sqrt{d}+\sqrt{n}\right)\right] \leq c_7\exp\left(-c_8 n\right) \leq c_7\exp\left(-c_8 n^{\frac{1}{12}}\right).$$

This completes the basis for induction.
Step 2. Let $|I| = k$, i.e., $I = \{1,\dots,k\} = [k]$. We want to bound $\left\|\Pi_{P^{(h)}_I}(\mathbf{V},\dots,\mathbf{V})\right\|$. By inductive hypothesis, we assume that for every $L \subset [k]$,

$$\Pr\left[\left\|\Pi_{P^{(h)}_L}(\mathbf{V},\dots,\mathbf{V})\right\| \geq c_6\left(d^{|L|/2}+\sqrt{n}\right)\right] \leq c_7\exp\left(-c_8\left(n^{\frac{1}{12|L|}}\right)\right), \tag{17}$$

where constants c_6, c_7, c_8 depend only on $|L|$. From Theorem 3.5, we know that the k-times entry wise product $\mathbf{V}\odot\cdots\odot\mathbf{V}$ satisfies the following norm condition (as $\mathbf{V}$ is matrix with independent τ-random entries)

$$\Pr\left[\|\mathbf{V}\odot\cdots\odot\mathbf{V}\| \geq c_3\left(\sqrt{d^k}+\sqrt{n}\right)\right] \leq \exp\left(-c_4 n^{\frac{1}{12k}}\right), \tag{18}$$

where again the constants c_3, c_4 depend only on k. From Equations (17) and (18), we get that there exists constants c_6, c_7, c_8 (depending only on k) such that

$$\Pr[\sum_{L\subset[k]} \left\|\Pi_{P_L^{(h)}}(\mathbf{V},\ldots,\mathbf{V})\right\| + \|\mathbf{V}\odot\cdots\odot\mathbf{V}\| \geq c_6\left(\sqrt{d^k}+\sqrt{n}\right)] \leq c_7\exp\left(-c_8 n^{\frac{1}{12k}}\right). \tag{19}$$

From Equation (16),

$$\begin{aligned}
&\Pr\left[\left\|\Pi_{P_I^{(h)}}(\mathbf{V},\ldots,\mathbf{V})\right\| \geq c_6\left(\sqrt{d^k}+\sqrt{n}\right)\right] \\
&= \Pr\left[\|\mathbf{V}\odot\ldots\odot\mathbf{V} - \sum_{L\subset[k]}(-1)^{k-|L|}\Pi_{P_L^{(h)}}(\mathbf{V},\ldots,\mathbf{V})\| \geq c_6(\sqrt{d^k}+\sqrt{n})\right] \\
&\leq c_7\exp\left(-c_8\left(n^{\frac{1}{12k}}\right)\right).
\end{aligned}$$

The last inequality comes by applying triangle inequality and then using from Equation (19). Note that for $I=[k]$, $\left\|\Pi_{P_I^{(h)}}(\mathbf{V},\ldots,\mathbf{V})\right\| = \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\|$. This completes the proof of the theorem. □

Theorem B.4. *Let k,q,n,d be natural numbers. Assume that $n \leq cd^k/\log_{(q)} d$. Consider a $d\times n$ matrix $\mathbf{V}$ with independent Bernoulli entries taking values -1 and 1 with probability $1/2$. Let h be a function $\{-1,1\}^k \to \{-1,0,1\}$ having a representation as a multilinear polynomial of degree k. Then the matrix $\Pi_h(\mathbf{V},\ldots,\mathbf{V})$ is with exponentially high probability a Euclidean section.*

Proof. Firstly note that $\Pi_h(\mathbf{V},\ldots,\mathbf{V})$ is an operator from $\mathbb{R}^n \to \mathbb{R}^{d^k}$. As mentioned before, by Cauchy-Schwarz's inequality

$$\forall \mathbf{x}\in S^{n-1} \;\; \sqrt{d^k}\,\|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_2 \geq \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_1.$$

Note that the proof idea of Theorem 3.8 also works for h. That is if $n \leq cd^k/\log_{(q)} d$ then

$$\Pr\left[\exists \mathbf{x}\in S^{n-1} \;\; \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_1 \leq C'd^k\right] \leq c_1\exp\left(-c_2 d\right).$$

In other words,

$$\Pr\left[\forall \mathbf{x}\in S^{n-1} \;\; \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_1 \geq C'd^k\right] \geq 1 - c_1\exp\left(-c_2 d\right).$$

Theorem B.3 implies that

$$\begin{aligned}
&\Pr[\forall \mathbf{x}\in S^{n-1} \;\; \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_2 \leq c_6\left(\sqrt{d^k}+\sqrt{n}\right)] \\
&\geq 1 - c_7\exp\left(-c_8\left(n^{\frac{1}{12k}}\right)\right).
\end{aligned}$$

If $n \leq cd^k/\log_{(q)} d$, then there exists a constant c_{13} (depending only on k) such that

$$\Pr\left[\forall \mathbf{x}\in S^{n-1} \;\; \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_2 \leq c_{13}\sqrt{d^k}\right] \geq 1 - c_7\exp\left(-c_8\left(n^{\frac{1}{12k}}\right)\right).$$

Therefore, with probability at least $1 - c_1\exp\left(-c_2 d\right) - c_7\exp\left(-c_8\left(n^{\frac{1}{12k}}\right)\right)$, there exists a α (depending only on k and q) such that

$$\forall \mathbf{x}\in S^{n-1} \;\; \|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_1 \geq \alpha\sqrt{d^k}\,\|\Pi_h(\mathbf{V},\ldots,\mathbf{V})\mathbf{x}\|_2.$$

This shows that the matrix $\Pi_h(\mathbf{V},\ldots,\mathbf{V})$ is with exponentially high probability a Euclidean section. □

C MLE's for Linear and Logistic Regression

Consider the linear regression problem $\mathbf{s} = X\theta + \varepsilon$. The likelihood function for linear regression (under the assumption that the entries in $\mathbf{s}$ are normally distributed) is:

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{s_i - \langle X_i, \theta\rangle}{2\sigma^2}\right),$$

where X_i is the ith row in X. The log-likelihood is:

$$\begin{aligned}\mathcal{L}_{\text{lin}}(\theta; \mathbf{s}, X, \sigma^2) &= \sum_{i=1}^{n} \ln(\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{s_i - \langle X_i, \theta\rangle}{2\sigma^2})) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\left(\frac{(\mathbf{s} - X\theta)^\top(\mathbf{s} - X\theta)}{\sigma^2}\right).\end{aligned}$$

For the logistic regression problem, the likelihood function (under the assumption that the entries in $\mathbf{s}$ are binomially distributed) is:

$$\prod_{i=1}^{n} \zeta_i^{s_i}(1-\zeta_i)^{1-s_i} = \prod_{i=1}^{n} \left(\frac{\zeta_i}{1-\zeta_i}\right)^{s_i} (1-\zeta_i).$$

The log-likelihood is:

$$\mathcal{L}_{\log}(\theta; \mathbf{s}, X) = \sum_{i=1}^{n} s_i\langle X_i, \theta\rangle - \ln(1 + \exp(\langle X_i, \theta\rangle)).$$

D Least Singular Value of Perturbed Random Matrices

In this section, we bound the least singular value of a random matrix perturbed by a low rank matrix. We need the following fact.

Lemma D.1. *Let $\mathbf{R}$ be a $d \times n$ random matrix with independent centered subgaussian entries with variances uniformly bounded below (τ-random entries fall into this category) and with $d \geq c_{14}n$. For any $\mathbf{z} \in \mathbb{R}^d$,*

$$\Pr\left[\exists \mathbf{x} \in S^{n-1} \ \|\mathbf{R}\mathbf{x} + \mathbf{z}\| \leq c_{15}\sqrt{d}\right] \leq \exp(-c_{16}d).$$

The constants c_{14}, c_{15}, and c_{16} are all independent of n and d.

For $\mathbf{z} = 0$ this follows from Proposition 2.5 [19] and the standard estimate of the norm of a subgaussian matrix (see e.g., Proposition 2.3 [18]). The proof for a general $\mathbf{z}$ follows the same lines.

The lemma below gives an estimate of the smallest singular value of a perturbed random matrix.

Lemma D.2. *Let $\mathbf{R}$ be a $d \times n$ random matrix with independent centered subgaussian entries with variances uniformly bounded below and $d \geq c_{14}n$. Let D be a deterministic $d \times n$ matrix such that rank$(D) = K$, $\|D\| \leq d^a$, where $a > 0$ is a constant. If*

$$\begin{cases} K \leq \frac{c_{17}d}{(a-1/2)\log d} & \text{if } a > 1/2, \\ K \leq c_{18}d & \text{if } a \leq 1/2. \end{cases}$$

then

$$\Pr[\sigma_n(\mathbf{R}+D) \le c_{19}\sqrt{d}] \le \exp(-c_{20}d).$$

The constants c_{17}, c_{18}, c_{19}, and c_{20} are all independent of n and d.

Proof. Set $Q = DB_2^n$ (where B_2^n denotes the unit Euclidean ball in $\mathbb{R}^n$), and let $\epsilon = c_{15}\sqrt{d}/2$. By the volumetric estimate, there exists an ϵ-net $\mathcal{N}$ in Q of cardinality at most

$$|\mathcal{N}| \le \left(\frac{3}{\epsilon}\right)^K \|D\| \le (c_{21}d)^{(a-1/2)K}$$

for $a > 1/2$ and $|\mathcal{N}| \le c_{22}^K$ for $a \le 1/2$. Assume that there exists $\mathbf{x} \in S^{n-1}$ such that $\|(\mathbf{R}+D)\mathbf{x}\| \le c_{15}\sqrt{d}/2$. Choose $\mathbf{z} \in \mathcal{N}$ so that $\|D\mathbf{x} - \mathbf{z}\| < \epsilon$. Then

$$\begin{aligned}\|\mathbf{R}\mathbf{x} + \mathbf{z}\| &\le \|(\mathbf{R}+D)\mathbf{x}\| + \|\mathbf{z} - D\mathbf{x}\| \\ &< c_{15}\sqrt{d}/2 + \epsilon = c_{15}\sqrt{d}.\end{aligned}$$

Lemma D.1 and the union bound yield

$$\begin{aligned}&\Pr[\exists \mathbf{x} \in S^{n-1} \ \|(\mathbf{R}+D)\mathbf{x}\|_2 \le c_{15}\sqrt{d}/2] \\ &\le \Pr[\exists \mathbf{x} \in S^{n-1} \exists z \in \mathcal{N} \ \|\mathbf{R}\mathbf{x}+\mathbf{z}\|_2 \le c_{15}\sqrt{d}] \\ &\le |\mathcal{N}| \cdot \exp(-c_{16}d) \\ &\le \exp(-c_{20}d).\end{aligned}$$

The last inequality follows from the assumption on K. Here, c_{21} and c_{22} are constants independent of n and d. □