Differentially Private Linear Algebra in the Streaming Model

Jalaj Upadhyay
University of Waterloo.
jalaj.upadhyay@uwaterloo.ca

Abstract

The focus of this paper is a systematic study of differential privacy on streaming data using sketch-based algorithms. Previous works, like Dwork *et al.* (ICS 2010, STOC 2010), explored random sampling based streaming algorithms. We work in the well studied streaming model of computation, where the database is stored in the form of a matrix and a curator can access the database row-wise or column-wise. Dwork *et al.* (STOC 2010) gave impossibility result for any non-trivial query on a streamed data with respect to the user level privacy. Therefore, in this paper, we work with the event level privacy. We provide optimal, up to logarithmic factor, space data-structure in the streaming model for three basic linear algebraic tasks in a differentially private manner: matrix multiplication, linear regression, and low rank approximation, while incurring significantly less additive error.

The mechanisms for matrix multiplication and linear regression can be seen as the private analogues of known non-private algorithms, and have some similarities with Blocki *et al.* (FOCS 2012) and Upadhyay (ASIACRYPT 2013) on the superficial level, but there are some subtle differences. For example, they perform an affine transformation to convert the private matrix in to a set of $\{\sqrt{w/n}, 1\}^n$ vectors for some appropriate w, while we perform a perturbation that raises the singular values of the private matrix. In order to get a streaming algorithm for low rank approximation, we have to reuse the random Gaussian matrix in a specific way. We prove that the resulting distribution also preserve differential privacy. We do not make any assumptions, like singular value separation, as made in the earlier works of Hardt and Roth (STOC 2013) and Kapralov and Talwar (SODA 2013). Further, we do not assume normalized row as in the work of Dwork *et al.* (STOC 2014). All our mechanisms, in the form presented, can also be computed in the distributed setting of Biemel, Nissim, and Omri (CRYPTO 2008).

Keywords. Differential Privacy, Linear Algebra, Random Projection.

Table of Contents

1 Introduction

In the setting of large data analysis, one of the desired goals is to design a sub-linear space data-structure to perform computation while receiving the data online. There are many non-private algorithms in this setting. Such algorithms are called *streaming algorithms*. However, these data often contain sensitive information and privacy is as important as correct computation. Considering these two issues, a natural problem that a database curator faces is to generate a data structure that could be used to provide useful information without leaking sensitive information about an individual. The focus of this paper is space restricted streaming algorithms for answering linear algebraic queries in a private manner.

For an input matrix A, the standard techniques used in (non-private) streaming algorithms either does *random sampling* or compute a *sketch*, which has the form ΩA for some choice of random matrix Ω. Dwork *et al.* [19] gave a private analogue of streaming algorithms that uses *sampling based approach* for various statistical queries, and gave an impossibility result for private analogues of *sketch based approaches* for specific "statistical queries." This raises doubts over the applicability of sketch based approach in privacy. These doubts, if true, would be unfortunate because sketch generation is one of the major techniques (and often provide better utility guarantee) in the non-private setting. In this paper, *we show the first set of positive results for sketch based approach*. We deviate from the traditional mechanisms of differential privacy: instead of perturbing the output of the query by adding noise, we reversibly perturb the input and then multiply noise (analogous to [5, 6, 59]). We give almost optimal (in terms of space required by the data-structure) differentially private sketch based streaming algorithms for three basic algorithmic problems in linear algebra.

A natural question one might ask is, why should one care about private streaming algorithms for linear algebra? To answer this, let us consider the following scenario. A large database is streamed to the curator such that, unless we store the data, it is irretrievably gone. On the other hand, the curator has limited memory and it cannot store the whole database, but expects queries on the streamed data. In such a setting, the curator would like to store a data-structure with enough information about the database to (approximately) answer the queries. This idea has been used in the streaming model without any privacy concern. However, if a curator is handling a confidential data-base, it has to store a data-structure that, in addition to providing useful information to the query-maker, does not leak any information about the individual entry of the database as per the specific requirements of a robust privacy guarantee.

The scenario mentioned above is not an artificial problem. An $n \times d$ real-valued matrix is a natural structure for storing data about n entities described by d features. For example, Hardt and Roth [33, 34] motivated the problem of differentially private low-rank approximation (LRA) by citing the Netflix competition. In addition to the Netflix type scenario, there are many other areas of large data analysis that have natural privacy concerns, like, genetics engineering finance. Computations in financial market often use various linear algebraic tasks as subroutine, like matrix multiplication or linear regression. Likewise, in genetic engineering, it is common to perform a procedure that computes a full or partial singular value decomposition (SVD) of the covariance matrix corresponding to the input data matrix, and then appeal to standard statistical model selection criterion, like getting the top half of the spectrum of the matrix, to quantify its significance.

On the other side, the traditional methods of *Krylov subspace iteration* (on which some of the recent works like [31, 34, 40] are based) and *rank revealing factorization method* requires a lot of space and are slow when matrices have high dimension. In fact, even storing the whole data during the computation is not always possible. Therefore, in such scenarios, computations are done in the streaming model. Many of these tasks have been studied in the non-private setting [2, 12, 24, 47, 54]. However, recent privacy violations in large data analyses have exemplified how sensitive these data are. This raises the question of whether one can perform all these tasks on a streamed data while giving a robust guarantee of privacy, like differential privacy. We do a principled study of private analogues of the known streaming algorithms for such tasks.

PRIVACY MODEL USED IN THIS PAPER. In this paper, we consider differential privacy in the streaming model where the data is streamed either row-wise or column-wise and the space available to the curator is sub-linear in the size of the dataset. There are two notions of differential privacy: *event level privacy*, where guarantees are at the granularity of individual records in the datasets, and *user level privacy*, where guarantees are at the granularity of each user whose data is present in the dataset. Dwork *et al.* [18] showed that it is impossible to obtain any non-trivial result with respect to the user level privacy on a streamed data. Therefore, in this paper, we restrict our attention to the event level privacy, i.e., D_1 and D_2 are *neighbouring* if $\|D_1 - D_2\| \leq 1$ (see [33, 34] for the motivation of this notion).We work with the following notion of privacy.

Definition 1. A randomized mechanism $\mathcal{K}$ gives (ε, δ)-*differential privacy*, if for all neighbouring data-sets D_1 and D_2, and all range $S \subset Range(\mathcal{K})$, $\mathsf{Pr}[\mathcal{K}(D_1) \in S] \leq \exp(\varepsilon)\mathsf{Pr}[\mathcal{K}(D_2) \in S] + \delta$, where the probability is over the coin tosses of $\mathcal{K}$. When $\delta = 0$, we get the traditional definition of *differential privacy*.

1.1 Problem statements and our contributions.

In this section, we give the formal description of the problems we investigate in this paper. We also state the best known space lower bounds for non-private streaming algorithms and present our results. We note that stronger space lower bounds are achievable for each of the problems considered because of the non-zero additive error; however, the bounds stated below are presently the best known lower bounds that we can use to make any sort of optimality comparison.

We reserve the letter n for the number of rows and d for the number of columns of a private matrix. We assume $d < n$. The performance of a streaming algorithm is measured by three basic factors: the number of passes over the data stream, the space used by the data-structure, and the time taken to update the data-structure. All our private mechanisms for performing linear algebraic tasks are single-pass, achieve almost optimal space bound for one-pass algorithms, and almost linear update time. Jumping ahead, we generate linear sketch, so our mechanisms extends naturally to turnstile updates. For *bit complexity*, we use the convention of Clarkson-Woodruff [12]–the entries of a matrix can be represented by $\kappa = \log(nd)$-bit integers.

MATRIX MULTIPLICATION. The first problem we consider is matrix multiplication of two conforming matrices. It is one of the most important tools in numerical analysis (for example, every linear differential equations solver uses matrix product), and, therefore, wherever private data are analyzed numerically.

Problem 1. *((α, β, τ)-*MAT-MULT*).* An $n \times d_1$ matrix A and $n \times d_2$ matrix B are given. Output a matrix C so that $\|A^{\mathsf{T}}B - C\|_F \leq \alpha\|A\|_F \cdot \|B\|_F + \tau$ with probability at least $1 - \beta$.

Theorem 1. *[12]* Suppose $n \geq c\kappa/\alpha^2$ for an absolute constant $c > 0$. Then any randomized one-pass algorithm which solves$(\alpha, \beta, 0)$-MAT-MULT with probability at least $4/5$ uses $\Omega(\max\{d_1, d_2\}\,\alpha^{-2}\kappa)$ bits of space.

Kane and Neilson [39, Theorem 6.2] used the Minkowski inequality to prove the following.

Theorem 2. Given $\alpha, \beta > 0$, and conforming matrices A and B presented column-wise. There is an $r = O(1/\alpha^2)$ so that for $d = \max\{d_1, d_2\}$ large enough, there is data structure which maintains a sketch of size $O(rd\log(1/\beta)\log\kappa + \log(1/\alpha))$ and $\tilde{O}(r)$ update time and solve $(\alpha, \beta, 0)$-MAT-MULT for input matrix A and B.

Theorem 22 gives a data-structure which maintains a sketch of size $O\left(d\alpha^{-2}\kappa\log(1/\beta)\right)$ and provides (ε, δ)-differential privacy with $\tau = O(\alpha\sqrt{n})$. If we are ready to pay for space, then using Theorem 27, we have $\tilde{O}(n)$ update time at a cost of $O\left(d\alpha^{-2}\kappa\log(1/\beta)\log n\right)$ space data-structure.

LINEAR REGRESSION. The second problem we consider is *linear regression*, an important tool in statistical analysis. One of its major application is in finance, a domain where data are extremely sensitive. One

example is *capital asset pricing model* which is used to predict demands [14] and investment [23].

Problem 2. *((α, β, τ)-LIN-REG).* Given an $n \times d$ matrix A and a m set of $n \times 1$ column vectors $B = \{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$, output a set of vectors $X = \{x_1, \cdots x_m\}$ so that for all $i \in [m]$, $\|AX - \mathbf{b}_i\|_F \leq (1 + \alpha) \min_{Y \in \mathbb{R}^{d \times m}} \|AY - \mathbf{b}_i\|_F + \tau$ with probability at least $1 - \beta/m$.

LIN-REG has also been studied in online learning model [37], which, though a different model, is closest to ours if one wishes to make any comparison. Jain, Kothari, and Thakurta [37] gave a private learning algorithm for LIN-REG with a bound of $\tilde{O}\left((R^6 \log(1/\delta)\sqrt{n} \log^{1.5} T)/\sqrt{\varepsilon}\alpha^3\right)$, where R is the maximum 2-norm of query input and T is the number of training data set. We better this bound by factor of $\alpha^4\sqrt{\varepsilon}$ in Theorem 24.

Theorem 3. *[12]* Suppose $n \geq \kappa d/36\alpha$ and d is sufficiently large. Then any randomized one-pass algorithm which solves the (α, β, τ)-LIN-REG problem with probability at least $7/9$ needs $(d^2\alpha^{-1}\kappa)$ bits of space.

Kane and Neilson [39, Theorem 6.5] showed the following in the non-private setting.

Theorem 4. There is a one-pass streaming algorithm for $(\alpha, \beta, 0)$-LIN-REG where one maintains a sketch of size $O(d^2\alpha^{-1} \log(1/\beta) \log(nd))$. Processing each update requires $O(d + \sqrt{d/\alpha} \log(1/\beta))$ computations.

Theorem 22 gives a data-structure that uses $O\left(d^2\alpha^{-1}\kappa \log(1/\beta)\right)$ bits of space and provides (ε, δ)-differential privacy with $\tau = O(\alpha\sqrt{n})$. If we are ready to pay for space, then using Theorem 27, we have $\tilde{O}(n)$ update time at a cost of $O\left(d^2\alpha^{-1}\kappa \log(1/\beta) \log n\right)$ space data-structure.

LOW-RANK APPROXIMATION. The last problem we consider is *low rank approximation* [15]. A partial list of its applications includes principal component analysis [35], fast multipole methods [26], and $\mathcal{H}$-matrices [25].

Problem 3. *((α, β, τ)-LRA).* Given an $n \times d$ matrix A and a target rank k, construct a matrix Ψ with k orthonormal columns such that $\|A - \Psi\Psi^{\mathsf{T}}A\| \leq (1 + \alpha) \min_{\mathsf{rank}(A_k) \leq k} \|A - A_k\| + \tau$ with probability at least $1 - \beta$ for both Frobenius and spectral norm.

Theorem 5. *[12]* Suppose $n \geq ck/\epsilon$ for an absolute constant $c > 0$, . Then any randomized 1-pass algorithm which solves *k-rank approximation* with probability at least $5/6$ uses $\Omega(nk\alpha^{-1})$ bits of space.

Clarkson and Woodruff [12] gave a (non-private) one-pass algorithm for LRA. They used Rademacher matrices, i.e., matrices with entries ± 1 with probability $1/2$, to produce the sketch and an observation that the *projection* step can be emulated by the information gathered during the first stage if one uses Rademacher matrices. Using their idea, Kane and Neilson [39, Theorem 6.6] showed the following in the non-private setting.

Theorem 6. There is a one-pass streaming algorithm for approximate LRA with respect to the Frobenius norm with row/column-wise updates where one maintains a sketch of size $O(k\alpha^{-1}(n+d) \log(1/\delta) \log(nd))$. Processing each update requires $O(k + \sqrt{k/\alpha} log(1/\beta))$ amortized operations per entry of the input matrix.

We do not know a way to prove differential privacy using [12], except to use additive noise mechanisms, much like Hardt and Roth [33]. In that case, when we try to emulate the second step, a simple analysis shows an error bound of order $k^{3/2}$. In this paper, we show how to emulate the projection step when using random Gaussian matrices. We give an $O(k\varepsilon^{-1}(n + d)\kappa)$ bits data structure (Theorem 26) that could be used to publish (ε, δ)-differentially private k-rank approximation of an $n \times d$ private matrix in a single pass with $\tau \leq O(k\sqrt{n \ln(2/\delta)}/\varepsilon)$ for the Frobenius and $\tau \leq O(\sqrt{nk \ln(2/\delta)}/\varepsilon)$ for the spectral norm.

Note that our data-structure has an improved space requirement by a factor of $\log(1/\beta)$; thereby, just off by a factor of $1/\alpha$ from the optimal. This is not the first time that a method used in private setting has helped to improve the result in non-private setting. Dwork *et al.* [21] have recently illustrated another case. The reason behind of our improvement is simple. The bound on the single pass algorithm of Clarkson and Woodruff [12] and the two-pass algorithm of Sarlos [54] use their bound for MAT-MULT, and, therefore, had to rely on the number of rows in the projection matrix used to bound the error in MAT-MULT. On the

Method	Norm	Additive noise	Privacy Notion	Streaming	# Passes
Hardt-Roth [33]	F	$\frac{\sqrt{kn}\log(k/\delta)}{\varepsilon} + \sqrt{\frac{\mu\|A\|_F \log(k/\delta)}{\varepsilon}}$	Event level	No	2
Subspace Iteration [34]	S	$O(\frac{k^2}{\varepsilon}\sqrt{(\mathsf{rk}(A)\mu + k\log n)\log\left(\frac{1}{\delta}\right)\log n})$	Event level	No	$k\sqrt{\log \lambda}$
Kapralov-Talwar [40]	S	$O(dk^3/(\varepsilon\gamma^2\delta^2))$	Spectral norm	No	k
Hardt [31]	S	$\frac{\lambda_1\sqrt{kn\mu\log(1/\delta)\log(n/\gamma)\log\log(n/\gamma)}}{\varepsilon\gamma^{1.5}\lambda_k}$	User level	No	$k\sqrt{\log \lambda}$
Dwork *et al.* [21]	S	$O((k\sqrt{n}\ln(1/\delta))/\varepsilon) + \tilde{O}(\sqrt{k^3n^{3/2}}/\varepsilon^2)$	User level	Yes	1
This paper	F	$O(\sqrt{nk\ln(2/\delta)}/\varepsilon)$	Event level	Yes	1
This paper	S	$O(\sqrt{nk\ln(2/\delta)}/\varepsilon)$	Event level	Yes	1

Table 1: Comparison Between our Mechanism and Previous Mechanisms for k-Rank Approximation of an $n \times d$ matrix A. μ denotes the coherence of A, $\gamma = (\lambda_k/\lambda_{k+1}) - 1$, S stands for spectral norm and F for Frobenius norm.

	Space Bound on the Data-Structure			
	Lower Bound [12]	Non-private [39]	Private	Additive Noise
(α,β,τ)-MAT-MULT	$\Omega\left(d\alpha^{-2}\kappa\right)$	$\tilde{O}(d\alpha^{-2}\kappa)$	$\tilde{O}(d\alpha^{-2}\kappa)$	$O\left(\sqrt{n}\alpha\right)$
(α,β,τ)-LIN-REG	$\Omega\left(d^2\alpha^{-1}\kappa\right)$	$\tilde{O}\left(d^2\alpha^{-1}\kappa\right)$	$\tilde{O}\left(d^2\alpha^{-1}\kappa\right)$	$O\left(\sqrt{n}\alpha\right)$
Frobenius (α,β,τ)-LRA	$\Omega(nk/\alpha)$	$\tilde{O}(k\alpha^{-1}(n+d)\kappa)$	$O(k\alpha^{-1}(n+d)\kappa)$	$O(k\sqrt{n\ln(2/\delta)}/\varepsilon)$
Spectral (α,β,τ)-LRA	–	–	$O(k\alpha^{-1}(n+d)\kappa)$	$O(\sqrt{nk\ln(2/\delta)}/\varepsilon)$

Table 2: Our Results with Respect to the Best Known Space Bounds in Non-private Setting ($\tilde{O}$ hides $\log(1/\beta)$ term).

other hand, as we discuss later, we use perturbation theory along the line of [21, 30].

Let $\lambda_1, \cdots, \lambda_{\mathsf{rank}(A)}$ be the singular values of a matrix A. We compare our results with the earlier known results in Table 1 and 2. In Table 1, we compare our result for LRA with the previous works. We do not make any assumptions, like singular-value separation, normalized rows, etc, on the input matrix to compute LRA as in the previous works [21, 34, 40]. If we do not make any coherence assumption, one should set $\mu = n$ in Table 1. A detail comparison is done in Section 3.3.1 taking into account the difference in privacy model and the assumptions. In Table 2, we give the best known lower bound for the space required for each of the problems, the space required by the data-structures in the non-private setting and by our private data-structures. The other parameters used in the tables are as defined above. Note that, we explicitly require Ω to be stored for the mechanism of LRA; in the other two cases, Ω can be picked from the distribution on the fly.

Our mechanism for LRA can be easily compiled to give differentially private *principal component analysis* using standard algorithms that use LRA in the first step. Another important application of our mechanism for private sketch generation is in manifold learning. We do not formally state these mechanisms as there are standard algorithms for these applications that only use private matrix for one of the problems stated above and rest of the steps are deterministic function of these computations. One can also implement our mechanisms *as distributed algorithms*, a desirable feature as argued by [3]. This is because every operations used in our mechanism have efficient distributed algorithms–Jacobi method for singular value decomposition [42, Chapter 4], Cannon's algorithm for multiplication [10], and GMRES for residual method [53].

1.2 Our Techniques.

The two standard techniques for streaming algorithms are (i) random sampling of the rows or columns of the streamed matrix and (ii) generating a random *sketch* of the matrix. In this paper, we use the sketch based approach. A sketch of a matrix A has the form ΩA for some appropriate choice of random matrix Ω. The known sketch based streaming algorithms for MAT-MULT and LIN-REG, to our knowledge, use tug-of-war matrices [12, 54] with Rademacher entries (this helps in improving the update time). Adding Gaussian noise to it to ensure differential privacy amounts to a large additive error.

In order to get a better utility bound, we use an idea analogous to Blocki *et al.* [5, 6] and Upadhyay [59]. We devise a private-sketch generation (PSG) mechanism to generate a private sketch of the private matrix, where the privacy depends on the spectral property of the input matrix. At a high level, all our mechanisms use this basic mechanism while maintaining the spectral property of the input matrix (to guarantee privacy). If we use affine transformation as in [5, 59], we would end up with τ dependent on $\|A\|_F$. Therefore, we use different method to maintain the required spectral property. If we instantiate PSG with random Gaussian matrix, we get the same bound on the space used by the data-structure as achieved by Clarkson and Woodruff [12].

Due to its generic construction, we can instantiate PSG with other known projection matrices that preserves privacy. For example, we can use fast-Johnson-Lindenstrauss transform [1], the diagonal matrix based construction or circulant matrix based construction, which were proven to preserve privacy in [59, 58, 60], respectively. In this paper, we explicitly use the last two constructions, wherein, the author proved two constructions of the form PWD, where W is the Walsh-Hadamard matrix, D is a diagonal Rademacher matrix, and P chosen appropriately, preserves privacy. For example, let $\mathbf{g} := (\mathbf{g}_1, \cdots, \mathbf{g}_n)$ be n i.d.d. Gaussian samples. Then if we use $P = \Pi_{1..r}C$, where C be a circulant matrix corresponding to $\mathbf{g}$, i.e, for $i \in [n]$, $C_{i:} = (\mathbf{g}_i, \cdots, \mathbf{g}_n, \mathbf{g}_1, \cdots, \mathbf{g}_{i-1})$ and $\Pi_{1..r}$ is a permutation matrix Π truncated to r rows, we get the construction in [60]. On the other hand, if we use $P = \Pi_{1..r}\mathsf{Diag}(\mathbf{g})\Pi$, we get the construction in [58]. The proof of privacy when using these constructions requires an entirely different analysis than [5, 59]. This is because the published matrix is not r-independent multi-variate Gaussian due to subtle correlations between the rows of the published matrix.

At a high level, our mechanisms for (α, β, τ)-MAT-MULT and (α, β, τ)-LIN-REG are private analogues of Clarkson-Woodruff [12]. For the privacy proof to go through, we need to lift the singular value of the input matrix. Our choice to lift the singular values is constrained by keeping a check on τ as well as to keep the mechanism one-pass. Once we find an optimal choice, the utility proof follows by considering the effect of our perturbation in the analysis of Kane and Neilson [39] (combined with our bound in Theorem 27 for projection matrices for mechanisms in Section 4 and Gaussian matrices bound stated in Theorem 18 for Section 2 and 3). Note that we cannot use the variance bound for 4-wise independent matrices used by other non-private algorithms [12, 54] as privacy would not necessarily hold for random matrices with dependent entries in its row.

The mechanism for LRA is more complicated and markedly different from the recent private mechanisms [21, 31, 33, 34, 40] (the online version of Dwork *et al.* [21] uses binary tree technique [18] and assumes a lower bound condition on the optimal value, see [21, Theorem 8]) as well as the non-private algorithms [12, 54]. All the previous works perturb the output by adding noise to it, while we perturb the input matrix and then multiply noise matrix. We follow the general prototype to compute a LRA, i.e., first computes a projection matrix (*range finding* step) followed by the computation of a LRA by operating the projection matrix on the input matrix (*projection* step). In the most naive form, both the steps require the input matrix.

Clarkson and Woodruff [12] have shown how to emulate the projection step by using the matrices formed in the range finding step, without reusing the input matrix. Unfortunately, it only works for Rademacher matrices, while our projection matrices are generated using random Gaussian variables.

The first key observation is that, by a clever use of linear algebra, information gathered in the range-finding step using a random Gaussian matrix can be used to emulate the projection step without using the input matrix explicitly. However, we need to reuse the random Gaussian matrix. Therefore, the privacy is not as straightforward as for the other two problems. Fortunately, the Gaussian matrix is reused in a specific manner for which one can prove privacy under certain spectral property of the input matrix. We believe that this could be of independent interest. The second observation, also done by Blocki *et al.* [5], is that the mechanism for PSG already gives considerable improvement in the range finding step. Following Hardt [31], use an oversampling parameter p, which helps in getting much sharper bounds for both the

spectral and Frobenius norm. However, we have to use a different method than used in matrix multiplication and linear regression to lift the singular value to keep α in check.

The analysis for LRA used in this paper differs a lot from the analyses of Clarkson and Woodruff [12] and Sarlos [54], and is more align with perturbation theory based approach of [21, 30]. The non-private algorithms in [12, 54] use the trick that a good bound on matrix multiplication allows a bound on LRA. This limits its applicability to the case when the approximation metric is the Frobenius norm. In this paper, we are also concerned with the spectral norm. Moreover, the algorithm of Sarlos [54] perform two-passes over the private input matrix. The proof of Theorem 26 presented here use perturbation theory, which allows us to give bounds for both the norms in an unified manner. We note that all the previous proofs were tailored for specific norm.

RELATED WORKS. The first formal definition of Differential Privacy was given by Dwork *et al.* [17]. They used Laplacian distribution to guarantee differential privacy for bounded *sensitivity* query functions. The Gaussian variant of this basic sanitizer was proven to preserve differential privacy by Dwork *et al.* [16] in a follow-up work. Since then, many mechanisms for preserving differential privacy have been proposed in the literature [8, 20, 27, 28, 41, 32, 33, 44, 51]. All these sanitizers have a common theme: they perturb the output before responding to queries. Blocki *et al.* [5, 6] and Upadhyay [59] took a complementary approach. They perturb the input reversibly and then perform a random projection of the perturbed matrix.

There are some recent works on differentially private low-rank approximation and differentially private streaming algorithm for statistical queries. Blum *et al.* [7] first studied this problem and gave a simple "input perturbation" algorithm that adds noise to the covariance matrix, an approach also taken recently by Dwork *et al.* [21]. This was improved by Hardt and Roth [33] who studied the low rank approximation in Frobenius norm under the low coherence assumption. Kapralov and Talwar [40] and Chaudhary *et al.* [11] studied the spectral low rank approximation of a matrix by giving a matching upper and lower bounds for privately computing the top k eigenvector of a matrix. Hardt and Roth [34] improved their noise bound by proposing robust private subspace iteration mechanism. All these works [11, 34, 40] uses some eigenvalue separation assumption. Recently, Dwork *et al.* [21] revisited randomized mechanism to give a tighter bound. They also gave an online learning version of their mechanism under a normalized row assumption.

The literature of performing (non-private) linear algebra using streaming algorithms, started by Munro and Paterson [47], is so extensive that we cannot hope to cover it in any detail here. In the private setting, Dwork *et al.* [19] studied *pan-privacy*, where the internal state is known to the adversary, to answer various counting tasks, like estimating distinct elements, cropped means, number of heavy hitters, and frequency counts. All these mechanisms uses private version of various sampling based streaming algorithms. Subsequently, there have been some works on online differential privacy [21, 37] for various tasks.

NOTATIONS AND BASIC PRELIMINARIES We reserve the letters A and B for private input matrices, Ω for a random projection matrix (Gaussian or the one defined in Section 2). For an $n \times d$ matrix A, we let $\mathbb{A}$ denote n copies of A stack row-wise, $A_{i:}$ to denote the i-th row of A, $A_{:j}$ to denote the j-th column of A, and A' to denote the symmetric matrix $\begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix}$ corresponding to A. We let A_t denote the matrix received after t time epochs. When we wish to refer to both the Frobenius as well as the spectral norm, we overload the symbol $\|\cdot\|$ and drop the subscript. We let $\mathbf{e}_1, \cdots, \mathbf{e}_d$ denote the standard basis vectors in $\mathbb{R}^d$. We use bold face symbols to denote vectors and $\mathbf{0}^n$ to denote an n-dimensional 0-vector. For a matrix M, we write $M \succ 0$ if all its eigenvalues are positive.

1.3 Useful Facts Used in the Paper

In the course of this paper, we use many standard results from linear algebra, random matrices, perturbation theory, and differential privacy literature.

LINEAR ALGEBRA. Our analysis make extensive use of linear algebra and statistical properties of Gaussian distribution. We give an exposition to the level required to understand this paper. Let A be an $n \times d$ matrix. The singular value decomposition (SVD) of A is $A = V\Lambda U^{\mathsf{T}}$, where U and V are left and right eigenvectors of A, and Λ is a diagonal matrix. The entries of Λ are called the *singular values* of A. Since U and V are unitary matrices, one can write $A^i = V\Lambda^i U^{\mathsf{T}}$ for any real value i. We let $\mathsf{rank}(A)$ denotes the rank of the matrix A and $\lambda_i(A)$ its singular values. Where it is clear from context, we simply write λ_i for the singular values.

We use various matrix norm. We use the notation $\|\cdot\|_F$ for Frobenius norm. A Frobenius norm for a matrix $A = (a_{ij})_{i\in[n],j\in[d]}$ is defined as following $\|A\|_F = \sum_{ij} |a_{ij}|^2$. For a matrix A, we let $\|A\|_2$ denote the 2-norm, i.e., $\max_{x\in\mathbb{R}^d} \|Ax\|_2/\|x\|_2$. We use the symbol $\|\cdot\|$ when we wish to refer to both Frobenius as well as 2-norm. We explicitly or implicitly use the fact that matrix norms are Lipschitz. We let $e_1, \cdots, e_n$ denote the standard basis vectors in $\mathbb{R}^n$. We denote by $a|b$ the vector formed by appending the vectors a and b. A matrix M is *positive semi-definite* if all its eigenvalues are non-negative, i.e., if for all $\mathbf{x} \in \mathbb{R}^n$, we have $\mathbf{x}^{\mathsf{T}} M\mathbf{x} \geq 0$. For two $n \times n$ matrices M and N, we denote by $M \succeq N$ if $M - N$ is a positive semi-definite matrix. We write $M \succ 0$ if all its eigenvalues are positive. We note few of the key lemmata of linear algebra used in this paper.

Lemma 7. If matrix A and B are conforming, then $\|AB\|_F \leq \|A\|_2\|B\|_F$.

Lemma 8. *[29].* Let A and B be Hermittian matrices with only $0, 1$ entries. Then $\mathrm{Tr}(AB) \leq \mathrm{Tr}(A)\mathrm{Tr}(B)$. Moreover, if $\|A - B\| \leq 1$, then $\mathrm{Tr}(A^{\mathsf{T}}A) - \mathrm{Tr}(B^{\mathsf{T}}B) \leq 2$.

Lemma 9. For matrices A, B, C, D, E, and X of appropriate dimensions, we have

1. $\mathsf{vec}(AXB) = (B^{\mathsf{T}} \otimes A)\mathsf{vec}(X)$,
2. $\mathrm{Tr}(CXB) = (\mathsf{vec}(C^{\mathsf{T}}))^{\mathsf{T}}(\mathbb{I}_q \otimes X)\mathsf{vec}(B)$,
3. $\mathrm{Tr}(DX^{\mathsf{T}}EXB) = (\mathsf{vec}(X))^{\mathsf{T}}(D^{\mathsf{T}}B^{\mathsf{T}} \otimes E)(\mathsf{vec}(X)) = (\mathsf{vec}(X))^{\mathsf{T}}(BD \otimes E^{\mathsf{T}})\mathsf{vec}(X)$.

STATISTICAL MODEL SELECTION AND PROBABILITY THEORY. One of the main methods to prove concentration inequalities is the following two step process: control the moment generating function of a random variable and then minimize the upper bound resulting from the Markov's inequality. Though simple, it is extremely powerful. In this paper, we use the result by Birge-Massart [4] stated by Corollary 14, but we need some concepts in statistical model selection to understand the result.

Let ζ be a real valued centered random variable, then the log-moment generating function is defined as $\psi_\zeta(\lambda) := \ln(\mathbb{E}[\exp(\lambda\zeta)]), \forall\lambda \in \mathbb{R}_+$, and the *Cramer's transform* is defined as $\psi^*_\zeta(x) := \sup_{\lambda\in\mathbb{R}_+}(\lambda x - \psi_\zeta(\lambda)$. The *generalized inverse* of ψ^* at a point t is defined by $\psi^{*-1}(f) := \inf\{x \geq 0 : \psi^*(x) > f\}$.

The log generating function for centered random variable has some nice properties. It is continuously differentiable in a half-open interval $I = [0, b)$, where $0 < b \leq \infty$, and both ψ_ζ and its differentiation at 0 equals 0. There is a nice characterization of the generalized inverse in the form of following lemma.

Lemma 10. Let ψ be a convex continuously differentialable function on I. Assume that $\psi(0) = \psi'(0) = 0$. Then ψ^* is non-negative non-decreasing convex function on $\mathbb{R}_+$. Moreover, its generalized inverse can be written as $\psi^{*-1} = \inf_{\lambda\in I}\left[(f + \psi(\lambda))/\lambda\right]$.

This lemma follows from the definition and basic calculus. In the area of model selection, Lemma 10 is often used to control the expectation of the supremum of a finite family of exponentially integrable variables. Pisier [48] proved the following fundamental lemma.

Lemma 11. (Pisier [48]) Let $\{\zeta_f\}_{f\in F}$ be a finite family of random variables and ψ be as in Lemma 10. Let $\mathbb{E}^A[\zeta] = \mathbb{E}[\zeta\chi_A]/\mathsf{Pr}[A]$ for a non-zero measurable set A. Then, for any non-zero measurable set A, we have $\mathbb{E}^A\left[\sup_{f\in F}\zeta_f\right] \leq \psi^{*-1}\left(\ln\left(|F|/\mathsf{Pr}[A]\right)\right)$.

If we take $A = (\zeta \geq \phi(x))$ and applying Markov's inequality, then using the property that ϕ is an increasing function, this immediately gives us that $x \leq \ln(1/\mathsf{Pr}[A])$. This gives the following key lemma.

Lemma 12. Let A be a set with non-zero measure and ζ be a centered random variable. Let ϕ be an increasing function on positive reals such that $\mathbb{E}^A[\zeta] \leq \phi(\ln(1/\mathsf{Pr}[A]))$. Then $\mathsf{Pr}[\zeta \geq \phi(x)] \leq \exp(-x)$.

We refer the interested readers to the book by Massart and Picard [43]. In this paper, we use the following result by Birge and Massart [4] for this purpose.

Theorem 13. (Birge-Massart [4]) Let $(\zeta_f)_{\mathcal{F}}$ be a finite family of random variable and ψ be a convex and continuously differentiable function on $[0, b)$ with $0 \leq b \leq \infty$ such that $\psi(0) = \psi'(0) = 0$ and for every $u \in [0, b)$ and $f \in \mathcal{F}$, we have $\log(\mathbb{E}[\exp(u\zeta_f)]) \leq \psi(u)$. If N denotes the cardinality of $\mathcal{F}$. Then $\mathbb{E}\left[\sup_{f\in\mathcal{F}}\zeta_f\right] \leq \psi^{*-1}(\ln N)$, where ψ^* is the Cramer's transformation.

Using Lemma 12 and Talagrand inequality, the authors also proved the following corollary to Theorem 13.

Corollary 14. (Birge-Massart [4]) Let $0 < \lambda < 1/b$ for some b. If ζ be a real valued integrable variable, and a and b be constants such that $\log(\mathbb{E}[\exp(\lambda\zeta)]) \leq \frac{a\lambda^2}{2(1-b\lambda)}$. Then $\mathsf{Pr}\left[\zeta \geq \sqrt{2a\tau} + b\tau\right] \leq \exp(-\tau)$.

A random variable, X, distributed according to a Gaussian distribution has the probability density function, $\mathsf{PDF}_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. We denote it by $X \sim \mathcal{N}(\mu, \sigma^2)$. The Gaussian distribution is invariant under affine transformation, i.e., if $X \sim \mathcal{N}(\mu_x, \sigma_x)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y)$, then $Z = aX + bY$ has the distribution $Z \sim \mathcal{N}(a\mu_x + b\mu_y, a\sigma_x^2 + b\sigma_y^2)$.

The multivariate Gaussian distribution is a generalization of univariate Gaussian distribution. Given a m dimensional multivariate random variable, $X \sim \mathcal{N}(\mu, \Sigma)$, the PDF of a multivariate Gaussian is given by $\mathsf{PDF}_{\mathbf{X}}(\mathbf{x}) := \frac{1}{\sqrt{(2\pi)^{\mathsf{rank}(\Sigma)}\mathsf{Det}(\Sigma)}}\exp\left(-\frac{1}{2}\mathbf{x}^{\mathsf{T}}\Sigma^{\dagger}\mathbf{x}\right)$ with mean $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma = \mathbb{E}[(X - \mu)(X - \mu)^{\mathsf{T}}]$. If Σ has a non-trivial kernel space, then the PDF is undefined. However, in this paper, we only need to compare the probability distribution of two random variables which are defined over the same subspace. Therefore, wherever required, we restrict our attention to the (sub)space orthogonal to the kernel space of Σ. Multivariate Gaussian distribution maintains many key properties of univariate Gaussian distribution. For example, any (non-empty) subset of multivariate Gaussians is a multivariate Gaussian and linear functions of multivariate Gaussian random variables are multivariate Gaussian random variables, i.e., if $\mathbf{y} = A\mathbf{x} + \mathbf{b}$, where $A \in \mathbb{R}^{n\times n}$ is a non-singular matrix and $\mathbf{b} \in \mathbb{R}^n$, then $\mathbf{y} \sim \mathcal{N}(A\mu + \mathbf{b}, A\Sigma A^{\mathsf{T}})$.

We use the following properties of Gaussian matrices, a proof could be found in Tao [57] or Muirhead [46].

Lemma 15. *[57].* Let Ω be a Gaussian matrix. Then $\mathbb{E}[\|A\Omega B\|_2] \leq \|A\|_2\|B\|_F + \|A\|_F\|B\|_2$ for fixed matrices A and B.

Lemma 16. *[46].* Let Ω be a $n \times (k + p)$ Gaussian matrix. Then

$$\mathbb{E}[\|\Omega^{-1}\|_F^2] = \sqrt{\frac{k}{p-1}} \qquad \text{and} \qquad \mathbb{E}[\|\Omega^{-1}\|_2] \leq \frac{e\sqrt{k+p}}{p}.$$

Lemma 17. *[46].* Let Ω be a random $n \times (k + p)$ Gaussian matrix whose entries are picked from the distribution $\mathcal{N}(0, 1)$. Then $\mathbb{E}[\mathrm{Tr}((\Omega^{\mathsf{T}}\Omega)^{-1})] = k/(p - 1)$.

On input a streamed vector $\mathbf{v}$, parameters $r, \tilde{n}$, the mechanism samples an $r \times 2\tilde{n}$ random matrix Ω, and

Variant 1: Compute $Y_{\mathbf{v}} = \Omega\mathbf{v}$, and return $Y_{\mathbf{v}}$.

Variant 2: Compute $Y_{\mathbf{v}} = \Omega^{\mathsf{T}}\Omega\mathbf{v}$, and return $Y_{\mathbf{v}}$.

Figure 1: Private Sketch Generation (PSG) Algorithm

Theorem 18. *(Johnson-Lindenstrauss lemma)* Fix any $\eta < 1/2$ and let m be a positive integer. Let Ω be a $k \times n$ matrix with entries picked from a Gaussian distribution $\mathcal{N}(0,1)$, where $k \geq 4(\alpha^2/2 - \alpha^3/3)^{-1} \ln m$. Then for any m unit vector set S in $\mathbb{R}^n$

$$\forall \mathbf{x} \in S, \mathsf{Pr}_M \left[\|\Omega\mathbf{x}\|_2 \in (1 \pm \alpha)\|\mathbf{x}\|_2\right] \geq 1 - 2\exp(-\alpha^2 k/8).$$

DIFFERENTIAL PRIVACY. We use the following in our analysis explicitly or implicitly.

Theorem 19. *(Composition Theorem [20]).* Let $\varepsilon, \delta \in (0,1)$, and $\delta' > 0$. If $\mathcal{K}_1, \cdots, \mathcal{K}_\ell$ are each (ε, δ)-differential private mechanism, then the mechanism $\mathcal{K}(D) := (\mathcal{K}_1(D), \cdots, \mathcal{K}_\ell(D))$ releasing the concatenation of each algorithm is $(\varepsilon', \ell\delta + \delta')$-differentially private for $\varepsilon' < \sqrt{2\ell \ln(1/\delta')}\varepsilon + 2\ell\varepsilon^2$.

Lemma 20. Let $M(D)$ be a (ε, δ)-differential private mechanism for a database D , and let h be any function, then any mechanism $M' := h(M(D))$ is also (ε, δ)-differentially private for the same set of queries.

2 Differentially Private Sketch Generation

We study differential privacy in the well known streaming model of computation [2, 47]. We present it at the level required to understand this paper (a more formal definition appears in Alon *et al.* [2]). This model has three entities: a stream generator or database owner $\mathcal{S}$, a (database) curator $\mathcal{K}$, and a query maker $\mathcal{Q}$. $\mathcal{S}$ starts the process at time $t = 0$ and the curator initializes its data structure to $\mathcal{D}_0$. Thereafter, the curator is allowed only one-pass over the input matrix, i.e., it can access any entry of the data-base during exactly one time epoch, and update its data structure to $\mathcal{D}_t$ using $\mathcal{D}_{t-1}$ and the newly accessed data-points of the matrix. At certain time, t, the query maker $\mathcal{Q}$ makes a query q. The curator responds with $q(\mathcal{D}_t)$.

The streaming model has a resource bound on the curator. A curator is only allowed to use total *polynomial in the size of the data base* time to construct the data structure and the size of data-structure *sub-linear in the size of the data base*. For *differential privacy, we further require that the response of $\mathcal{K}$ to the query of $\mathcal{Q}$ should satisfy Definition 1 with the two neighbouring streams differing in at most one entry of norm* 1. Researchers have also studied differential privacy on streaming data in private learning theory model [37], where the emphasis is on learning about the streamed input–one bears a regret on a hypothesis evaluated against a data point which is not yet streamed. Due to lack of space, we do not delve into the detail comparison of the two models.

We give the generic construction of PSG in Figure 1. If we use random Gaussian matrix as Ω, then the first variant has some resemblance to the mechanism of Blocki *et al.* [5] and Upadhyay [59], while the second variant can be seen as its extension in the sense that two successive application of a Gaussian matrix in a defined form also preserve privacy. However, the analogy ends here. For example, [5, 59] perform an affine transformation to convert the private matrix in to a set of $\{\sqrt{w/n}, 1\}^n$ vectors, while we perform the perturbation to raise the singular values before invoking PSG (see Section 3). As argued by Blocki *et al.* [5], their mechanism does not give a guarantee that the singular values of $A^{\mathsf{T}}A$ and their published matrix is close or their eigenvalues are comparable. In other words, it does not give a LRA. Apart from these major differences, there are couple of subtle differences: (i) they project the entries of the columns of private matrix to a higher dimensional space; here, we perform embedding to a lower dimensional subspace in the similar vein as other applications of the JL-transform, and (ii) their mechanism uses multiple pass over the input matrix in the way it is presented (they require at least two-pass over the input matrix even with the

streaming algorithms for computing the SVD [50, 56] (see [5, Algorithm 3]). Our first observation is that we do not need to subtract the entries of the matrix because of the type of queries we are dealing with. We prove the following for the mechanism in Figure 1 when initiated by random Gaussian matrices.

Theorem 21. If the singular values of the streamed matrix to the first variant of the PSG algorithm are at least $\sigma_{\mathsf{min}} := \left(4\sqrt{r\log(2/\delta)}\log(r/\delta)\right)/\varepsilon$ and for second variant are at least $\sigma_{\mathsf{min}} := (4r\log(r/\delta))/\varepsilon$. Then PSG using $r \times 2n$ Gaussian matrix preserves (ε, δ)-differential privacy.

The proof of the first variant follows the idea of Blocki *et al.* [5] taking into account the subtle differences mentioned above; however, the proof of variants 2 is more involved. We show that, for a streamed matrix A, the probability density function of the published matrix when using the second variant is

$$\frac{\exp(-\mathrm{Tr}((A^{\mathsf{T}}A)^{-1}\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{r(r-1)/4}\Delta(A^{\mathsf{T}}A)^{r/2}\prod_{i=1}^{r}\Gamma((n-i+1)/2)}, \tag{1}$$

where $\Delta(A)$ is the product of the singular values of A and $\Phi = \sum_{i=1}^{r}|\mathbf{a}_i\rangle\langle\mathbf{a}_i|$ for n-variate Gaussians, $\mathbf{a}_1, \cdots, \mathbf{a}_r$. This is the technical part of the proof; rest of the proof follows by evaluating the pdf for neighbouring matrices. The detail proof follows.

Proof. We now prove that the second variant preserves privacy if the singular values of the streamed matrix follows the hypothesis of the theorem. In this section, we use Dirac notation to denote vectors, i.e., $|\cdot\rangle$ for column and $\langle\cdot|$ for row vector. We start by computing the probability density function when the underlying multivariate Gaussian distribution is $\mathcal{N}(0, \mathbb{I})$. The case for arbitrary positive definite covariance matrix follows like the transition from identity to arbitrary positive definite covariance matrices in the multivariate Gaussian distribution. Let $\mathbf{g}_1, \cdots, \mathbf{g}_r$ be i.i.d. $\mathcal{N}(0, \mathbb{I})$ be r multivariate Gaussian distribution, i.e., $\mathbf{g}_{ij} \sim \mathcal{N}(0, 1)$ for $1 \le i \le r, 1 \le j \le n$. The distribution we are interested in is $\Phi = \sum_{i=1}^{r}|\mathbf{g}_i\rangle\langle\mathbf{g}_i|$. We use the notation $\mathsf{PDF}(\Phi; \mathbb{I})$ to denote the probability density function of Φ when each random variable is picked using a normal distribution, i.e., when the covariance matrix of the random variables is $\mathbb{I}$.

Using the chain rule, the joint distribution of the entries of Φ is as follows.

$$\begin{aligned}\mathsf{PDF}(\Phi; \mathbb{I}) = \mathsf{PDF}(\langle\mathbf{g}_1, \mathbf{g}_1\rangle; \mathbb{I})\mathsf{PDF}((\langle\mathbf{g}_2, \mathbf{g}_1\rangle, \langle\mathbf{g}_2, \mathbf{g}_2\rangle)|\langle\mathbf{g}_1, \mathbf{g}_1\rangle; \mathbb{I}) \cdots \\ \mathsf{PDF}((\langle\mathbf{g}_r, \mathbf{g}_1\rangle, \cdots \langle\mathbf{g}_r, \mathbf{g}_r\rangle)|\Phi_{[r-1]}; \mathbb{I}).\end{aligned} \tag{2}$$

There are $r(r+1)/2$ distinct entries, $\langle\mathbf{g}_1, \mathbf{g}_1\rangle, (\langle\mathbf{g}_2, \mathbf{g}_1\rangle, \langle\mathbf{g}_2, \mathbf{g}_2\rangle), \cdots, (\langle\mathbf{g}_r, \mathbf{g}_1\rangle, \cdots \langle\mathbf{g}_r, \mathbf{g}_r\rangle)$. Our aim is to compute each individual term in the product form of the above chain rule. For this, we use the facts we first analyze the distribution of $\mathbf{h}_{i-1}^{\mathsf{T}} = (\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{i-1}\rangle)$. Then we use the fact that there is a transformation of Jacobian one from $\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{(i-1)}\rangle, \langle\mathbf{g}_i, \mathbf{g}_i\rangle - \mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\mathbf{h}_{i-1}\right)$ to $\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{(i)}\rangle\right)$ to compute each term in the chain rule [46, 49].

Distribution of $\mathbf{h}_{i-1}$. We first prove that $\mathbf{h}_{i-1}$ is $(i-1)$ variate Gaussian distribution. Since the covariance matrix is $\mathbb{I}$, and $\mathbf{g}_{11}, \cdots, \mathbf{g}_{1n}, \cdots, \mathbf{g}_{r1}, \cdots, \mathbf{g}_{rn}$ are i.i.d. $\mathcal{N}(0, 1)$, from the elementary property of linear functions of normal variables, conditional on $\mathbf{g}_{kj}$ for $1 \le k \le i-1$ and $1 \le j \le n$, $\mathbf{h}_{i-1}$ is $(i-1)$-variate Gaussian distribution with

$$\Phi_{[i]} = \begin{pmatrix}\langle\mathbf{g}_1, \mathbf{g}_1\rangle & \cdots & \langle\mathbf{g}_1, \mathbf{g}_i\rangle \\ \vdots & \ddots & \vdots \\ \langle\mathbf{g}_i, \mathbf{g}_1\rangle & \cdots & \langle\mathbf{g}_i, \mathbf{g}_i\rangle\end{pmatrix}$$

Now $\mathbf{g}_{11}, \cdots, \mathbf{g}_{rn}$, for every $j = 1, \cdots, n$ are mutually independent; therefore, we have

$$\mathsf{COV}(\mathbf{h}_{i-1}, \mathbf{g}_{ij}) = \left[\mathsf{COV}\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle\mathbf{g}_{ij}\right), \cdots, \mathsf{COV}\left(\langle\mathbf{g}_i, \mathbf{g}_{(i-1)}\rangle\mathbf{g}_{ij}\right)\right] = (\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})^{\mathsf{T}}$$

and $\mathbb{E}[|\mathbf{h}_{i-1}\rangle\langle\mathbf{h}_{i-1}||\mathbf{g}_{kj}] = \Phi_{[i-1]}$ for $1 \leq j < i$. This implies

$$\mathsf{COV}\left[\mathbf{h}_{i-1}, \mathbf{g}_{ij} - \mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\mathbf{g}_1 j, \cdots \mathbf{g}_{i-1,j})^{\mathsf{T}} | \mathbf{g}_{kj}\right] = 0 \qquad \forall 1 \leq k \leq i-1, \tag{3}$$

as the left hand side equals $(\mathbf{g}_1 j, \cdots \mathbf{g}_{i-1,j})^{\mathsf{T}} - \Phi_{[i-1]}\Phi_{[i-1]}^{-1}(\mathbf{g}_1 j, \cdots \mathbf{g}_{i-1,j})^{\mathsf{T}}$.

Therefore, $\mathbf{h}_{i-1}$ is independent of $\sum_{j=1}^{k}\left(\mathbf{g}_{ij} - (\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\mathbf{g}_{1j,\cdots,\mathbf{g}_{i-1,j}})\right)^2$. Rao [49] proved that

$$\sum_{j=1}^{k}\left(\mathbf{g}_{ij} - (\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})\right)^2 \sim \chi^2_{n-i+1}, \tag{4}$$

the standard χ^2 distribution with $(n-i+1)$ degrees of freedom.

Computing every term in the chain rule. From the fact that $\mathbf{h}_{i-1}$ is a $(i-1)$-variate Gaussian distribution, equation (26), equation (4), and the identity

$$\Delta(\Phi_{[i]}) = \Delta(\Phi_{[i-1]})\sum_{j=1}^{k}\left(\mathbf{g}_{ij} - (\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\mathbf{g}_{1j}, \cdots, \mathbf{g}_{i-1,j})\right)^2,$$

where $\Delta(\cdot)$ denotes the determinant, we first calculate the joint pdf of

$$\begin{aligned}
&\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{(i-1)}\rangle, \langle\mathbf{g}_i, \mathbf{g}_i\rangle - \langle\mathbf{h}_{i-1}|\Phi_{[i-1]}^{-1}|\mathbf{h}_{i-1}\rangle\right)^{\mathsf{T}} \\
&\quad= \frac{\exp\left(-\frac{1}{2}\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{(i-1)}\rangle\right)^{\mathsf{T}}\Phi_{[i-1]}^{-1}\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i\mathbf{g}_{(i-1)}\rangle\right)\right)}{(2\pi)^{(i-1)/2}\Delta(\Phi_{[i-1})^{1/2}} \\
&\qquad\times \frac{\exp\left(-\frac{\langle\mathbf{g}_i,\mathbf{g}_i\rangle - \mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\mathbf{h}_{i-1}}{2}\right)\left(\langle\mathbf{g}_i, \mathbf{g}_i\rangle - \mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\mathbf{h}_{i-1}\right)^{(n-i+1)/2-1}}{2^{(n-i+1)/2}\Gamma((n-i+1)/2)} \\
&\quad= \frac{\exp\left(-\frac{\mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\mathbf{h}_{i-1}}{2}\right)\Delta(\Phi_i)^{(n-i-1)/2}}{2^{m/2}\pi^{(i-1)/2}\Gamma((n-i+1)/2)\Delta(\Phi_{[i-1]})^{(n-i)/2}} \\
&\quad= \mathsf{PDF}((\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{i-1}\rangle, \langle\mathbf{g}_i, \mathbf{g}_i\rangle)|\Phi_{[i-1]}),
\end{aligned} \tag{5}$$

where the last step uses the fact that the transformation from $\left(\mathbf{h}_i^{\mathsf{T}}, \langle\mathbf{g}_i, \mathbf{g}_i\rangle - \mathbf{h}_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\mathbf{h}_{i-1}\right)$ to $\left(\langle\mathbf{g}_i, \mathbf{g}_1\rangle, \cdots, \langle\mathbf{g}_i, \mathbf{g}_{(i)}\rangle\right)$ is one-to-one with Jacobian 1.

Computing the joint distribution of Φ. A simple arithmetic followed by plugging equation (5) in equation (2) gives the closed formed expression of the pdf of Φ as

$$\frac{\exp(-\mathrm{Tr}(\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{\sum_i(i-1)/2}\prod_{i=1}^{r}\Gamma((n-i+1)/2)} \times \prod_{i=1}^{r}\left(\frac{\Delta(\Phi_{[i]})^{(n-i-1)/2}}{\Delta(\Phi_{[i-1]})^{(n-i)/2}}\right) = \frac{\exp(-\mathrm{Tr}(\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{n(n-1)/4}\prod_{i=1}^{r}\Gamma((n-i+1)/2)}.$$

Using standard techniques (that could be found in any standard textbook, including Rao [49]) of the transformation method and the factorization theorem yields the PDF for arbitrary linear translation. That is, if $X \sim \mathsf{PDF}(\Phi; \mathbb{I})$, then $Y = AX$ is distributed as $\mathsf{PDF}(\Phi; A^{\mathsf{T}}A)$. The proof is similar to the similar

transformation for multivariate Gaussian distribution. It is easy to verify that it does not matter if we multiply A from right or left of vectors $\mathbf{g}_1, \cdots, \mathbf{g}_r$, i.e., $\sum_{i=1}^r A|\mathbf{g}_i\rangle\langle\mathbf{g}_i|$ and $\sum_{i=1}^r |\mathbf{g}_i\rangle\langle\mathbf{g}_i|A^\mathsf{T}$ have the same distribution. More concretely, for $\sum_{i=1}^r A|\mathbf{g}_i\rangle\langle\mathbf{g}_i|$, the distribution is

$$\mathsf{PDF}(\Phi; A^\mathsf{T}A) = \mathsf{PDF}(\Phi; \mathbb{I})\frac{\mathsf{PDF}(\langle \mathbf{g}_1, \cdots, \mathbf{g}_r\rangle; A^\mathsf{T}A)}{\mathsf{PDF}(\langle \mathbf{g}_1, \cdots, \mathbf{g}_r\rangle; \mathbb{I})} = \frac{\exp(-\mathrm{Tr}((A^\mathsf{T}A)^{-1}\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{r(r-1)/4}\Delta(A^\mathsf{T}A)^{r/2}\prod_{i=1}^r \Gamma((n-i+1)/2)}.$$

We can now prove the privacy guarantee. Let $\delta_0 = \delta/r$. Let A and $\tilde{A}$ be the matrix such that $A - \tilde{A} = E = |v\rangle\langle e_i|$ for an unit vector $|v\rangle$ and some i. The published matrices corresponding to the two neighboring matrices have the following probability density function

$$\begin{aligned}
\mathsf{PDF}(\Phi; A^\mathsf{T}A) &= \frac{\exp(-\mathrm{Tr}((A^\mathsf{T}A)^{-1}\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{n(n-1)/4}\Delta(A^\mathsf{T}A)^{r/2}\prod_{i=1}^r \Gamma((n-i+1)/2)} = C\frac{\exp(-\mathrm{Tr}((A^\mathsf{T}A)^{-1}\Phi)/2)}{\Delta(A^\mathsf{T}A)^{r/2}},\\
\mathsf{PDF}(\Phi; \tilde{A}^\mathsf{T}\tilde{A}) &= \frac{\exp(-\mathrm{Tr}((\tilde{A}^\mathsf{T}\tilde{A})^{-1}\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{n(n-1)/4}\Delta(\tilde{A}^\mathsf{T}\tilde{A})^{r/2}\prod_{i=1}^r \Gamma((n-i+1)/2)} = C\frac{\exp(-\mathrm{Tr}((\tilde{A}^\mathsf{T}\tilde{A})^{-1}\Phi)/2)}{\Delta(\tilde{A}^\mathsf{T}\tilde{A})^{r/2}},
\end{aligned}$$

where $C = \Delta(\Phi)^{(n-r-1)/2}/\left(2^{rn/2}\pi^{n(n-1)/4}\prod_{i=1}^r \Gamma((n-i+1)/2)\right)$. As in Blocki *et al.* [5], it is straightforward to see that combination of the following proves differential privacy of the published matrix:

$$\exp(-\varepsilon/r) \leq \sqrt{\frac{\tilde{\Delta}(A^\mathsf{T}A)}{\tilde{\Delta}(\bar{A}^\mathsf{T}\bar{A})}} \leq \exp(\varepsilon/r) \quad \text{and} \quad \mathsf{Pr}\left[|\mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right)| \leq \varepsilon\right] \geq 1-\delta. \tag{6}$$

Let $\sigma_1 \geq \cdots, \geq \sigma_d \geq \sigma_{\mathsf{min}}$ be the singular values of A. Let $\lambda_1, \geq \cdots, \geq \lambda_d \geq \sigma_{\mathsf{min}}$ be the singular value for $\bar{A}$. Since the singular values of $A - \bar{A}$ and $\bar{A} - A$ are the same, $\sum_{i\in G}(\sigma_i - \lambda_i) \leq 1$ using Linskii's theorem, where G is the set of indices for which $\sigma_i > \lambda_i$. The first bound follows similarly as in Blocki *et al.* [5]. For the second bound required for the privacy, we first bound the following

$$\begin{aligned}
\mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right) &= \mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1}(\tilde{A}^\mathsf{T}\tilde{A})(\tilde{A}^\mathsf{T}\tilde{A})^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right)\\
&= \mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1}(A+E)^\mathsf{T}(A+E)(\tilde{A}^\mathsf{T}\tilde{A})^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right)\\
&= \mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1}(A^\mathsf{T}E + E^\mathsf{T}\tilde{A})(\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right).
\end{aligned}$$

Using the singular value decomposition of $A = U\Sigma V^\mathsf{T}$ and $\tilde{A} = \tilde{U}\Lambda\tilde{V}^\mathsf{T}$, and the fact that $E = |v\rangle\langle e_i|$ for some i, we can further solve the above expression.

$$\begin{aligned}
\mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right) &= \mathrm{Tr}\left(\left(V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T} + V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}\right)\Phi\right)\\
&= \mathrm{Tr}\left(V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}\Phi\right) + \mathrm{Tr}\left(V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}\Phi\right)\\
&= \sum_{j=1}^r \mathrm{Tr}\left(\langle\mathbf{g}_j|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle\right)\\
&\quad + \sum_{j=1}^r \mathrm{Tr}\left(\langle\mathbf{g}_j|V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle\right).
\end{aligned}$$

Fix a j. We bound the following.

$$|\mathrm{Tr}\left(\langle\mathbf{g}_j|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle\right) + \mathrm{Tr}\left(\langle\mathbf{g}_j|V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle\right)|. \tag{7}$$

We now look at each term in the above expression. $\langle \mathbf{g}_j | V\Sigma^{-2}V^\mathsf{T} | v \rangle$ is distributed as $\mathcal{N}(0, \|V\Sigma^{-2}V^\mathsf{T}|v\rangle\|^2)$, $\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle$ as $\mathcal{N}(0, \|\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}\|^2)$, $\langle \mathbf{g}_j|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle$ as $\mathcal{N}(0, \|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\|^2)$, and $\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\mathbf{g}_j\rangle$ as $\mathcal{N}(0, \|\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}\|^2)$. Since v and e_i are unit vectors, the norm of the above four quantities are less than $1/\sigma_{\mathsf{min}}^2$, $1/\sigma_{\mathsf{min}}^2 + 1/\sigma_{\mathsf{min}}^3$, $1/\sigma_{\mathsf{min}}$, and $1/\sigma_{\mathsf{min}} + 1/\sigma_{\mathsf{min}}^2$, respectively.

Therefore, from the concentration inequality of Gaussian distribution, we have

$$\mathsf{Pr}\left[(7) \leq 2\left(\frac{1}{\sigma_{\mathsf{min}}^2} + \frac{1}{\sigma_{\mathsf{min}}^3}\right)\ln(4/\delta_0) \leq \varepsilon\right] \geq 1 - \delta_0.$$

Taking union bound, we have with probability $1-\delta$, $-\varepsilon \leq \mathrm{Tr}\left(\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right) \leq \varepsilon$. □

3 Applications of Private Sketch Generation

Our mechanisms assume that the matrix A is provided as the symmetric $\mathsf{rank}(A)$ matrix A' corresponding to A (note the update steps in Figure 2 and 3), and stop updating the data structure once all the columns of A are streamed. This simplifies the presentation as well as the analysis. By the argument of Hardt and Roth [33, Fact 2.8], this leads to a depreciation of the privacy guarantee by half (both ε and δ). The utility proof does not change for MAT-MULT and LIN-REG if we use A or A'. However, the analysis for utility gets more complicated in the case of LRA because the right and the left singular vectors of the original matrix might be different. Keeping this in mind, we present its analysis in the most general form by working with its SVD.

We need more care to design the mechanism for MAT-MULT and LIN-REG. For example, consider MAT-MULT. If we just feed the streamed vector (after making an affine transformation) in PSG, we end up getting an additive error proportional to the Frobenius norm of input matrices. On the other hand, after the transformation to A' and B' of any conforming matrices A and B, respectively, we can use the identity, $\begin{pmatrix}\mathbb{I} & A'\end{pmatrix}\begin{pmatrix}\mathbb{I} & B'\end{pmatrix}^\mathsf{T} = \left(\mathbb{I} + A'B'^T\right)$, to perturb the input matrix with a careful choice of parameters. Intuitively, τ is due to the identity term of the published matrix. The utility bound follows from either the application of Theorem 18 or one of its corollary. We use the same idea for LIN-REG as well. To prove the privacy of MAT-MULT and LIN-REG, we use the first variant, while for LRA, we need both the variants defined in Figure 1.

We note that this is not the only method that could be used for the first two problems. For example, one could compute the SVD using one of the standard streaming algorithms [50, 56]–by say, setting the target rank in the algorithm of [50] to be $\mathsf{rank}(A)$, and then concatenate $\mathbf{e}_t$ to the matrix. The privacy proof remains the same, but this makes the mechanisms more complicated. Moreover, in disguise, this is also an affine transformation, so leads to an added utility loss. We do not explore this idea any further.

3.1 Matrix Multiplication

Our mechanism for MAT-MULT (Figure 2) lifts the singular values of the input matrices above the threshold of the first variant of Theorem 21 and Theorem 27, while not introducing error independent of Frobenius norm.

Theorem 22. Let Ω be the random matrix used by PSG. Then, the data-structure generated by mechanism stated in Figure 2 uses $O(d\alpha^{-2}\kappa\log(1/\beta))$ bits of space, and on input conforming matrices A and B, computes (α, β, τ)-MAT-MULT with $\tau \leq \sqrt{n}\alpha$ additive error and (ε, δ)-differential privacy.

The space requirement of the data-structure is straightforward by the choice of r. The privacy guarantee follows because $\bar{A}\bar{A}^\mathsf{T} \succeq U\,(16r\ln(2/\delta)/\varepsilon)\ln(16r/\delta)^2\mathbb{I}U^\mathsf{T} = \sigma_{\mathsf{min}}^2\mathbb{I}$, i.e., the singular values of the perturbed matrices are above the threshold of Theorem 21. The proof of utility readily follows using the

Initialization. On input parameters $\alpha, \beta, \varepsilon, \delta$, set $r = O(\log(1/\beta)/\alpha^2)$. Set $s = \sqrt{16r\ln(2/\delta)}\varepsilon^{-1}\ln(16r/\delta)$. Set the intial sketches of A and B to be all zero matrices Y_{A_0} and Y_{B_0}.

Data-structure Update. On input a column a of an $n \times d_1$ matrix A and column b of an $n \times d_2$ matrix B at time epoch t, set $d = \max\{d_1, d_2\}$, $\bar{A}_{:a} = s\begin{pmatrix}\mathbf{e}_a & \mathbf{0}^{n+d} & A_{:a}\end{pmatrix}$ and $\bar{B}_{:b} = s\begin{pmatrix}\mathbf{e}_b & \mathbf{0}^{n+d} & B_{:b}\end{pmatrix}$. Invoke variant 1 of PSG with inputs $(\bar{A}_{:a}, r, n+d)$ and $(\bar{B}_{:b}, r, n+d)$. Update the sketches by replacing the columns a of $Y_{A_{t-1}}$ and b of $Y_{B_{t-1}}$ by the respective returned sketches to get the sketch Y_{A_t}, Y_{B_t}.

Answering Matrix Product: On request to compute the matrix product at time t, compute the product $Y_{A_t}^{\mathsf{T}} Y_{B_t}/s^2$.

Figure 2: The Mechanism for (α, β, τ)-MAT-MULT

variance bound on $\|A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B - AB\|_F^2$ by Kane and Neilson [39], bounding $\|\Omega^{\mathsf{T}}\Omega - \mathbb{I}\|_2$ that results from taking in to account the perturbation made to the input streams in Step 1, and norm inequalities. The details follow.

Proof. The proof of the utility of mechanism in Figure 2 follows readily from Lemma 23, which is the bound computed by Kane-Neilson [39].

Lemma 23. Let Ω be a $k \times d$ matrix as constructed in Section 2 with every entries picked from the distribution $\mathcal{N}(0,1)$, then for a set of m vectors, $\mathbf{v}_1, \cdots, \mathbf{v}_m \in \mathbb{R}^n$, with probability at least $1 - 2\exp(-k\alpha^2/8)$, for any pair $\mathbf{v}_i, \mathbf{v}_j$, we have $|\langle \Omega\mathbf{v}_i, \Omega\mathbf{v}_j\rangle - \langle \mathbf{v}_i, \mathbf{v}_j\rangle| \leq \alpha\|\mathbf{v}_i\| \cdot \|\mathbf{v}_j\|$.

We need to upper bound the quantity $\|A^{\mathsf{T}}B - \bar{A}^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega\bar{B}/s^2\|_F$. First note that

$$\bar{A}^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega\bar{B} = s^2\begin{pmatrix}\mathbb{I} & 0 & 0 & A^{\mathsf{T}}\end{pmatrix}\Omega^{\mathsf{T}}\Omega\begin{pmatrix}\mathbb{I} & 0 & 0 & B^{\mathsf{T}}\end{pmatrix}^{\mathsf{T}} = s^2(\mathbb{I}\Omega^{\mathsf{T}}\Omega\mathbb{I} + A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B).$$

Therefore,

$$\|A^{\mathsf{T}}B - \bar{A}^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega\bar{B}/s^2\|_F = \|A^{\mathsf{T}}B - A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B - \mathbb{I}\Omega^{\mathsf{T}}\Omega\mathbb{I}\|_F \leq \|A^{\mathsf{T}}B - A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B\|_F + \|\mathbb{I} - \Omega^{\mathsf{T}}\Omega\|_F. \quad (8)$$

To bound the first term, let random variable X_{ij} denote $(A^{\mathsf{T}}B)_{ij} - (A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B)_{ij}$. Then, with probability at least $1 - 2\exp(-k\alpha^2/8)$, we have $|X_{ij}| \leq \alpha\|A_{:i}\|_2 \cdot \|B_{:j}\|_2$. Using Lemma 23, this results in

$$\|A^{\mathsf{T}}B - A^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega B\|_F^{\,2} = \sum |X_{ij}|^2 \leq \sum \alpha^2\|A_{i:}\|_2^2\|B_{:j}\|_2^2 \leq \alpha^2\|A\|_F^2\|B\|_F^2. \quad (9)$$

For the second term, we need to bound the variance on unitaries. This follows from the following set of inequalities.

$$\begin{aligned}\|U_1^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega U_2 - U_1U_2\|_2 &= \|U_1(\Omega^{\mathsf{T}}\Omega - \mathbb{I})U_2\|_2 = \|\Omega^{\mathsf{T}}\Omega - \mathbb{I}\|_2 \\ &= \left(\max\frac{x^{\mathsf{T}}(\Omega^{\mathsf{T}}\Omega - \mathbb{I})x}{\langle x, x\rangle}\right) \leq ((1+\alpha) - 1) = \alpha, \end{aligned} \quad (10)$$

where the inequality follows from Theorem 18 and noting that r still satisfies the theorem requirement. The result follows by adjusting the value of α after plugging this and equation (28) in equation (8), using the fact that $\|X\|_2 \leq \|X\|_F \leq \sqrt{n}\|X\|_2$ for any $n \times n$ matrix X.

□

3.2 Linear Regression

Our mechanism for LIN-REG (Figure 3) lifts the singular values of the input matrices above the threshold of the first variant of Theorem 21 and Theorem 27, while not introducing error independent of Frobenius norm.

Initialization. On input parameters $\alpha, \beta, \varepsilon, \delta$, set $r = O(d\log(1/\beta)/\alpha)$, $s = \sqrt{16r\ln(2/\delta)}\varepsilon^{-1}\ln(16r/\delta)$, and Y_{A_0} to be all zero matrix.

Data-structure Update. On input a column c of an $n \times d$ matrix A at time epoch t, set $\bar{A}_{:c} = s\begin{pmatrix}\mathbf{e}_c & \mathbf{0}^{n+d} & A_{:c}\end{pmatrix}$, and invoke variant 1 of PSG with inputs $(\bar{A}_{:c}, r, n+d)$. Update the sketch of A by replacing the column c of $Y_{A_{t-1}}$ by the returned sketches to get the sketch Y_{A_t}.

Answering Queries. On being queried with a vector b_i, set $\bar{\mathbf{b}}_i = \begin{pmatrix}\mathbf{e}_i & \mathbf{0}^{n+d} & \mathbf{b}_i\end{pmatrix}$, and invoke PSG with input matrix $(\bar{\mathbf{b}}_i, r, n+d)$ to get the sketch $Y_{\mathbf{b}_i}$. Compute a vector $\mathbf{x}_i$ satisfying $\min_{\mathbf{x}} \|Y_{A_t}\mathbf{x}_i - Y_{\mathbf{b}_i}\|$.

Figure 3: The Mechanism for (α, β, τ)-LIN-REG

Theorem 24. Let Ω be $r \times 2(n+d)$ matrix used by PSG, where $r = O(d\log(1/\beta)/\alpha)$. Then the data-structure generated in the mechanism of Figure 3 requires $O(d^2\alpha^{-1}\kappa\log(1/\beta))$ bits and allows to solve (α, β, τ)-LIN-REG problem in an (ε, δ)-differentially private manner with $\tau \leq O(\sqrt{n}\alpha)$.

The utility bound for LIN-REG follows from taking in to account the changes made to the input matrices by our mechanism and a straightforward application of the result of Kane and Nelson [39] to the proof idea of Sarlos [54]. Sarlos [54] observed that the utility guarantee for LIN-REG is provided as long as Ω provides MAT-MULT with multiplicative error $\sqrt{\alpha/d}$, and Ω is an $O(d)$-space embedding. The latter follows from the standard result on any Johnson-Lindenstrauss transform with $r = O(\mathsf{rank}(A) + \log(1/\beta)/\alpha^2)$. Using Indyk and Motwani [36], we only need $r = O(d^2\alpha^{-1}\log(1/\beta))$ as in Theorem 24. The details follow.

Proof. The formal proof of Theorem 24 is identical to Clarkson and Woodruff [12] modulo the analysis to consider the lift of singular value and using the bound of Kane and Neilson [39]. We present it next. Let B be the matrix formed by the set of queries $\{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$. Since the columns of U is a set of orthonormal vectors, we have $UU^\mathsf{T}U = U$ and $\|U^\mathsf{T}UC\|_F = \|UC\|_F$ for any matrix C. Therefore, it suffices for the utility bound to prove a bound on $\|U^\mathsf{T}\bar{A}(\bar{X} - X^*)\|$. For this, we first prove that $U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*)$ has a small norm. We have

$$\begin{aligned} U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*) &= U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*) + U^\mathsf{T}\Omega^\mathsf{T}\Omega(B - \bar{A}\bar{X}) \\ &= U^\mathsf{T}\Omega^\mathsf{T}\Omega(B - \bar{A}X^*). \end{aligned}$$

This is because $U^\mathsf{T}\Omega^\mathsf{T}\Omega(\bar{A}\bar{X} - B) = \bar{A}^\mathsf{T}\Omega^\mathsf{T}\Omega(\bar{A}\bar{X} - B) = 0$. Therefore, from Theorem 22 with $\alpha' = \sqrt{\alpha/d}$ (since we chose r in Theorem 22 which differs by a factor of α and $1/d$ with respect to that in Theorem 24), we have

$$\begin{aligned} \|U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*)\|_F &= \|U^\mathsf{T}\Omega^\mathsf{T}\Omega(B - \bar{A}X^*)\|_F \\ &\leq \sqrt{\alpha}\|B - AX^*\|_F + \sqrt{\tau} \end{aligned}$$

From the sub-additivity of the norm and property of conforming matrices, we have

$$\begin{aligned} \|U^\mathsf{T}A(\bar{X} - X^*)\|_F &\leq \|U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*)\|_F + \|U^\mathsf{T}\Omega^\mathsf{T}\Omega\bar{A}(\bar{X} - X^*) - U^\mathsf{T}\bar{A}(\bar{X} - X^*)\|_F \\ &\leq \sqrt{\alpha}\|B - AX^*\|_F + \sqrt{\tau} + \|U^\mathsf{T}\Omega^\mathsf{T}\Omega U - \mathbb{I}\|_2 \cdot \|U^\mathsf{T}\bar{A}(\bar{X} - X^*)\|_F \end{aligned}$$

To move forward, we need the following lemma, an equivalent for equation (10) with this value of r.

Lemma 25. *(Kane and Neilson [39])* Given $r = O(d\log(1/\beta)/\alpha)$. Let U be any unitary matrix. If Ω satisfies the Johnson-Lindenstrauss bound, then with probability at least $1 - \beta$, we have $\|U\Omega^\mathsf{T}\Omega U^\mathsf{T} - \mathbb{I}\|_2 \leq \alpha$.

Now rearranging the terms, we get $\|U^\mathsf{T}\bar{A}(\bar{X} - X^*)\|_F \leq 2\sqrt{\alpha}\|B - AX^*\|_F + \sqrt{\tau}$. The utility proof is now immediate by observing that the column-space of A and U are the same, and the Pythagorus theorem

On input parameters $\alpha, \beta, \varepsilon, \delta$, the target rank k, set $w = \left(ck\varepsilon^{-1}\ln(2/\delta)\right)$ for a global constant c. Pick a $2n \times k$ standard Gaussian matrix Ω. On input an $n \times n$ matrix A of rank r, the mechanism does the following:

Range Finding. Compute $Y_A = \begin{pmatrix} w\mathbb{I} & A \end{pmatrix} \Omega$ by computing $\begin{pmatrix} w\mathbf{e}_i & A_{i:} \end{pmatrix} \Omega$ for all streams $A_{i:}, 1 \leq i \leq n$ and appending them row-wise. Let $\Pi_{Y_A} = \Psi\Psi^{\mathsf{T}}$ be the projection matrix corresponding to the range of Y_A.

Projection: When the whole matrix is streamed, the curator does the following:

1. Let the (unknown) matrix $B = \Psi^{\mathsf{T}} A_t \Psi$.
2. Use the minimal residual method to find a solution to $B\Psi_t^{\mathsf{T}}\Omega = \Psi_t^{\mathsf{T}} Y_t$.
3. Compute the decomposition of $B_t = \bar{U}_t \Lambda_t \bar{U}_t^{\mathsf{T}}$, form the product $\hat{U}_t = \Psi_t \bar{U}_t$, and publish $\hat{U}_t \Lambda_t \hat{U}_t^{\mathsf{T}}$.

Figure 4: The Mechanism for k-rank Approximation

on the norms. More concretely, with probability at least $1 - 2\beta$

$$\begin{aligned} \|A\bar{X} - B\|_F^2 &= \|AX^* - B\|_F^2 + \|A(\bar{X} - X^*)\|_F^2 \\ &\leq (1 + 4\alpha)\|AX^* + B\|_F + \tau. \end{aligned}$$

Adjusting and renaming the values of α and β, we get the claim of the theorem. □

3.3 Low Rank Approximation

We use the prototype mentioned in [30], which was also used by Hardt and Roth [33] to improve the worst-case lower bound under a *low coherence* assumption. The main prototype in [30] constructs a low-dimensional subspace that captures the action of the matrix (*range-finding*), and then restrict the matrix to that subspace to compute the required factorization (*projection*). More concretely, range finding finds a measurement matrix $Y = A\Omega$, where Ω is a Gaussian matrix and computes the orthonormal projection matrix Π_Y corresponding to the range defined by Y; projection then computes a rank k matrix $B = \Pi_Y A$. From this exposition, it seems that privacy preserving algorithms are required for both the stages; however, we show that the two-step prototype mentioned in [30] can be replaced by a two-step algorithm in which the input matrix is explicitly needed only in the first step at the expense of privacy proof requiring both the variants of PSG. We first note that if $\Psi\Psi^{\mathsf{T}}A$ is a LRA of A, i.e., $\|A - \Psi\Psi^T A\| \leq \eta$, then so is $\Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}$. This is because $\|A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| = \|A - \Psi\Psi^{\mathsf{T}}A + \Psi\Psi^{\mathsf{T}}A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| \leq \|A - \Psi\Psi^{\mathsf{T}}A\| + \|\Psi\Psi^{\mathsf{T}}A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| \leq 2\eta$. The crucial observation now for the single pass LRA when Ω is a Gaussian matrix is that Ω, Y, and the basis for the range of Y contains enough information to compute the matrix B. The range-finding is private as in Theorem 22 and 24, but as we are reuse Ω, we have to rely on privacy of the second variant of PSG algorithm to prove the privacy of projection step. In order to simplify the presentation, we state our mechanism for symmetric matrices in Figure 4 ([34, 40] also made this assumption) to compute a LRA of the SVD of A.

Theorem 26. Let $\lambda_1 \geq \cdots \geq \lambda_{\mathsf{rk}(A)}$ be the singular values of A. Then for an over-sampling parameter p, there is a single-pass mechanism that compute LRA using $O(k(n+d)\alpha^{-1}\log(nd))$ bits while preserving (ε, δ)-differential privacy such that

$$\|A - \bar{A}\|_F \leq \left(1 + \frac{k}{p-1}\right)^{1/2} \min_{rk(A')<k} \|A - A'\|_F + \frac{2k}{\varepsilon}\sqrt{\frac{(n+d)\ln(2/\delta)}{p}}, \quad \text{and} \tag{11}$$

$$\|A - \bar{A}\|_2 \leq \left(1 + \frac{k}{p-1}\right)^{1/2} \lambda_{k+1} + \frac{e\sqrt{(k+p)\sum_{j>k}\lambda_j^2}}{p} + \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}. \tag{12}$$

We need both the variants of PSG algorithm for the privacy proof–the first variant during the range finding stage and the second variant because we reuse Ω in the second step of the projection stage. Note that

the first step pushes the singular values above the threshold of Theorem 21. Since rest of the computations are deterministic, privacy follows from Lemma 20, Theorems 21 and 19, and our choice of w. The space guarantee is also straightforward. Now, Π_Y has a decomposition $\Psi\Psi^\mathsf{T}$ for some orthonormal basis Ψ, which also forms an orthonormal basis for the approximated matrix $\bar{A}$ after the run of the algorithm. Therefore, B must satisfy $B\Psi^\mathsf{T}\Omega \approx \Psi^\mathsf{T}Y$. This is what step 2 does. Therefore, $\bar{A}_t = \Pi_Y A_t$ for the projection operator Π_Y with the same range as Ψ. The rest of the utility proof relies heavily on the fact that $\mathrm{Tr}(\Omega\Omega^\mathsf{T})^{-1}$ is bounded, left singular vectors of A does not play any essential role in the concentration bound, and the rotational invariance of Gaussian vectors. We bound the norm of $\|(\mathbb{I} - \Pi_Y)A\|$ for both the spectral and Frobenius norm. Once we have the bound on $\|(\mathbb{I}-\Pi_Y)A\|$, we use standard results in random matrix theory to get the final bounds.

The proof is very different from the proof of Sarlos [54]'s two-pass algorithm and Clarkson-Woodruff [12]'s one-pass algorithm, wherein the authors use the error bound computed in the estimate of MAT-MULT. Here we use perturbation theory, more aligned with [21, 30]. The details proof follow.

Proof. The space complexity is easy to follow as we need to store the matrix Ω and the sketch. The privacy guarantee follows from Theorem 21 and noting that all the singular-values of the matrix on which Ω is operated from the right is greater than the threshold required for the statement of the Theorem 21, and the distribution of $\Psi^\mathsf{T}\Omega$ is the same as that of the second variant as we reuse Ω [9]. Now, it follows from the proof of the second variant (Theorem 21) that it does not matter if we multiply A (or A^T) from left (or right, respectively) of vectors $\mathbf{g}_1, \cdots, \mathbf{g}_r$, i.e., $\sum_{i=1}^r A|\mathbf{g}_i\rangle\langle\mathbf{g}_i| = A\Phi$ and $\sum_{i=1}^r |\mathbf{g}_i\rangle\langle\mathbf{g}_i|A^\mathsf{T} = \Phi A^\mathsf{T}$ have the same distribution. Combining all these arguments, we have the distribution of the second step of projection stage is identical to the second variant in Figure 1, modulo some deterministic computation. Since, any arbitrary post-processing preserves differential privacy, we can now complete the proof by invoking Theorem 21. The privacy guarantee due to Theorem 21 requires the minimum singular value to be greater than $\frac{4\sqrt{k\log(2/\delta)\log(k/\delta)}}{\varepsilon}$ for the first variant and $\frac{k\log(k/\delta)}{\varepsilon}$ for the second variant. By our choice of w, the singular values of the streamed matrix to the algorithm for PSG are at least the eigenvalues of $\sqrt{w^2\mathbb{I} + A^\mathsf{T}A}$, which are all greater than $\frac{16k\ln(2/\delta)}{\varepsilon}$. Since $\frac{4\sqrt{k\log(2/\delta)\log(k/\delta)}}{\varepsilon} \ll 16\frac{k\log(k/\delta)}{\varepsilon}$, the privacy guarantee follows from Theorem 21.

In more detail, it follows from Bura and Pfeiffer [9][1] and the proof in Theorem 21 that, for $\Phi = \sum_{i=1}^r |\mathbf{g}_i\rangle\langle\mathbf{g}_i|$ and large enough n, the distribution of $\Psi^\mathsf{T}\Omega$ is

$$\frac{\exp(-\mathrm{Tr}((A^\mathsf{T}A)^{-1}(D^{-1}R)(D^{-1}R)^\mathsf{T}\Phi)/2)\Delta(\Phi(D^{-1}R)(D^{-1}R)^\mathsf{T})^{(n-r-1)/2}}{2^{rn/2}\pi^{n(n-1)/4}\Delta(A^\mathsf{T}A)^{r/2}\prod_{i=1}^r\Gamma((n-i+1)/2)}.$$

The proof for the first expression of equation (6) is as before. Following the rest of the steps of Theorem 21, using Lemma 8, we compute for fixed $j \in [k+p]$, instead of equation (7), the second expression in equation (6) is bounded by the following expression

$$\left|\mathrm{Tr}\left(V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\alpha_j\rangle\langle\alpha_j| + V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\alpha_j\rangle\langle\alpha_j|\right)\mathrm{Tr}\left((D^{-1}R)(D^{-1}R)^\mathsf{T}\right)\right|.$$

Using the fact that $\mathrm{Tr}((D^\mathsf{T}D)^{-1}) = \mathrm{Tr}((\Omega^\mathsf{T}\Omega)^{-1})$, Lemma 17 with $p = k+1$, and following the remaining steps of Theorem 21, we have the privacy result for computing B. The complete proof follows from the discussion stated at the start of this proof.

Remark 1. We take the liberty to diverge a little to understand the intuition behind Bura and Pfeiffer [9]. One may skip this part without effect the readability of the rest of the section. The result of Bura and Pfeiffer [9]

[1]We use the following result of Bura and Pfeiffer [9]: for a random $n \times k$ normal matrix Ω, for large enough n, the vector formed by the entries of its left singular matrix is distributed normally with covariance matrix $(D^{-1}R \otimes \mathbb{I})(RD^{-1} \otimes \mathbb{I})^\mathsf{T}$, where $\Omega = LDR^\mathsf{T}$.

uses advance statistical tools, but the intuition can be argued using some basic statistics. On the other hand, it is well known that the singular values of random matrices are notoriously hard to compute (see Rudelson and Vershynin [52]). The basic reasoning behind their proof is the following line of argument. Since the entries of an $n \times r$ matrix Ω is $\mathcal{N}(0,1)$, then for any orthogonal matrices $G \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{r\times r}$, the entries of $G\Omega R^{\mathsf{T}}$ is also i.d.d. normal. This can be seen by translating to the vector form of the matrix, i.e., $\mathbf{v} = \mathsf{vec}(V)$ represent rn vector with entries $\mathbf{v}_{i+nj} = V_{ij}$. Then $\mathsf{vec}(G\Omega R^{\mathsf{T}}) = (G \otimes R)\mathsf{vec}(\Omega)$ using Lemma 9. Now $G \otimes R$ is also an orthogonal matrix, and multivariate Gaussian distribution is preserved if one multiply by an orthogonal matrix. Therefore, the distribution of the left singular vectors of Ω is the same as $G\Omega R^{\mathsf{T}}$. Consequently, for large enough n, the distribution of each singular vector is also spherically distributed. □

UTILITY GUARANTEE. In Section 3.3, we showed that the mechanism for symmetric matrices does what [30, Section 1.2] prototype algorithm achieves. To construct a mechanism for non-symmetric matrices, we construct two sketches Y_t and $\hat{Y}_t$ corresponding to A_1 and A_2 using a single pass over A and using two Gaussian matrices Ω and $\hat{\Omega}$ of appropriate dimension, where $A_1 = \begin{pmatrix} w\mathbb{I} & A \end{pmatrix}$ and $A_2 = \begin{pmatrix} A^{\mathsf{T}} & w\mathbb{I} \end{pmatrix}$ for appropriate dimension identity matrices in both A_1 and A_2. Basically, we do the following using a single pass over the matrix A:

$$\begin{pmatrix} Y \\ \hat{Y} \end{pmatrix} = \begin{pmatrix} w\mathbb{I}_n & A \\ A^{\mathsf{T}} & w\mathbb{I}_d \end{pmatrix} \begin{pmatrix} \Omega \\ \hat{\Omega} \end{pmatrix},$$

where $\mathbb{I}_n$ is an $n \times n$ identity matrix.

Since $\begin{pmatrix} Y & \hat{Y} \end{pmatrix}^{\mathsf{T}}$ corresponds to a symmetric matrix, we can use the steps used in the projection stage in Figure 4. Note that Clarkson and Woodruff [12] compute $A^{\mathsf{T}}A$ in a single-pass over the matrix A.

Keeping the most general case in mind, we first show that the left singular vectors have hardly any role to play in bounding the perturbation. We assume that we perform SVD. Let the SVD of A be $U\Sigma V^{\mathsf{T}}$. In the following discussion, we compute the approximation of $\begin{pmatrix} w\mathbb{I} & A \end{pmatrix}$ and denote it by A. This is because $\begin{pmatrix} w\mathbb{I} & A \end{pmatrix}$ is more manageable and any upper bound on the approximation of this matrix is an upper bound on the approximation of the original matrix. The actual bound as computed in Figure 4 can be obtained by simply performing the computation on the singular value decomposition as performed in the last step of mechanism and using the sub-additivity of norms. From the discussion above, we know that $\bar{A} = \Pi_Y A$; therefore, we need to bound $\|(\mathbb{I} - \Pi_Y)A\|$, where, unless specified, in this section $\|\cdot\|$ refers to both the Frobenius as well as the spectral norm. From the Hölder's inequality on the second moment, we have

$$\mathbb{E}[\|(\mathbb{I} - \Pi_Y)A\|_F] \leq \left(\mathbb{E}\left[\|(\mathbb{I} - \Pi_Y)A\|_F^2\right]\right)^{1/2}. \tag{13}$$

We now bound $\| (\mathbb{I} - \Pi_Y)A \|$. Let $\Lambda = \sqrt{\Sigma^2 + w^2}$. Let Λ_1 denote the diagonal matrix formed by the first k singular values and Λ_2 be the diagonal matrix for the rest of the singular values. We decompose V^{T} similarly. Let the matrix formed by the first k rows of V^{T} be V_1^{T} and by the rest of the rows be V_2^{T}.

Left singular vectors have essentially no role in the approximation bound. Recall that Ω is an $2n \times k$ matrix; therefore, $Y = A\Omega = U\begin{pmatrix} \Lambda_1 V_1^{\mathsf{T}}\Omega & \Lambda_2 V_2^{\mathsf{T}}\Omega \end{pmatrix}^{\mathsf{T}}$. It would be useful to consider the first k rows of Y to be the one that mimics the action of A and the rest of the rows of Y as a small perturbation that we wish to bound. We first prove that the left singular vectors have essentially no role to play in bounding the error. Let $A' = UA$, then the following chain of equalities are straightforward.

$$\|(\mathbb{I} - \Pi_Y)A\| = \|U^{\mathsf{T}}(\mathbb{I} - \Pi_Y)A\| = \|U^{\mathsf{T}}(\mathbb{I} - \Pi_Y)UA'\| = \|\mathbb{I}A' - U^{\mathsf{T}}\Pi_Y UA'\|. \tag{14}$$

Now note that the projection matrix corresponding to a matrix Y is uniquely defined by $\mathsf{range}(Y)$, the range of Y. Therefore, $\mathsf{range}(U^{\mathsf{T}}\Pi_Y U) = U^{\mathsf{T}}\mathsf{range}(\Pi_Y) = \mathsf{range}(U^{\mathsf{T}}\Pi_Y)$.

Therefore, $\|A' - U^{\mathsf{T}}\Pi_Y U A'\| = \|(\mathbb{I} - \Pi_{U^{\intercal}Y})A'\|$. A useful way to understand the above expression is to view this geometrically and recall that unitary are just rotation in the space: projection by an unitary, followed by any projection operator, followed by the inverse of unitary brings us to the same space as projection by an operator followed by the inverse of the unitary.

Finding and bounding an appropriate perturbed matrix. We now use the identity that, for two operators O_1 and O_2, if the range of O_1 is a subset of the range of O_2, then the projection of any matrix using O_1 will have all its norm smaller than the projection by O_2. More concretely, we find a matrix C such that its range is a strict subset of the range of $U^{\mathsf{T}}Y$. We obtain this matrix by flattening out the first k rows of $U^{\mathsf{T}}Y$. This is in correspondence with our earlier observation that the first k rows mimic the action of A and other rows are the perturbation that we wish to bound. Since the first k rows of $U^{\mathsf{T}}Y$ is $\Lambda_1 V_1^{\mathsf{T}}\Omega$, let $C := U^{\mathsf{T}}Y\Omega^{-1}V_1\Lambda_1^{-1}$. The rows corresponding to the perturbation are $\Lambda_2 V_2\Omega$. Thus, $C = \begin{pmatrix}\mathbb{I} & \Lambda_2 V_2^{\mathsf{T}} V_1 \Lambda_1^{-1}\end{pmatrix}^{\mathsf{T}}$.

Let us denote by $S = \Lambda_2 V_2^{\mathsf{T}} V_1 \Lambda_1^{-1}$. It is not difficult to see that $\mathsf{range}(C) \subset \mathsf{range}(U^{\mathsf{T}}Y)$. Moreover, $\Pi_C \preceq \mathbb{I}$, $\Pi_{U^{\intercal}Y}\Pi_C\Pi_{U^{\intercal}Y} \preceq \Pi_{U^{\intercal}Y}$. This follows from the fact that $\mathsf{range}(C) \subset \mathsf{range}(U^{\mathsf{T}}Y)$ and the following derivation

$$\Pi_{U^{\intercal}Y} \succeq \Pi_{U^{\intercal}Y}\Pi_C\Pi_{U^{\intercal}Y} = \Pi_C\Pi_{U^{\intercal}Y} = (\Pi_{U^{\intercal}Y}\Pi_C)^{\mathsf{T}} = \Pi_C.$$

An immediate result of the above is the following:

$$\|(\mathbb{I} - \Pi_{U^{\intercal}Y})A'\| \leq \|(\mathbb{I} - \Pi_C)A'\|. \tag{15}$$

Since, $\Pi_C = C(C^{\mathsf{T}}C)^{-1}C^{\mathsf{T}}$, we have the following set of derivations.

$$\begin{aligned}
\Pi_C &= \begin{pmatrix}\mathbb{I}\\ S\end{pmatrix}\left[\begin{pmatrix}\mathbb{I} & S^{\mathsf{T}}\end{pmatrix}\begin{pmatrix}\mathbb{I}\\ S\end{pmatrix}\right]^{-1}\begin{pmatrix}\mathbb{I} & S^{\mathsf{T}}\end{pmatrix} = \begin{pmatrix}\mathbb{I}\\ S\end{pmatrix}\left[\left(\mathbb{I} + S^{\mathsf{T}}S\right)\right]^{-1}\begin{pmatrix}\mathbb{I} & S^{\mathsf{T}}\end{pmatrix}\\
&= \begin{pmatrix}\mathbb{I}(\mathbb{I} + S^{\mathsf{T}}S)^{-1}\\ S(\mathbb{I} + S^{\mathsf{T}}S)^{-1}\end{pmatrix}\begin{pmatrix}\mathbb{I} & S^{\mathsf{T}}\end{pmatrix} = \begin{pmatrix}(\mathbb{I} + S^{\mathsf{T}}S)^{-1} & (\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}}\\ S(\mathbb{I} + S^{\mathsf{T}}S)^{-1} & S(\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}}\end{pmatrix}\\
&\succeq \begin{pmatrix}(\mathbb{I} - S^{\mathsf{T}}S) & (\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}}\\ S(\mathbb{I} + S^{\mathsf{T}}S)^{-1} & 0\end{pmatrix},
\end{aligned}$$

where the last inequality uses the fact that $\mathbb{I} - S^{\mathsf{T}}S \preceq (\mathbb{I} + S^{\mathsf{T}}S)^{-1}$ and $S(\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}} \succeq 0$. Therefore,

$$\mathbb{I} - \Pi_C \preceq \begin{pmatrix}S^{\mathsf{T}}S & \mathbb{I} - (\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}}\\ \mathbb{I} - ((\mathbb{I} + S^{\mathsf{T}}S)^{-1}S^{\mathsf{T}})^{\mathsf{T}} & \mathbb{I}\end{pmatrix}.$$

Conjugating $\mathbb{I} - \Pi_C$ with Λ, and applying the fact that for every positive definite matrix, $P = \begin{pmatrix}X & Y\\ Y^{\mathsf{T}} & Z\end{pmatrix}$, we have $\|P\| \leq \|X\| + \|Z\|$, we get

$$\|(\mathbb{I} - \Pi_C)A'\| \leq \|S^{\mathsf{T}}SA'\| + \|A'\| \tag{16}$$

for any norm. From here on, it is easy arithmetic to show that

$$\|(\mathbb{I} - \Pi_C)A'\| \leq \sqrt{\|\Lambda_2'\| + \|\Lambda_2' V_2^{\mathsf{T}}\Omega(V_1^{\mathsf{T}}\Omega)^{-1}\|} \tag{17}$$

for both the required norms. Using equation (14), equation (15) and equation (16), this gives us a bound on the approximation of matrix A'. Till this point, our analysis closely follows the ideas of [30], accommodating the steps of our algorithm. Now, all that remains is to bound $\|\Lambda_2 V_2^{\mathsf{T}}\Omega(V_1^{\mathsf{T}}\Omega)^{-1}\|$, and for this, we have to analyze the matrix Ω.

ERROR BOUND FOR FROBENIUS NORM We now exploit the rotational invariance of a Gaussian distribution. An important point to note is that $(\Omega\Omega^{\mathsf{T}})^{-1}$ exists and has a well defined trace. The first part of the right hand side of equation (17) is immediate. Thus, if we bound $\mathbb{E}[\|\Lambda_2'\mathbb{V}_2^{\mathsf{T}}\Omega(V_1\Omega)^{-1}\|]$, we are done. This could be accomplished as below.

$$\begin{aligned}
\mathbb{E}[\|\Lambda_2\mathbb{V}_2^{\mathsf{T}}\Omega(V_1\Omega)^{-1}\|] &\leq \sqrt{\mathbb{E}\left[\sum_{ij}|(\Lambda_2)'_{ij}\Pi_{ij}(V_1\Omega^{-1})_{jj}|\right]} \\
&\leq \sqrt{\|\Lambda_2'\|_F\|\Omega^{-1}\|_F} \\
&= \sqrt{\|\Lambda_2'\|_F\mathsf{Tr}\left((\Omega\Omega^{-1})^{\mathsf{T}}\right)} = \sqrt{\|\Lambda_2'\|_F\mathsf{Tr}(\Omega\Omega^{\mathsf{T}})^{-1}} \\
&\leq \sqrt{\mathsf{Tr}(\Omega\Omega^{\mathsf{T}})^{-1}}\min_{rk(A')\leq k}\|A-A'\|_F + \sqrt{(n+d)w\mathsf{Tr}(\Omega\Omega^{\mathsf{T}})^{-1}}.
\end{aligned}$$

The utility guarantee follows by plugging this value in equation (17), and combining equation (13) and the fact that $(\Omega\Omega^{\mathsf{T}})^{-1}$ has a well defined trace $k/(p-1)$ [46].

ERROR BOUND FOR SPECTRAL NORM In order to bound the second term, we use few well known facts in the theory of random matrices to simplify equation (17). In particular, using Lemma 15 and 16, and Holder's inequality, the statement of the theorem for the spectral norm follows. The utility bound then follows using the same arithmetic of representing Λ' in terms of Λ as done in the case of Frobenius norm. In more details, we first bound

$$\begin{aligned}
\mathbb{E}\left[\|\Lambda_2'V_2^{\mathsf{T}}\Omega(V_1^{\mathsf{T}}\Omega)^{-1}\|\right] &\leq \|\Lambda_2'\|\left(\mathbb{E}\left[\|(V_1^{\mathsf{T}}\Omega)^{-1}\|_F^2\|\right]\right)^{1/2} + \|\Lambda_2'\|_F\left(\mathbb{E}\left[\|(V_1^{\mathsf{T}}\Omega)^{-1}\|\right]\right) \\
&\leq \|\Lambda_2'\|\left(\mathbb{E}\left[\|\Omega^{-1}\|_F^2\|\right]\right)^{1/2} + \|\Lambda_2'\|_F\left(\mathbb{E}\left[\|\Omega^{-1}\|\right]\right),
\end{aligned}$$

and then invoke Lemma 16 followed by the sub-additivity of norms. Making these substitution and on simplification, we get the bound stated in equation (12) of Theorem 26.

$$(17) \leq \left(1+\frac{k}{p-1}\right)^{1/2}\|\Lambda_2\|_2 + \frac{e\sqrt{(k+p)\sum_{j>k}\lambda_j^2}}{p} + \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}.$$

□

3.3.1 Tightness of Bounds and Comparison with Earlier Works

We compare our results with the best possible results in the non-private setting. Eckart and Young [22] have shown that the quantity $\min_{rk(A')<k}\|A-A'\|_F$ in the first term of equation (11) is optimal. Likewise, Mirsky [45] proved that λ_{k+1} is the minimum spectral error for k-rank approximation. The second term in equation (12) shows that we also pay for the Frobenius norm error when doing a unified analysis. However, when the oversampling parameter $p \approx k$, then the factor on λ_{k+1} is constant and that on the second term is of order $k^{-1/2}$. In fact, on closer analysis,

$$\|A-\bar{A}\|_2 \leq \left(1+\frac{k}{p-1}+\frac{e(\sqrt{(k+p)\min\{d,n\}}-k)}{p}\right)\lambda_{k+1} + \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon},$$

therefore, the error always lies within some polynomial factor of λ_{k+1}, modulo some additive error. As pointed out of Halko *et al.* [30], one can improve this by power-iteration, the method used by [31, 34] and multiple pass mechanism. However, it seems unlikely to improve it in a single-pass.

Kapralov-Talwar [40] showed a lower bound on additive error when $\delta = 0$ for neighbouring data differing by unit spectral norm. Our privacy proof depends strongly on the fact that $\delta \neq 0$. In fact, our bound is vacuous if $\delta = 0$. Though incomparable due to difference in the notion of neighbouring data, this separation gap further strengthen the belief that better bounds are possible for $\delta \neq 0$.

In the last couple of years, there have been five major works that give a tight bound on differentially private low-rank approximation. We now compare our results with these works in more detail.

Hardt-Roth [33]: The authors use two passes over the input matrix; therefore, it does not fall in our one-pass streaming model of computation. They use the same notion of neighboring data-sets as we do in this paper. This makes their coherence conditions and notion of neighbouring data sets rotationally invariant. As argued by Blocki *et al.* [5], we achieve a better utility bound in the range finding step. Intuitively, this could be seen as a consequence of the absence of additive Gaussian noise. More concretely, Hardt and Roth [33] achieved an error bound of $\sqrt{kn}\log(k/\delta)/\varepsilon + \sqrt{\mu\|A\|_F(n/d)^{1/2}\log(k/\delta)/\varepsilon}$. Their error bound depends on $\|A\|_F$, which can be as large as $\sqrt{nd}$ for binary matrices in the worst case when the matrix is not as well-behaved as captured by low-coherence assumption. On the other hand, we achieve a bound that is independent of $\|A\|_F$.

Hardt-Roth [34]: In some sense, this paper is based on Krylov subspace iteration combined with powering method of Halko *et al.* [30]. They define two data-sets as neighbouring in the same manner as in Hardt and Roth [34]. The coherence definition used in this paper depends on the maximum value of the left or right singular vectors, and is, therefore, rotationally variant. This work assumes that the singular value are well separated, i.e., the first and the k-th singular value has a non-trivial separation, and give LRA in spectral norm. Their bound, however, depends on the rank of the input matrix. Their mechanism uses k rounds of subspace generation, each of which depends on the spectrum of the matrix and uses the power-iteration method of Halko *et al.* [30]. This helps them in achieving better multiplicative bound, but make them unsuitable in a streaming model. A note on multiplicative bound is due here. We believe that the general application of LRA is for thin matrices with very small tail singular values as argued by Deerwester [15] in the work where he initiated the study of LRA. Therefore, we feel that in practical scenario, polynomial multiplicative error would not be that big an issue. The additive error bound computed by [34] $\tau \leq O(k^2\varepsilon^{-1}\sqrt{(\mathsf{rank}(A)\mu + k\log n)\log(1/\delta)}\log n)$ compared to $\tau \leq \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}$ (equation (12)).

Kapralov-Talwar [40]: The only assumption this paper makes is that of singular value separation of the same form as in Hardt and Roth [34]. They also give low-rank approximation in the spectral norm. Additionally, they achieve $(\varepsilon, 0)$-differential privacy, which is not achieved by any other work, including ours. Their definition of neighbouring data sets can be (arguably) considered the most general in the sense that they consider two data sets neighbouring if they differ by at most one in the spectral norm. On the negative side, their mechanism uses k rounds; therefore, it cannot be implemented in a streaming fashion. Since the notion of neighbouring data-sets and privacy guarantee achieved is different from that of ours, we believe our result is incomparable to that of Kapralov and Talwar [40]. However, if we just concentrate on the additive error bound, they achieve a bound of $O(dk^3/(\varepsilon\gamma^2\delta^2))$ compared to our bound $\tau \leq \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}$ (equation (12)).

Hardt [31]: In this recent work, Hardt [31] gave a robust subspace iteration mechanism that allows to publish LRA with noise independent of the rank of the input matrix, thereby, resolving one of the

open problems in Hardt-Roth [34]. They define neighbouring data-sets in the same manner as in Hardt and Roth [33, 34]. However, they also make an assumption on the singular value separation–a separation between the k-th and $(k+1)$-th singular value of the input matrix. Their mechanism uses k rounds of subspace generation, each of which depends on the spectrum of the matrix to reduce the multiplicative error; therefore, it cannot be implemented in a streaming fashion. We achieve a bound of $\frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}$ (equation (12)) compared to $\lambda_1\sqrt{kn\mu\log(1/\delta)\log(n/\gamma)\log\log(n/\gamma)}/\varepsilon\gamma^{1.5}\lambda_k$ of Hardt [31].

Dwork *et al.* [21]: Dwork *et al.* [21] gave the first single-pass online learning algorithm for private low-rank approximation under the assumption that the rows of the input matrix are normalized. They consider the online-learning model, which is very different from our model, and we do not see any natural way to compare. They use the *follow the perturbed leader* (FTL) algorithm of [38] with the binary tree technique of Dwork *et al.* [18]. This idea was previously used by Jain, Kothari, and Thakurta [37] as well. They give a bound that assumes a lower bound of $k\sqrt{n}\log^2(m/\delta)/\varepsilon^2$ on the optimal value, where $\delta < 1/\operatorname{poly}(n)$. More concretely, if OPT is the optimal value, then their error bound is $O(\sqrt{k\mathsf{OPT}}n^{1/4}\log^2(m/\delta))$. We do not make any of the assumptions made by them and, if we just consider the end result, we achieve a bound which is factor $k\sqrt{n}$ better than theirs (see equation (12)). The case that we are able to bypass their lower bound gives a mathematical indication that the unit norm notion of neighbouring data is strictly weaker than user-level privacy.

4 On the Update Time Efficiency

The flexibility of the generic construction in Section 2 is that various other choices of Ω can be used to obtain various guarantees. For example, if we use the projection matrix of Upadhyay [58, 60], one gets an improvement over naive private sketch generation mechanism as far as update time is concerned. One can also use the fast-Johnson Lindenstrauss transform which also preserves differential privacy [59]. For example, let $\mathbf{g} := (\mathbf{g}_1, \cdots, \mathbf{g}_n)$ be n i.d.d. Gaussian samples, $W^{(n)}$ be $n \times n$ Walsh-Hadamard matrix, and $D = \mathsf{Diag}(\mathbf{d})$ a diagonal Rademacher matrix. Using [58, 60], we would replace Ω in Figure 1 by $\frac{1}{\sqrt{r}}PW^{(n)}D$, where (i) in the case of [60], $P = \Pi_{1..r}C$, with C a circulant matrix formed using the vector $\mathbf{g}$, i.e, for $1 \leq i \leq n$, $C_{i:} = (\mathbf{g}_i, \cdots, \mathbf{g}_n, \mathbf{g}_1, \cdots, \mathbf{g}_{i-1})$ and $\Pi_{1..r}$ is the matrix formed by first r rows of Π, and (ii) if we use [58], we use $P = \Pi_{1..r}\mathsf{Diag}(\mathbf{g})\Pi'$ for an n-dimensional Gaussian vector $\mathbf{g}$ and permutations Π and Π'. Upadhyay [60, 58] proved the following ([58, Theorem 10]).

Theorem 27. [58, 60] Let $\eta \in (0,1)$ be a constant, $\mathcal{S}$ a set of m vectors $\mathbf{x} \in \mathbb{R}^n$. Let $\Omega = \frac{1}{\sqrt{r}}PW^{(n)}D$ be an $r \times n$ matrix as constructed above, where $r = c\eta^{-2}\log m \log n$ for a global constant c. Then with probability at least $2/3$, $(1-\eta)\|\mathbf{x}\|_2^2 \leq \|\Omega\mathbf{x}\|_2^2 \leq (1+\eta)\|\mathbf{x}\|_2^2$. Moreover, if the singular values of an $n \times d$ matrix A are at least $(16n/\varepsilon)\log(2/\delta)$, then ΩA is (ε, δ)-differentially private. The time taken to compute ΩA is $O(nd\log n)$.

The sketch of the matrix for MAT-MULT and LIN-REG is now generated row (or column) wise using $\Omega = r^{-1/2}PW^{(n)}D$ instead of a Gaussian matrix in Figure 1. Note that the way in which we reversibly lift the singular values of the private input matrix, the singular values of the private matrix only effects the privacy guarantee and not the utility guarantee; therefore, the bound on τ remains the same in both MAT-MULT and LIN-REG. Since the time taken by matrix-vector multiplication when using $r^{-1/2}PW^{(n)}D$ is $O(n\log n)$, we get an improved update time. The privacy proof requires fresh analysis.

Proof. We break the analysis in two parts. We first use the fact that $W^{(n)}D$ is an isometry [1], i.e., for any vector $\mathbf{x}$ of unit length, $\tilde{\mathbf{x}} = W^{(n)}D\mathbf{x}$ has bounded co-ordinates. Then, we use this promise to prove the

following: when we multiply a circulant matrix formed by Gaussian vector from the right to this smoothen vector and sample r rows, then this preserves the Euclidean norm with high probability.

Fix a $\mathbf{x} \in \mathcal{S}$. Since the transformation is linear, without loss of generality, we can assume $\mathbf{x}$ is a unit vector. The first step follows simply from the simple corollary to Ailon and Chazelle [1].

Theorem 28. *(Ailon-Chazelle [1])* Let $\mathbf{x} \in \mathbb{R}^n$ and $t > 0$. Let $W^{(n)}$ and D be as in Section 4. Then, for any $\kappa > 0$, we have $\mathsf{Pr}\left[\|WD\mathbf{x}\|_\infty \geq \sqrt{2/n \log(2n/\kappa)}\langle x, x\rangle\right] \leq \kappa$.

Proof. Ailon and Chazelle [1] proved that $\mathsf{Pr}\left[|(W^{(n)}D\mathbf{x})_i| \geq t\right] \leq 2\exp(-t^2 n/2)$ for $i \in [n]$. Setting $t = \sqrt{2/n \log(2n/\kappa)}$, we get $\mathsf{Pr}\left[|\tilde{\mathbf{x}}_i| \geq t\right] \leq \kappa/n$. Using union bound gives the result. □

The second step is to use Theorem 28 to get the desired concentration bound. The naive method to work with the permutation in the matrix Ω to get the concentration result makes the proof very lengthy. Let $\tilde{\mathbf{x}} = W^{(n)}D\mathbf{x}$. As in the proof of Dasgupta and Gupta [13], the crucial observation here is that a circulant matrix formed by a vector of i.d.d. Gaussian is very symmetric; therefore, picking any set of r would have the same concentration properties as picking the first r rows. We follow the approach taken by Upadhyay [58]. This is possible because of the nice representation of a circulant matrices by a discrete Fourier transform matrix (see equation (31)). We first get around the problem of dealing with the permutation matrices by making a substitution, i.e., for the matrix Ω and any vector $\mathbf{x} \in \mathbb{R}^n$, $\|\Omega\mathbf{x}\|_2^2 = \sum_{i=1}^r \left|\sum_{j=1}^n \mathbf{g}_{(j-i+1) \mod n}\tilde{\mathbf{x}}_j\right|^2 = \sum_{i=1}^r \left|\sum_{j=1}^n \tilde{\mathbf{x}}_{(i+j+1) \mod n}\right|^2 = \|Z\mathbf{g}\|_2^2$. The substitution becomes much clearer with the following toy example:

$$\left\|\begin{pmatrix}\mathbf{g}_1 & \mathbf{g}_2 & \mathbf{g}_3\\ \mathbf{g}_2 & \mathbf{g}_3 & \mathbf{g}_1\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}}_1\\ \tilde{\mathbf{x}}_2\\ \tilde{\mathbf{x}}_3\end{pmatrix}\right\|_2^2 = \left\|\begin{pmatrix}\mathbf{g}_1\tilde{\mathbf{x}}_1 & \mathbf{g}_2\tilde{\mathbf{x}}_2 & \mathbf{g}_3\tilde{\mathbf{x}}_3\\ \mathbf{g}_2\tilde{\mathbf{x}}_1 & \mathbf{g}_3\tilde{\mathbf{x}}_2 & \mathbf{g}_1\tilde{\mathbf{x}}_3\end{pmatrix}\right\|_2^2 = \left\|\begin{pmatrix}\tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2 & \tilde{\mathbf{x}}_3\\ \tilde{\mathbf{x}}_3 & \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_2\end{pmatrix}\begin{pmatrix}\mathbf{g}_1\\ \mathbf{g}_2\\ \mathbf{g}_3\end{pmatrix}\right\|_2^2$$

The rest of the proof is very similar to [58, 61] owing to our observation in equation (31); we include it for the sake of completion.

Let $U\Sigma V^\mathsf{T}$ be the SVD of Z, $\gamma = V^\mathsf{T}\mathbf{g}$. Let $\sigma := (\sigma_1, \cdots, \sigma_r)$ be the singular values of Z and $(\gamma_1, \cdots, \gamma_r)$ be the co-ordinates of $V^\mathsf{T}\mathbf{g}$. The following equalities are immediate.

$$\begin{aligned}\mathsf{Pr}_{\mathbf{g}}\left[\|\Omega\mathbf{x}\|_2^2 \geq (1+\eta)\right] &= \mathsf{Pr}_{\mathbf{g}}\left[\|Z\mathbf{g}\|_2^2 \geq (1+\eta)r\right] = \mathsf{Pr}_{\mathbf{g}}\left[\|U\Sigma V^\mathsf{T}\mathbf{g}\|_2^2 \geq (1+\eta)r\right]\\ &= \mathsf{Pr}_{\gamma}\left[\|U\Sigma\gamma\|_2^2 \geq (1+\eta)r\right] = \mathsf{Pr}_{\gamma}\left[\|\Sigma\gamma\|_2^2 \geq (1+\eta)r\right]. \qquad (18)\end{aligned}$$

Our problem thus reduces to proving the required concentration bound on $\sum \sigma_i^2|\gamma_i|^2$. We use Corollary 14 by computing the function ψ corresponding to our case through the following proposition.

Proposition 29. Let $(Y_1, \cdots Y_r)$ be distributed according to $\mathcal{N}(\mathbf{0}^r, \mathbb{I})$ and $\sigma = (\sigma_1, \cdots, \sigma_r)$ be any r dimensional vector. Let λ be an arbitrary constant such that $0 < \lambda < 1/2\|\sigma\|_\infty$. Then

$$\sum_{i=1}^r \log\left(\mathbb{E}_{Y_i}\left[\exp\left(\lambda\sigma_i^2(Y_i^2 - 1)\right)\right]\right) \leq \frac{\lambda^2\sum_{i=1}^r \sigma_i^4}{1 - 2\lambda\max_i|\sigma_i|^2}.$$

Proof. Let $(Y_1, \cdots, Y_r)$ be random variable $\mathcal{N}(\mathbf{0}^r, \mathbb{I})$. Equivalently, they are i.d.d. Gaussian. We first consider the simple case when $Y \sim \mathcal{N}(0,1)$ and $R := \log(\mathbb{E}[\exp(a\lambda(Y^2 - 1))])$, then

$$R = a^2\lambda^2\sum_{i\geq 0}\frac{(2\lambda a)^i}{i+1} \leq a\lambda^2\sum_{i\geq 0}(2a\lambda)^i = \frac{a^2\lambda^2}{1-2a\lambda}. \qquad (19)$$

From the linearity of expectation and equation (19), we have the following set of inequalities.

$$\sum_{j=1}^{r} \log\left(\mathbb{E}_{Y_i}\left[\exp(\lambda\sigma_j^2(Y_j^2-1))\right]\right) = \sum_{j=1}^{r} \lambda^2\sigma_j^4 \sum_{i\geq 0} \frac{(2\lambda\sigma_j^2)^i}{i+1}$$
$$\leq \lambda^2 \sum_{j=1}^{r} \sigma_j^4 \sum_{i\geq 0} (2\lambda\sigma_j^2)^i \leq \frac{\lambda^2 \sum_{i=1}^{r} \sigma_i^4}{1-2\lambda \max_i |\sigma_i|^2}.$$

□

Recall that $\Sigma = \mathsf{Diag}(\sigma_1, \cdots, \sigma_r)$ and $Z = U\Sigma V^T$. Since V has orthonormal columns, $V^{\mathsf{T}}\mathbf{g}$ is distributed according to Gaussian distribution. Also, the right hand side of the Proposition 29 has the form $\psi(u) = \frac{a\lambda^2}{2(1-b\lambda)}$ for $a = 2\sum_i \sigma_i^4$ and $b = 2\max_i\left\{\sigma_i^2\right\}$. Using Corollary 14, we have

$$\mathsf{Pr}_\gamma\left[\sum_{i=1}^{r} \sigma_i^2\left(\gamma_i^2-1\right) \geq 2\tau \max_i\left\{\sigma^2\right\} + 2\sqrt{2\tau\sum_i \sigma^4}\right] \leq \exp(-\tau). \tag{20}$$

To apply equation (20), we need an estimate on $\max_i\left\{\sigma_i^2\right\}$ and $\sum_{i=1}^{r}\sigma_i^4$. This is where the isometry of $W^{(n)}D$ is useful. First note that it only gives a stronger guarantee if we consider an upper bound on $\sum \sigma_i^4$ and $\max_i\left\{\sigma_i^2\right\}$ instead of an exact value in equation (20). This is where the guarantee on $\|\tilde{\mathbf{x}}\|_\infty$ is useful. From Theorem 28 and the symmetry of our construction, with probability at least $19/20$,

$$\max_i\left\{\sigma_i^2\right\} = \|Z\|_\infty^2 = \max_{\mathbf{y}\in\mathbb{R}^n, \|\mathbf{y}\|_2=1} \|Z\mathbf{y}\|_2^2 \leq n\|\Pi' W \mathsf{Diag}(\mathbf{d}')\mathbf{y}\|_\infty^2 \leq n\|\tilde{\mathbf{x}}'\|_\infty^2 \leq 2\log(40n), \tag{21}$$

where $\mathbf{d}' = (\mathbf{d}_1\mathbf{x}_1, \cdots \mathbf{d}_n\mathbf{x}_n)$, $\mathbf{x}' = (\mathbf{x}_1\mathbf{y}_1, \cdots, \mathbf{x}_n\mathbf{y}_n)$ and $\tilde{\mathbf{x}}' = W^{(n)}D\mathbf{x}'$ with Rademacher entries $(\mathbf{d}_1, \cdots \mathbf{d}_n)$.

Since $\|Z\|_F^2 = \sum_{i=1}^{r}\sigma_i^2 = r$. Thus,

$$\sum_i \sigma_i^4 \leq \max_i\left\{\sigma_i^2\right\} \cdot \sum_i \sigma_i^2 = 2r\log(40n). \tag{22}$$

Setting $\tau = cr\varepsilon^2/\log(40n)$ for a small constant c, and using equations (18), (20), and (22), we have

$$\mathsf{Pr}_{\mathbf{g}}[\|\Omega\mathbf{x}\|_2^2 \geq (1+\varepsilon)] = \mathsf{Pr}_\gamma\left[\sum_{j=1}^{r}\sigma_j^2(|\gamma_j|-1) \geq \varepsilon r\right] < \exp\left(-\frac{r\varepsilon^2}{\log(40n)}\right). \tag{23}$$

For $(23) < 1/6m$, we need $r = O\left(\varepsilon^{-2}\log n \log m\right)$. The result follows using the union bound and similar analysis for the negative side of the tail, i.e., for the value of r, we have

$$\mathsf{Pr}_{\mathbf{g}}[\|\Omega\mathbf{x}\|_2^2 \leq (1-\varepsilon)] = \mathsf{Pr}_\gamma\left[\sum_{j=1}^{r}\sigma_j^2(|\gamma_j|-1) \leq -\varepsilon r\right] < \frac{1}{6m}. \tag{24}$$

Combining equation (23) and equation (24) and union bound over all $x \in \mathcal{S}$, the result follows.

Note that the way in which we reversibly lift the singular values of the private input matrix, the singular values of the private matrix only effects the privacy guarantee and not the utility guarantee; therefore, the bound on τ remains the same in both MAT-MULT and LIN-REG when using Theorem 27 instead of random Gaussian matrix in Figure 1.

For the run-time guarantee, note that a circulant matrix has a singular value decomposition in the form of discrete Fourier transform (equation (31)). Therefore, in disguise, all the matrices used in the projection matrix allow fast matrix-vector multiplication. The final truncated permutation matrix takes $O(r)$ time to sample the corresponding entry after the application of CWD; therefore, the run-time of a single matrix-vector multiplication takes $O(n\log n)$ time. Since, there are d columns in the matrix A, it takes $O(nd\log n)$ time to publish the sanitized matrix.

PRIVACY PROOF. Before we give our proof, we state the reason why the proof of Blocki *et al.* [5] does not extend to our case. One of the reasons why the proof of Blocki *et al.* [5] does not generalize to any Johnson-Lindenstrauss transform in general is due to its strong dependency on the fact that each samples in a dense Gaussian matrix is picked i.d.d. More concretely, each row of their published matrix is a multivariate Gaussian and preserves differential privacy. This allows them to use the composition theorem of Dwork, Rothblum and Vadhan [20] to prove differential privacy of the entire published matrix. Unfortunately, we cannot invoke the composition theorem once we reuse the random samples due to the introduction of correlations between the entries of our projection matrices. Therefore, applying our projection matrix to a private matrix does not yield r independent multivariate distribution, but one matrix-variate distribution. We compute the resulting distribution and then prove that it preserves differential privacy. In this sense, our proof uses the same idea as used by Upadhyay [60]; however, the analogy ends here as the probability distribution are very different and requires a fresh analysis.

Our starting point is an alternate way to look at any matrix P as a matrix formed by sampling r rows of a fully circulant matrix independently. In other words, one can see P as a product of a truncated permutation matrix and a circulant matrix formed by $\mathbf{g}$. Since differential privacy is preserved under any arbitrary post-processing, we can just concentrate on the distribution when a fully circulant matrix is used instead of the matrix P (the truncated permutation can be seen as a post-processing step). Let $\mathbb{I}_k$ denote the $k \times k$ identity matrix. Then for a partial circulant matrix, we have,

$$P = \begin{pmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_n \\ \mathbf{g}_2 & \cdots & \mathbf{g}_n & \mathbf{g}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}_r & \mathbf{g}_1 & \cdots & \mathbf{g}_{n-1} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbb{I}_r & 0 \end{pmatrix}}_{n \text{ columns}} \begin{pmatrix} \mathbf{g}_1 & \mathbf{g}_2 & \cdots & \mathbf{g}_n \\ \mathbf{g}_2 & \cdots & \mathbf{g}_n & \mathbf{g}_1 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{g}_n & \mathbf{g}_1 & \cdots & \mathbf{g}_{n-1} \end{pmatrix} = \begin{pmatrix} \mathbb{I}_r & 0 \end{pmatrix} C. \tag{25}$$

Therefore, for the rest of this proof, we just concentrate on circulant matrices. Let $\mathsf{vec}(C)$ denote the vector formed by the entries of C. Then, the covariance matrix of $\mathsf{vec}(C)$ is,

$$\Lambda := \mathsf{COV}(\mathsf{vec}(C)) = \underbrace{\begin{pmatrix} \mathbb{I}_{n/2} & 0 & 0 & \mathbb{I}_{n-1} & 0 & \mathbb{I}_{n-2} & \cdots & 0 & \mathbb{I}_1 \\ 0 & \mathbb{I}_{n/2} & \mathbb{I}_1 & 0 & \mathbb{I}_2 & 0 & \cdots & \mathbb{I}_{n-1} & 0 \\ 0 & \mathbb{I}_1 & \mathbb{I}_{n/2} & 0 & \cdots & \cdots & \cdots & \vdots & \vdots \\ \mathbb{I}_{n-1} & 0 & 0 & \mathbb{I}_{n/2} & & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbb{I}_{n-2} & \vdots & \vdots & \vdots & \ddots & \cdots & \mathbb{I}_{n/2+1} & 0 \\ \mathbb{I}_2 & 0 & \cdots & \cdots & \cdots & \cdots & \ddots & 0 & \mathbb{I}_{n/2-1} \\ 0 & \mathbb{I}_{n-1} & 0 & \mathbb{I}_{n-2} & \cdots & \cdots & \cdots & \mathbb{I}_{n/2} & 0 \\ \mathbb{I}_1 & 0 & \mathbb{I}_2 & 0 & \cdots & & \cdots & 0 & \mathbb{I}_{n/2} \end{pmatrix}}_{n^2 \text{ columns}} n^2 \text{ rows}, \tag{26}$$

where 0 are block zero matrices of appropriate dimension. To simplify the presentation of the proof, we make couple of changes. We assume that our projection matrix is an $n \times r$ matrix instead of $r \times n$ matrix by considering the transpose of matrix constructed in Section 4, i.e., we use Ω instead of the transpose of the construction stated in Section 4, i.e., we consider our projection matrix is $n \times r$ and is applied on the right of A^{T}. With this change of notation, this implies that we now publish $A^{\mathsf{T}}\Omega$. Since WD is an isometry, $\|(A - \tilde{A})^{\mathsf{T}} DW^{\mathsf{T}}\| = \|A' - \tilde{A}'\| \leq \|A - \tilde{A}\| \leq 1$ for $A' = A^{\mathsf{T}} DW^{\mathsf{T}}$ and $\tilde{A}' = \tilde{A}^{\mathsf{T}} DW^{\mathsf{T}}$, so without any loss of generality, we can analyze the distribution $A'^{\mathsf{T}} C$ instead of $A^{\mathsf{T}}\Omega$. From hereon, whenever we say A, we mean A' and $\tilde{A}$ to mean $\tilde{A}'$.

In order to compute the PDF of the matrix-variate distribution corresponding to the published matrix, we follow the standard technique. We look at the published matrix as a vector and analyze the corresponding multivariate distribution. A covariance matrix is a positive semi-definite matrix; therefore, we can write equation (26) succinctly in the form of its Cholesky decomposition, say $\Lambda = LL^{\mathsf{T}}$.

Note that $\Lambda\Lambda^T = n\mathbb{I}$; therefore all the singular values of L is $\sqrt{n}$. Using the left spherical symmetry of Gaussian distribution, and noting that the Jacobian of the transformation $A^{\mathsf{T}}C$ is $\sqrt{\det(A^{\mathsf{T}}A)}$, the resulting matrix variate distribution for $X \sim A^{\mathsf{T}}Y$ for Y picked from a distribution with mean vector 0 and covariance matrix Λ has the covariance matrix $\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A}$, where $\mathbb{A}$ is the matrix formed by stacking n copies of A row-wise. Therefore,

$$\mathsf{PDF}_{A^{\mathsf{T}}C}(X) = \frac{1}{\sqrt{\det(\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A})}} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(X^{\mathsf{T}}(\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A})^{-1}X\right)\right). \tag{27}$$

For the sake of simplicity, let us denote by $\mathbb{B} = L^{\mathsf{T}}\mathbb{A}$. Then from equation (27), we can write the distribution of the published matrices corresponding to the neighbouring matrices A and $\tilde{A}$ as follows.

$$\mathsf{PDF}_{A^{\mathsf{T}}C}(X) = \frac{1}{\sqrt{\det(\mathbb{B}^{\mathsf{T}}\mathbb{B})}} \exp\left(-\frac{1}{2}\mathrm{Tr}(X^{\mathsf{T}}(\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}X)\right),$$
$$\mathsf{PDF}_{\tilde{A}^{\mathsf{T}}C}(X) = \frac{1}{\sqrt{\det(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})}} \exp\left(-\frac{1}{2}\mathrm{Tr}(X^{\mathsf{T}}(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}X)\right).$$

Let the singular value decomposition of $\mathbb{B} = \mathbb{U}\Sigma\mathbb{V}^{\mathsf{T}}$. Similarly, define $\tilde{\mathbb{B}} = L^{\mathsf{T}}\tilde{\mathbb{A}} = \tilde{\mathbb{U}}\tilde{\Sigma}\tilde{\mathbb{V}}^{\mathsf{T}}$. First note that the singular values of $\mathbb{B}$ are $\sqrt{n}$ times the singular values of A because $\Lambda\Lambda^{\mathsf{T}} = n\mathbb{I}$: L is a scaled orthonormal matrix that multiplied with n-copies of the left singular vectors of A forms the matrix $\sqrt{n}\mathbb{U}$ (recall that unitary group is a subgroup of general linear group $\mathbb{GL}(n)$). Similarly, for $\tilde{\mathbb{B}}$. We prove the following.

Lemma 30. For a matrix A with all singular values greater than $\Omega\left(\frac{n}{\varepsilon}\log(4/\delta)\right)$, the following holds

$$\exp(-\varepsilon) \leq \sqrt{\frac{\det(\mathbb{B}^{\mathsf{T}}\mathbb{B})}{\det(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})}} \leq \exp(\varepsilon). \tag{28}$$

If $X = A^{\mathsf{T}}C$, then

$$\mathsf{Pr}\left[\left|\mathrm{Tr}\left(X^{\mathsf{T}}\left((\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1} - (\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}\right)X\right)\right| \leq \varepsilon\right] \geq 1 - \delta. \tag{29}$$

Proof. The first part (equation (28)) is exactly as in Blocki *et al.* [5]. We include it for the sake of completion. More concretely, we have $\det(A^{\mathsf{T}}A) = \prod_i \sigma_i^2$, where $\sigma_1 \geq \cdots \geq \sigma_d \geq \sigma_{\mathsf{min}}$ are the singular values of A. Let $\lambda_1 \geq \cdots \geq \lambda_d \geq \sigma_{\mathsf{min}}$ be the singular value for $\tilde{A}$. Let $G = \{i : \lambda_i > \sigma_i\}$. Since the singular values of $A - \tilde{A}$ and $\tilde{A} - A$ are the same, $\sum_{i\in G}(\sigma_i - \lambda_i) \leq 1$ using Linskii's theorem. Therefore,

$$\sqrt{\frac{\det(A^{\mathsf{T}}A)}{\det(\tilde{A}^{\mathsf{T}}\tilde{A})}} = \sqrt{\prod_i \frac{\lambda_i^2}{\sigma_i^2}} \leq \exp\left(\frac{\varepsilon}{16n\log(2/\delta)}\right)\sum_i(\lambda_i - \sigma_i) \leq \exp\left(\varepsilon/2n\right).$$

Similarly, we can bound $\frac{\det(\tilde{A}^{\mathsf{T}}\tilde{A})}{\det(A^{\mathsf{T}}A)} \leq \exp(\varepsilon/2n)$.

PROOF OF EQUATION (29). In this part, we bound the following expression.

$$\left|\mathrm{Tr}\left(X^{\mathsf{T}}\left(\left(\mathbb{B}^{\mathsf{T}}\mathbb{B}\right)^{-1}-\left(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}}\right)^{-1}\right)X\right)\right|. \tag{30}$$

We can write $\tilde{A} = A + \mathbf{e}_i\mathbf{v}^{\mathsf{T}}$ for some i and a unit vector v. Let $\mathcal{E}$ be a matrix formed by n-copies of $\mathbf{e}_i\mathbf{v}^{\mathsf{T}}$ stacked together. Then $\tilde{\mathbb{B}} = L^{\mathsf{T}}\tilde{\mathbb{A}} = L^{\mathsf{T}}(\mathbb{A}+\mathcal{E}) = \mathbb{B} + L^{\mathsf{T}}\mathcal{E}$. The following is immediate.

$$\begin{aligned}
\mathrm{Tr}\left(X^{\mathsf{T}}\left(\left(\mathbb{B}^{\mathsf{T}}\mathbb{B}\right)^{-1}-\left(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}}\right)^{-1}\right)X\right) &= \mathrm{Tr}\left(X^{\mathsf{T}}\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}-(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)X\right)\\
&= \mathrm{Tr}\left(X^{\mathsf{T}}\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}+L^{\mathsf{T}}\mathcal{E})^{\mathsf{T}}(\mathbb{B}+L^{\mathsf{T}}\mathcal{E})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}-(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)X\right)\\
&= \mathrm{Tr}\left(X^{\mathsf{T}}\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}L^{\mathsf{T}}\mathcal{E}+\mathcal{E}^{\mathsf{T}}L\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)X\right)\\
&= \mathrm{Tr}\left(C^{\mathsf{T}}A\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}L^{\mathsf{T}}\mathcal{E}+\mathcal{E}^{\mathsf{T}}L\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)A^{\mathsf{T}}C\right)\\
&\le \mathrm{Tr}\left(CC^{\mathsf{T}}\right)\mathrm{Tr}\left(A\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}L^{\mathsf{T}}\mathcal{E}+\mathcal{E}^{\mathsf{T}}L\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)A^{\mathsf{T}}\right)\\
&= \mathrm{Tr}\left(CC^{\mathsf{T}}\right)\mathrm{Tr}\left(A^{\mathsf{T}}A\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}L^{\mathsf{T}}\mathcal{E}+\mathcal{E}^{\mathsf{T}}L\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)\right)\\
&\le \underbrace{\mathrm{Tr}\left(CC^{\mathsf{T}}\right)}_{S}\underbrace{\mathrm{Tr}\left(A^{\mathsf{T}}A\right)\mathrm{Tr}\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})\right)^{-1}\mathrm{Tr}\left((\tilde{\mathbb{B}}^{\mathsf{T}}L^{\mathsf{T}}\mathcal{E}+\mathcal{E}^{\mathsf{T}}L\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)}_{Q}
\end{aligned}$$

where the inequalities follows from the fact that $\mathrm{Tr}(XY) \le \mathrm{Tr}(X)\mathrm{Tr}(Y)$ for positive definite Hermitian matrices X and Y. We bound each of the above trace terms. To bound S, we recall the fundamental relation between discrete Fourier transform and circulant matrices. Recall that C is made by circulating the Gaussian vectors $(\mathbf{g}_1, \cdots, \mathbf{g}_n)$ to form an $n \times n$ matrix. Then $C = \mathcal{F}_n\mathsf{Diag}(\mathcal{F}_n\mathbf{g})\mathcal{F}_n^{-1}$. Therefore, to bound the trace of CC^{T}, we have to bound the following.

$$CC^{\mathsf{T}} = \mathcal{F}_n\mathsf{Diag}(\mathcal{F}_n\mathbf{g})\mathcal{F}_n^{-1}\left[\mathcal{F}_n\mathsf{Diag}(\mathcal{F}_n\mathbf{g})\mathcal{F}_n^{-1}\right]^{\mathsf{T}} = \mathcal{F}_n\mathsf{Diag}(|\mathcal{F}_n\mathbf{g}|^2)\mathcal{F}_n^{-1}. \tag{31}$$

Since $\mathbf{g}$ and $Z\mathbf{g}$ are equi-distributed when the rows of Z are orthonormal, we need the following lemma to bound S.

Lemma 31. Let $\mathbf{h}_1, \cdots, \mathbf{h}_n$ be n i.i.d. $\mathcal{N}(0,1)$ random variables. Then,

$$\mathsf{Pr}\left[\sum_{i=1}^{n}\mathbf{h}_i^2 > 2(1+\theta)n\right] \le 2^{-\theta n/2}.$$

Proof. Now $\mathsf{Pr}[\mathbf{h}_i = t] = \frac{1}{\sqrt{2\pi}}\exp\left(-t^2/2\right)$. Then consider the random variable $Z_i = \exp(\mathbf{h}_i^2/4)$. Then

$$\mathbb{E}[Z_i] = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-t^2/2\right)\exp(-t^2/4)\,\mathrm{d}t = \sqrt{2}.$$

Now, observe that,

$$\begin{aligned}
\mathsf{Pr}_{\mathbf{h}_1,\cdots\mathbf{h}_n}[\mathbf{h}_1^2+\cdots\mathbf{h}_n^2 > \lambda] &= \mathsf{Pr}_{\mathbf{h}_1,\cdots,\mathbf{h}_n}\left[\frac{\mathbf{h}_1^2+\cdots\mathbf{h}_n^2}{4} > \frac{\lambda}{4}\right]\\
&= \mathsf{Pr}_{\mathbf{h}_1,\cdots,\mathbf{h}_n}\left[\exp\left(\frac{\mathbf{h}_1^2+\cdots\mathbf{h}_n^2}{4}\right) > \exp\left(\frac{\lambda}{4}\right)\right]\\
&\le \exp(-\lambda/4)\mathbb{E}_{\mathbf{h}_1,\cdots,\mathbf{h}_n}\left[\exp\left(\frac{\mathbf{h}_1^2+\cdots\mathbf{h}_n^2}{4}\right)\right].
\end{aligned}$$

Since all $\mathbf{h}_i$ are i.d.d., the above expression is bounded as

$$\prod_{i=1}^{n} \mathbb{E}\left[\exp\left(\frac{\mathbf{h}_i^2}{4}\right)\right] = \prod_{i=1}^{n} \mathbb{E}\left[Z_i\right] = 2^{n/2}.$$

Putting $\lambda = 2(1+\theta)n$, the lemma follows. □

We set $\theta = (1/n)\log(1/\delta)$ to get

$$\mathsf{Pr}\left[\sum_{i=1}^{n} \mathbf{h}_i^2 > 2(1+\theta)n\right] \leq 2^{-\theta n/2} \leq \delta/2 \quad \Rightarrow \quad \mathsf{Pr}\left[S > 2n^2\left(1 + \frac{1}{n}\log\left(\frac{1}{\delta}\right)\right)\right] \leq \delta/2.$$

The term Q is easy to compute once we note the following.

$$n\mathrm{Tr}\left(A^{\mathsf{T}}A\right) \leq \mathrm{Tr}\left(\mathbb{V}\Sigma^2\mathbb{V}^{\mathsf{T}}\right) = \mathrm{Tr}\left(\mathbb{B}^{\mathsf{T}}\mathbb{B}\right) \leq n^{3/2}\mathrm{Tr}\left(A^{\mathsf{T}}A\right)$$

since $\Lambda\Lambda^{\mathsf{T}} = n\mathbb{I}_{n^2\times n^2}$ and $\mathrm{Tr}(\Lambda) = n^{3/2}$. Now $\mathbf{e}_i$ and $\mathbf{v}$ are unit vectors which forms the matrix $\mathcal{E}$, and the singular values of $A, \tilde{A}$ are at least σ_{min} with $\tilde{A} - A = \mathbf{e}_i\mathbf{v}^{\mathsf{T}}$. Using the singular value decomposition of $\tilde{B}$, observing that L is a scaled unitary by $\sqrt{n}$, and combining with the invariance of trace of matrices under cyclic permutations with the singular value decomposition of $\tilde{B}$, it is easy to see that Q is bounded by $1/\sigma_{\mathsf{min}}$. Using Lemma 31, equation (31), and the bound on σ_{min}, we have

$$\mathsf{Pr}\left[(30) \leq \frac{4n\ln(4n/\delta)}{\sigma_{\mathsf{min}}} \leq 5\varepsilon\right] \geq 1 - \delta.$$

Rescaling the value of ε, the lemma follows. □

It is straightforward to see that Lemma 30 implies the privacy guarantee in Theorem 27.

□

ACKNOWLEDGEMENTS. I would like to thank Prateek Jain for the insightful discussion and suggestion of matrices of the form used in Section 4. My discussion with him led to the first two problems studied in this paper. I would like to thank Shweta Agarwal and Ragesh Jaiswal for the discussions in the early stages of this work, Or Sheffet for his input on the initial draft of this work, Abhradeep Guha Thakurta for clarifying the difference between our model and the online learning model, and Shitikanth Kashyap for proof-reading an earlier draft. I would also like to thank the anonymous reviewers of CRYPTO 2014 for various suggestions that helped in improving the presentation of this paper. I am extremely thankful to the anonymous reviewers of SODA 2014 for various references (like, Munro-Paterson [47], Flajolet-Martin [24], and Kane-Nelson [39]) and specifically pointing the improvements for utility bounds for MAT-MULT through Kane and Nelson [39].

References

[1] Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009. 5, 22, 23

[2] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 20–29. ACM, 1996. 1, 9

[3] Amos Beimel, Kobbi Nissim, and Eran Omri. Distributed private data analysis: Simultaneously solving how and what. In *Advances in Cryptology–CRYPTO 2008*, pages 451–468. Springer, 2008. 4

[4] Lucien Birgé and Pascal Massart. *From Model Selection to Adaptive Estimation*, pages 55–87. Springer New York, 1997. 7, 8

[5] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss Transform Itself Preserves Differential Privacy. In *FOCS*, pages 410–419, 2012. 1, 5, 6, 9, 10, 12, 21, 25, 26, 32

[6] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. Differentially private data analysis of social networks via restricted sensitivity. In Robert D. Kleinberg, editor, *ITCS*, pages 87–96. ACM, 2013. 1, 5, 6

[7] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In Chen Li, editor, *PODS*, pages 128–138. ACM, 2005. 6

[8] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *J. ACM*, 60(2):12, 2013. 6

[9] E Bura and R Pfeiffer. On the distribution of the left singular vectors of a random matrix and its applications. *Statistics & Probability Letters*, 78(15):2275–2280, 2008. 17, 18

[10] Lynn E Cannon. A cellular computer to implement the kalman filter algorithm. Technical report, DTIC Document, 1969. 4

[11] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *NIPS*, pages 998–1006, 2012. 6

[12] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *STOC*, pages 205–214, 2009. 1, 2, 3, 4, 5, 6, 15, 17, 18

[13] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003. 23

[14] Angus Deaton. *Understanding consumption*. Oxford University Press, 1992. 3

[15] Scott Deerwester. Improving information retrieval with latent semantic indexing. 1988. 3, 21

[16] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *EUROCRYPT*, pages 486–503, 2006. 6

[17] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In Shai Halevi and Tal Rabin, editors, *TCC*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006. 6

[18] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In Schulman [55], pages 715–724. 2, 5, 22

[19] Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N Rothblum, and Sergey Yekhanin. Pan-Private Streaming Algorithms. In *ICS*, pages 66–80, 2010. 1, 6

[20] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and Differential Privacy. In *FOCS*, pages 51–60. IEEE Computer Society, 2010. 6, 9, 25

[21] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze Gauss: Optimal Bounds for Privacy-Preserving Principal Component Analysis. In *STOC*, pages 11–20, 2014. 3, 4, 5, 6, 17, 22

[22] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 20

[23] Ronald G Ehrenberg and Robert S Smith. Modern labor economics. 2010. 3

[24] Philippe Flajolet and G Nigel Martin. Probabilistic counting algorithms for data base applications. *Journal of computer and system sciences*, 31(2):182–209, 1985. 1, 28

[25] Lars Grasedyck and Wolfgang Hackbusch. Construction and arithmetics of H-matrices. *Computing*, 70(4):295–334, 2003. 3

[26] Leslie Greengard and Vladimir Rokhlin. A new version of the fast multipole method for the laplace equation in three dimensions. *Acta numerica*, 6:229–269, 1997. 3

[27] Anupam Gupta, Moritz Hardt, Aaron Roth, and Jonathan Ullman. Privately Releasing Conjunctions and the Statistical Query Barrier. *SIAM J. Comput.*, 42(4):1494–1520, 2013. 6

[28] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative Constructions and Private Data Release. In *TCC*, pages 339–356, 2012. 6

[29] Arjun K Gupta and Daya K Nagar. *Matrix variate distributions*, volume 104. CRC Press, 1999. 7

[30] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011. 4, 6, 16, 17, 18, 19, 21

[31] Moritz Hardt. Robust subspace iteration and privacy-preserving spectral analysis. In *Allerton*, pages 1624–1626. IEEE, 2013. 1, 4, 5, 21, 22

[32] Moritz Hardt, Katrina Ligett, and Frank McSherry. A Simple and Practical Algorithm for Differentially Private Data Release. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger, editors, *NIPS*, pages 2348–2356, 2012. 6

[33] Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. In *STOC*, pages 1255–1268, 2012. 1, 2, 3, 4, 5, 6, 13, 16, 21, 22

[34] Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *STOC*, pages 331–340, 2013. 1, 2, 4, 5, 6, 16, 21, 22

[35] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005. 3

[36] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 604–613, 1998. 15

[37] Prateek Jain, Pravesh Kothari, and Abhradeep Thakurta. Differentially Private Online Learning. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *COLT*, volume 23 of *JMLR Proceedings*, pages 24.1–24.34. JMLR.org, 2012. 3, 6, 9, 22

[38] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005. 22

[39] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss Transforms. *J. ACM*, 61(1):4, 2014. 2, 3, 4, 5, 14, 15, 28

[40] Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *SODA*, volume 5, page 1. SIAM, 2013. 1, 4, 5, 6, 16, 21

[41] Krishnaram Kenthapadi, Aleksandra Korolova, Ilya Mironov, and Nina Mishra. Privacy via the johnson-lindenstrauss transform. *arXiv preprint arXiv:1204.2606*, 2012. 6

[42] Erricos John Kontoghiorghes. *Handbook of parallel computing and statistics*. CRC Press, 2010. 4

[43] Pascal Massart and Jean Picard. *Concentration inequalities and model selection*, volume 1896. Springer, 2007. 8

[44] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 94–103. IEEE, 2007. 6

[45] Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, 11(1):50–59, 1960. 20

[46] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009. 8, 10, 20

[47] J Ian Munro and Mike S Paterson. Selection and sorting with limited storage. *Theoretical computer science*, 12(3):315–323, 1980. 1, 6, 9, 28

[48] Gilles Pisier. Some applications of the metric entropy condition to harmonic analysis. In *Banach Spaces, Harmonic Analysis, and Probability Theory*, pages 123–154. Springer, 1983. 8

[49] C Radhakrishna Rao. *Linear statistical inference and its applications*, volume 22. John Wiley & Sons, 2009. 10, 11

[50] Radim Rehurek. Subspace tracking for latent semantic analysis. In *Advances in Information Retrieval*, pages 289–300. Springer, 2011. 10, 13

[51] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In Schulman [55], pages 765–774. 6

[52] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009. 18

[53] Youcef Saad and Martin H Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986. 4

[54] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 143–152. IEEE, 2006. 1, 3, 4, 5, 6, 15, 17

[55] Leonard J. Schulman, editor. *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*. ACM, 2010. 29, 31

[56] Volker Strumpen, Henry Hoffmann, and Anant Agarwal. A Stream Algorithm for the SVD. 2003. 10, 13

[57] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012. 8

[58] Jalaj Upadhyay. Randomness Efficient Johnson-Lindenstrauss Transform with Applications in Compressed Sensing, Differential Privacy, and Functional Analysis. *Manuscipt, Submitted to STOC, 2015.* 5, 22, 23

[59] Jalaj Upadhyay. Random Projections, Graph Sparsification, and Differential Privacy. In *ASIACRYPT (1)*, pages 276–295, 2013. 1, 5, 6, 9, 22

[60] Jalaj Upadhyay. Circulant matrices and differential privacy. *IACR Cryptology ePrint Archive*, 2014:818, 2014. 5, 22, 25

[61] Jan Vybíral. A variant of the Johnson–Lindenstrauss lemma for circulant matrices. *Journal of Functional Analysis*, 260(4):1096–1105, 2 2011. 23

A Missing Proofs

A.1 Differential Privacy of Variant 1

We give the proof Blocki *et al.* [5] for the sake of completion. We denote by $\tilde{A}$ the matrix that differs from A by at most one entry. In other word, if A and $\tilde{A}$ differs in row i, then there exists a unit vector v such that $A - \tilde{A} = E = \mathbf{v}\mathbf{e}_i^{\mathsf{T}}$. Let $U\Sigma V^{\mathsf{T}}$ ($\tilde{U}\Lambda\tilde{V}^{\mathsf{T}}$, respectively) be the SVD of A ($\tilde{A}$, respectively).

The PDF for the two distributions, corresponding to A and $\tilde{A}$, is just a linear transformation of $\mathcal{N}(0, \mathbb{I}_{n\times n})$. Therefore,

$$\mathsf{PDF}_{A^{\intercal}Y}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \Delta(A^{\mathsf{T}}A)}} \exp(-\frac{1}{2}\langle \mathbf{x}|(A^{\mathsf{T}}A)^{-1}|\mathbf{x}\rangle)$$
$$\mathsf{PDF}_{\tilde{A}^{\intercal}Y}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \Delta(\tilde{A}^{\mathsf{T}}\tilde{A})}} \exp(-\frac{1}{2}\langle \mathbf{x}|(\tilde{A}^{\mathsf{T}}\tilde{A})^{-1}|\mathbf{x}\rangle)$$

We prove the result for a row of the published matrix; the theorem follows from Theorem 19. It is straightforward to see that combination of the following proves differential privacy for a row of published matrix:

$$\sqrt{\frac{\det(A^{\mathsf{T}}A)}{\det(\tilde{A}^{\mathsf{T}}\tilde{A})}} \in \exp(\pm\varepsilon_0) \quad \text{and} \quad \Pr\left[|\langle \mathbf{x}|\Omega^{\mathsf{T}}(A^{\mathsf{T}}A)^{-1}\Omega|\mathbf{x}\rangle - \langle \mathbf{x}|\Omega^{\mathsf{T}}(\tilde{A}^{\mathsf{T}}\tilde{A})^{-1}\Omega|\mathbf{x}\rangle| \leq \varepsilon_0\right] \geq 1 - \delta_0,$$

where $\varepsilon_0 = \frac{\varepsilon}{\sqrt{4r\ln(2/\delta)}}$ and $\delta_0 = \delta/2r$.

The first part of the proof follows simply as in Blocki *et al.* [5]. More concretely, we have $\det(A^{\mathsf{T}}A) = \prod_i \sigma_i^2$, where $\sigma_1 \geq \cdots \geq \sigma_d \geq \sigma_{\mathsf{min}}$ are the singular values of A. Let $\lambda_1 \geq \cdots \geq \lambda_d \geq \sigma_{\mathsf{min}}$ be its singular value for $\tilde{A}$. Since the singular values of $A - \tilde{A}$ and $\tilde{A} - A$ are the same, $\sum_i(\sigma_i - \lambda_i) \leq 1$ using Linskii's theorem. Therefore,

$$\sqrt{\prod_i \frac{\lambda_i^2}{\sigma_i^2}} \leq \exp\left(\frac{\varepsilon}{32\sqrt{r\log(2/\delta)}\log(r/\delta)}\right)\sum_i(\lambda_i - \sigma_i) \leq e^{\varepsilon_0/2}.$$

The second part of the proof is slightly more involved. Each row i of the published matrix is distributed identically and is constructed by multiplying an n-dimensional vector Ω_i that has entries picked from a normal distribution $\mathcal{N}(0,1)$. Note that $\mathbb{E}[\Omega_{i:}] = \mathbf{0}^n$ and $\mathsf{COV}(\Omega_{i:}) = \mathbb{I}$. Then

$$\langle \mathbf{x}|\Omega^{\mathsf{T}}(A^{\mathsf{T}}A)^{-1}\Omega|\mathbf{x}\rangle - \langle \mathbf{x}|\Omega^{\mathsf{T}}(\tilde{A}^{\mathsf{T}}\tilde{A})^{-1}\Omega|\mathbf{x}\rangle = \langle \mathbf{x}|\Omega^{\mathsf{T}}\left[(A^{\mathsf{T}}A)^{-1}(A^{\mathsf{T}}E + E^{\mathsf{T}}\tilde{A})(\tilde{A}^{\mathsf{T}}\tilde{A})^{-1}\right]\Omega|\mathbf{x}\rangle.$$

Using the singular value decomposition of $A = U\Sigma U^{\mathsf{T}}$ and $\tilde{A} = \tilde{U}\Lambda\tilde{U}^{\mathsf{T}}$, this simplifies as

$$\left[\langle \mathbf{x}|\Omega^{\mathsf{T}}(V\Sigma^{-1}U^{\mathsf{T}})\mathbf{e}_i\right]\left[\mathbf{v}^{\mathsf{T}}(\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}})\Omega|\mathbf{x}\rangle\right] + \langle \mathbf{x}|\Omega^{\mathsf{T}}\left[(V\Sigma^{-2}V^{\mathsf{T}})\mathbf{v}\right]\left[\mathbf{e}_i^{\mathsf{T}}(\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}})\Omega|\mathbf{x}\rangle\right].$$

Since $\mathbf{x} \sim A^{\mathsf{T}}\mathbf{y}$, where $\mathbf{y} \sim \mathcal{N}(0,1)$, we can further simplify it as

$$\underbrace{\left[\mathbf{y}^{\mathsf{T}}A\Omega^{\mathsf{T}}(V\Sigma^{-1}U^{\mathsf{T}})\mathbf{e}_i\right]}_{t_1}\underbrace{\left[\mathbf{v}^{\mathsf{T}}(\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}})\Omega A^{\mathsf{T}}\mathbf{y}\right]}_{t_2} + \underbrace{\left[\mathbf{y}^{\mathsf{T}}A\Omega^{\mathsf{T}}(V\Sigma^{-2}V^{\mathsf{T}})\mathbf{v}\right]}_{t_3}\underbrace{\left[\mathbf{e}_i^{\mathsf{T}}(\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}})\Omega A^{\mathsf{T}}\mathbf{y}\right]}_{t_4}.$$

Now since $\|\Lambda\|_2, \|\Sigma\|_2 \geq w$, plugging in the SVD of A and $A - A' = \mathbf{e}_i\mathbf{v}^{\mathsf{T}}$, and that every term t_i in the above expression is a linear combination of a Gaussian, i.e., each term is distributed as per $\mathcal{N}(0, \|t_i\|^2)$, we calculate $\|t_i\|$ as below.

$$\|(U\Sigma V^{\mathsf{T}})\Omega^{\mathsf{T}}(V\Sigma^{-1}U^{\mathsf{T}})\mathbf{e}_i\|_2 \leq \|\Omega^{\mathsf{T}}\|_2 \leq 1, \qquad \|(U\Sigma V^{\mathsf{T}})\Omega^{\mathsf{T}}(V\Sigma^{-2}V^{\mathsf{T}})\mathbf{v}\|_2 \leq \|\Omega^{\mathsf{T}}\|_2\|\Sigma^{-1}\|_2 \leq \frac{1}{\sigma_{\mathsf{min}}},$$

$$\|v^{\mathsf{T}}(\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}})\Omega(\tilde{V}\Lambda\tilde{U}^{\mathsf{T}} - \mathbf{v}\mathbf{e}_i^{\mathsf{T}})\|_2 \leq \|v^{\mathsf{T}}(\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}})\Omega\tilde{V}\Lambda\tilde{U}^{\mathsf{T}}\|_2 + \|v^{\mathsf{T}}(\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}})\Omega\mathbf{v}\mathbf{e}_i^{\mathsf{T}}\|_2 \leq \frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2},$$

$$\|e_i^{\mathsf{T}}(\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}})\Omega(\tilde{V}\Lambda\tilde{U}^{\mathsf{T}} - \mathbf{v}\mathbf{e}_i^{\mathsf{T}})\|_2 \leq \|\mathbf{e}_i^{\mathsf{T}}(\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}})\Omega(\tilde{V}\Lambda\tilde{U}^{\mathsf{T}}\|_2 + \|\mathbf{e}_i^{\mathsf{T}}(\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}})\Omega\mathbf{v}\mathbf{e}_i^{\mathsf{T}}\|_2 \leq 1 + \frac{1}{\sigma_{\mathsf{min}}},$$

where $\sigma_{\mathsf{min}} = \left(\frac{\sqrt{r\log(2/\delta)}\log(r/\delta)}{\varepsilon}\right)$. Using the concentration bound on the Gaussian distribution, each term, t_1, t_2, t_3, and t_4, is less than $\|t_i\|\ln(4/\delta_0)$ with probability $1-\delta_0/2$. From the fact that $2\left(\frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2}\right)\ln(4/\delta_0) \leq \varepsilon_0$, we have the second part of the proof.