Differentially Private Linear Algebra in the Streaming Model

Jalaj Upadhyay
Center for Applied Cryptographic Research
University of Waterloo.
jalaj.upadhyay@uwaterloo.ca

Abstract

The focus of this paper is on differential privacy of 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 restrict our attention to the event level privacy. We provide optimal, up to logarithmic factor, space differentially private mechanism in the streaming model for three basic linear algebraic tasks: matrix multiplication, linear regression, and low rank approximation, while incurring significantly less additive error.

Our approach for matrix multiplication and linear regression has 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 an input perturbation that raises the singular value 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.

1 Introduction

In the setting of large data analysis, one of the desired goals is to design a sub-linear space algorithm to perform computation while receiving the data online. There are many non-private algorithms in this setting. Such algorithms are called *online algorithms* or *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.* [21] 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 the sketch based approach 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 Blocki *et al.* [6, 7] and Upadhyay [59]). We give almost optimal (in terms of space required) 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, each of which is described by d features. For example, Hardt and Roth [34, 35] 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 [32, 35, 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 [3, 15, 26, 49, 55]. However, recent privacy violations in health-care and genetic studies 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.* [20] showed that it is impossible to obtain any non-trivial result with respect to the user level privacy when the data is streamed online. Therefore, in this paper, we restrict our attention to the event level 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 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 space lower bounds that we can use to make any sort of optimality comparison.

We reserve the letter n for number of rows and d for 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 memory used by the algorithm, and the total time taken by the algorithm. All our differentially private mechanisms for performing linear algebraic tasks are single-pass and achieve almost optimal space bound for one-pass algorithms. We do not make any assumption on the input matrix to compute low-rank approximation as in the previous works [23, 34, 35, 40]. For *bit complexity*, we use the convention used by Clarkson-Woodruff [15], i.e., the entries of an $n \times d$ matrix are $\kappa = \log(nd)$-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)-Matrix Multiplication).* An $n \times d$ matrix A and $d \times n$ matrix B are given. Output a matrix C so that $\|AB - C\|_F \leq \alpha \|A\|_F \cdot \|B\|_F + \tau$ with probability at least $1 - \beta$.

Theorem 1.1. *[15]* Suppose $n \geq c\kappa/\alpha^2$ for an absolute constant $c > 0$. Then any randomized one-pass algorithm which solves matrix multiplication with probability at least $4/5$ uses $\Omega(d\alpha^{-2}\kappa)$ bits of space.

Clarkson-Woodruff [15, Theorem 2.4] extended the idea of Sarlos [55] to show the following in the non-private setting.

Theorem 1.2. Given $\alpha, \beta > 0$, and conforming matrices A and B. The matrices are presented row-wise and column-wise, respectively, with integer entries having κ bits. There is a data structure so that, at a given time, AB can be estimated, so that with probability at least $1 - \beta$, the Frobenius norm of the error is at most $\alpha \|A\|_F \cdot \|B\|_F$. There is an $r = O(1/\alpha^2)$ so that for d large enough, the data structure requires $O(rd \log(1/\beta) \log \kappa + \log(1/\alpha))$ bits of space and $O(r)$ update time.

Theorem 5.1 gives a data-structure that uses $O\left(d\alpha^{-2}\kappa \log(1/\beta) \log n\right)$ bits of space and provides (ε, δ)-differential privacy with $\tau = O(\alpha\sqrt{n})$ and $O(r)$ update time. If we agree to pay for update time, then we can remove the extra $\log n$ factor in the space used and achieve the space bound of Theorem 1.2.

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 [16], supplies [46], and investment [25].

Problem 2. *((α, β, τ)-Linear Regression).* Given an $n \times d$ matrix A and a m set of $n \times 1$ column vectors $\{b_1, \cdots, b_m\}$, output a set of vectors $X = \{x_1, \cdots x_m\}$ so that $\|AX - B\|_F \leq (1+\alpha) \min_{Y \in \mathbb{R}^{d \times m}} \|AY - B\|_F + \tau$ with probability at least $1 - \beta$, where $B = \{b_1 | \cdots | b_m\}$.

Theorem 1.3. *[15]* Suppose $n \geq \kappa d/36\alpha$ and d is sufficiently large. Then any randomized one-pass algorithm which solves the *Linear Regression* problem with probability at least $7/9$ needs $(d^2\alpha^{-1}\kappa)$ bits of space.

Clarkson-Woodruff [15, Theorem 3.2] showed the following in the non-private setting.

Theorem 1.4. Given $\alpha, \beta > 0$, an $n \times d$ matrix A, and an n-vector b. There is a data structure so that, at a given time, can be used to estimate the linear regression problem. There is $r = O(d/\alpha)$ so that for d large enough, the data structure requires $O(rd \log(1/\beta) \log \kappa + \log(1/\alpha))$ bits of space and update time $O(d/\alpha)$.

Theorem 5.3 gives a data-structure that uses $O\left(d^2\alpha^{-1}\kappa \log(1/\beta) \log n\right)$ bits of space and provides (ε, δ)-differential privacy with $\tau = O(\alpha\sqrt{n})$ while maintaining an update time $O(d/\alpha)$. If we agree to pay for update time, then we can remove the extra $\log n$ factor in the space used and achieve the space bound of Theorem 1.4.

LOW-RANK APPROXIMATION. The last problem we consider is *low rank approximation* [17]. A partial list of its applications includes principal component analysis [36], fast multipole methods [28], and $\mathcal{H}$-matrices [27].

Problem 3. *((α, β, τ)-Low-rank approximation).* Given an $n \times d$ matrix A, a target rank k, and an oversampling parameter p, construct a matrix Ψ with $k + p$ 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 1.5. *[15]* 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 [15] also 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. They showed the following [15, Theorem 4.5] in the non-private setting.

Theorem 1.6. Suppose input A is given as a sequence of columns or rows. There is an $r = O(k \log(1/\beta)/\alpha)$, such that with probability at least $1 - \beta$, a matrix $\tilde{A}_k$ satisfying the condition of Problem 3 can be obtained under Frobenius norm. The space needed is $O(k\alpha^{-1}(n + d)\kappa \log(1/\beta))$

Unfortunately, we do not know a way to prove differential privacy using their algorithm, except to use additive noise mechanisms, much like Hardt and Roth [34]. 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 $O(k\alpha^{-1}(n + d)\kappa)$ bits data structure (Theorem 5.5) that can be used to publish k-rank approximation of an $n \times d$ input matrix in a single pass with (ε, δ)-differential privacy while incurring an additive error $\tau \leq O(k\sqrt{n \ln(2/\delta)}/\varepsilon)$ for the Frobenius and $\tau \leq O(\sqrt{nk \ln(2/\delta)}/\varepsilon)$ for the spectral norm. Our multiplicative approximation factor is almost optimal because of Eckart-Young bound [24] in the case of Frobenius norm and Mirsky [47] in the case of 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 space bound. This is not the first time that methods used in private setting has helped to improve the results in non-private setting. Dwork *et al.* [23] has recently showed one such case. The reason behind this is simple. The bound on the single pass algorithm of Clarkson and Woodruff [15] and the two-pass algorithm of Sarlos [55] use the earlier result for matrix-multiplication, and, therefore, had to rely on the number of rows in the projection matrix used to prove the bound in matrix multiplication. On the other hand, as we discuss later, we use perturbation theory along the line of Halko *et al.* [31].

We next compare our results with the earlier known results. Let $\lambda_1, \cdots, \lambda_{\mathsf{rank}(A)}$ be the singular values of a matrix A. In Table 1, we compare our result stated for low-rank approximation with the previous works. For effective comparison, since we do not make any coherence assumption, one should put $\mu = n$ in Table 1. A detail comparison is done in Section 5.3.3 taking into account the difference in privacy model and assumptions made. In Table 2, we give the best known lower bound for the space required for each of the problems stated above to compare the space optimality of our results. The various parameters used are as defined above.

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 [4]. This is because our mechanism uses operations that have efficient distributed algorithms. For example, one could use Jacobi method for

Method	Norm	Additive noise	Privacy Notion	Streaming	# Passes
Hardt-Roth [34]	F	$\frac{\sqrt{kn}\log(k/\delta)}{\varepsilon} + \sqrt{\frac{\mu\|A\|_F \log(k/\delta)}{\varepsilon}}$	Event level	No	2
Subspace Iteration [35]	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	$O(dk^3/(\varepsilon\gamma^2\delta^2))$	User level	No	k
Hardt [32]	S	$\frac{\lambda_1\sqrt{kn\mu\log(1/\delta)\log\left(\frac{n}{\gamma}\right)\log\log\left(\frac{n}{\gamma}\right)}}{\varepsilon\gamma^{1.5}\lambda_k}$	User level	No	$k\sqrt{\log\lambda}$
Dwork *et al.* [23]	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	$O(k\sqrt{n\ln(2/\delta)}/\varepsilon)$	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			Additive Noise
	Lower Bound [15]	Non-private	Private	
Linear Regression	$\Omega\left(d^2\alpha^{-1}\kappa\right)$	$O\left(d^2\alpha^{-1}\kappa\log\left(1/\beta\right)\right)$	$O\left(d^2\alpha^{-1}\kappa\log\left(1/\beta\right)\right)$	$O\left(\sqrt{n}\alpha\right)$
Matrix Product	$\Omega\left(d\alpha^{-2}\kappa/\right)$	$O(d\alpha^{-2}\kappa\log(1/\beta))$	$O(d\alpha^{-2}\kappa\log(1/\beta))$	$O\left(\sqrt{n}\alpha\right)$
Frobenius Low-rank	$\Omega(nk/\alpha)$	$O(k\alpha^{-2}(n+d)\kappa)$	$O(k\alpha^{-2}(n+d)\kappa)$	$O(k\sqrt{n\ln(2/\delta)}/\varepsilon)$
Spectral Low-rank	–	–	$O(k\alpha^{-2}(n+d)\kappa)$	$O(\sqrt{nk\ln(2/\delta)}/\varepsilon)$

Table 2: Optimality of our Results with Respect to the Best Known Space Bounds in Non-private Setting.

SVD [42, Chapter 4], Cannon's algorithm for multiplication [13], and GMRES for residual method [54].

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 matrix product and linear regression, to our knowledge, use Rademacher matrix [15] or tug-of-war matrices [55]. 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.* [6, 7] and Upadhyay [59]. We devise a private-sketch generation (PSG) mechanism to generate a private sketch of the private matrix. Our PSG uses random Gaussian matrix. We prove the privacy of PSG under certain 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). The algorithm for both matrix product and linear regression involves deterministic use of the sketch generated to give an approximate matrix product and linear regression. This gives the same bound on the space used by the data-structure as achieved by Clarkson-Woodruff [15].

We still have not reached the required update time. The standard Gaussian matrices takes $O(nr)$ time for update. This has $O(n)$ gap with the bound we wish to achieve. For this, we use a projection matrix that mimics the action of standard Gaussian matrix. For this, we use the projection matrix proposed by Upadhyay [60]. More concretely, we use the idea of Krahmer and Ward [43] to randomly permute the rows of the matrix WD using a permutation matrix, where W is a Walsh-Hadamard matrix and D is an appropriately chosen diagonal random matrix. However, we still do not achieve the required construction. For this, we use the spherical symmetry of Gaussian matrix, but our crucial observation here is that even a diagonal Gaussian

matrix suffices. An intuition why our construction works can be seen from the following observation: any given row of Φ' formed as above (i.e., composition of diagonal Gaussian matrix, a permutation matrix, and WD in the order from left to right matrix multiplication) is an i.d.d. Gaussian. This is because, for any row i, entry Φ'_{ij} consists of a sum of zero-mean independent Gaussian random variables and sign change retains the Gaussian property. Moreover, $\mathsf{Var}(\Phi') = n$ and D and the random permutation ensures that different entries in the row i have zero correlation. Since, the first and the second moment suffices for Gaussian distribution, we have that every row is i.d.d. Gaussian. Therefore, Φ' can be seen as "mimicking" the nature of random dense Gaussian matrices. We then sample r rows to form the final projection matrix. This sampling of r rows have been used in work related to compressed sensing, more specifically, to construct a distribution of matrices that satisfies the *Restricted Isometry Property* [11, 12]. Therefore, at the cost of gross oversimplification, an intuitive way to see our construction is as a hybrid of the known constructions of projection matrix with *Restricted Isometry Property* and known constructions of JL transform. As an added benefit, our construction requires only $2n$ random samples.

At a high level, our mechanisms for matrix product and linear regression are private analogues of Ref. [15] with the utility proof using the analogous result guaranteed by our projection matrix instead of the variance bounds for 4-wise independent *tug-of-war* matrices used by earlier works [15, 55] (note that differential privacy would not necessarily hold if the random matrix has dependent entries in its row). One approach could be to prove concentration bound on the corresponding distribution of product matrices. We use the approach used by Sarlos [55] and Clarkson-Woodruff [15] combined with our bound for projection matrices.

The mechanism for LRA is more complicated and markedly different from the recent works [23, 32, 34, 35, 40] (the online version of Dwork *et al.* [23] uses binary tree technique [20] and assumes a lower bound condition on the optimal value, see [23, Theorem 8]). 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 of algorithms to compute a LRA, i.e., first computes a projection matrix (*range finding* step) and then computes a low rank matrix by operating the projection matrix (*projection* step) on the input matrix. In the most naive form, both the steps require the input matrix. However, Ref. [15] showed that one can emulate the projection step by using the matrices formed in the range finding step, eliminating the need of input matrix in the projection step. Unfortunately, it only works for Rademacher matrices. On the other hand, our projection matrices are generated using a random Gaussian matrix.

The first key observation is that, by a clever use of linear algebra, information gathered in the range-finding step 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.* [6], is that the mechanism for PSG already gives considerable improvement in the range finding step. On top of that, we use an oversampling parameter p. This extra oversampling parameter helps us in getting much sharper bounds for both spectral as well as Frobenius norm.

We now mention the difference between our analysis and the analyses of Clarkson-Woodruff [15] and Sarlos [55]. The two results that used Johnson-Lindenstrauss matrices to give low rank approximation uses the trick that a good bound on matrix multiplication allows them to give a good bound on low-rank approximation. This only give them a bound when the approximation metric is stated in terms of Frobenius norm. Moreover, the algorithm of Sarlos [55] requires two-pass over the private input matrix. Our utility proof uses perturbation theory along the line of [31]. This 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.* [19]. 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.* [18] in a follow-up work. Since then, many mechanisms for preserving differential privacy have been proposed in the literature [9, 22, 29, 30, 41, 33, 34, 45, 53]. All these sanitizers have a common theme: they perturb the output before responding to queries. Blocki *et al.* [6, 7] 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.* [8] 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.* [23]. This was improved by Hardt and Roth [34] who studied the low rank approximation in Frobenius norm under the low coherence assumption. Kapralov and Talwar [40] and Chaudhary *et al.* [14] 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. Recently, Hardt and Roth [35] improved their noise bound by proposing robust private subspace iteration mechanism. All these works [14, 35, 40] uses some eigenvalue separation assumption, i.e., the top eigenvalue and the k-th eigenvalue has some separation, and Hardt and Roth [35] used the low-coherence assumption. Recently, Dwork *et al.* [23] revisited randomized mechanism to give a tighter bound. They also gave an online version of their mechanism under a normalized row assumption.

The literature of performing (non-private) linear algebra using streaming algorithms, started by Alon *et al.* [3], is so extensive that we cannot hope to cover it in any detail here. In the private setting, Dwork *et al.* [21] 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 [23, 37] for various tasks.

ORGANIZATION OF THE PAPER. In Section 2, we cover the basic definitions and notations. In Section 3, we give our basic mechanism that we use for intuition of the utility proof. We give the mechanism which achieves a better update time in Section 4 and the mechanisms for solving various linear algebra task in Section 5. We conclude the paper by stating some open problems in Section 6.

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, 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 [49], Flajolet-Martin [26], and Kane-Nelson [39]) and specifically pointing the improvements for utility bounds for linear regression through Kane and Nelson [39].

2 Notations and Basic Preliminaries

NOTATIONS. We reserve the letters A and B for private input matrices, Ω for Gaussian matrix (i.e., matrix whose entries are picked independently from a Gaussian distribution $\mathcal{N}(0,1)$). For an $n \times d$ matrix A, we let $A_{i:}$ 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 $e_1, \cdots, e_d$ denote the standard basis vectors in $\mathbb{R}^d$. We use the Dirac notation to denote vectors, i.e., $|\cdot\rangle$ for column-vector and $\langle\cdot|$ for row-vector. We use $\langle 0^n|$ to denote the transpose of an n-dimensional 0-vector. For a matrix M, we write $M \succ 0$ if all its eigenvalues are positive.

PRIVACY. We work with the relaxed notion of privacy, known as *approximate differential privacy*. We consider two data-sets D_1 and D_2 *neighbouring* if $\|D_1 - D_2\| \leq 1$. This notion was used in many of the earlier works [34, 35]

Definition 2.1. 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*.

We use the following in our analysis explicitly or implicitly.

Theorem 2.1. *(Composition Theorem [22]).* 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 2.2. 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.

In the course of this paper, we use few standard results from the theory of linear algebra, random matrices, and perturbation theory. We follow up with the key concepts and results we would need in our proofs.

LINEAR ALGEBRA. Our analysis makes extensive use of linear algebra and statistical properties of the 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 $rk(A)$ denote the rank of the matrix 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 norms. We use the notation $\|\cdot\|_F$ for Frobenius norm. the 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 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 2.3. If matrix A and B are conforming, then $\|AB\|_F \leq \|A\|_2\|B\|_F$.

Lemma 2.4. Let $A = (a_{ij})$ be a symmetric $n \times n$ matrix. If $|a_{ii}| \geq |a_{ij}|$ for all $1 \leq i, j \leq n$, then A is a positive semi-definite matrix.

Lemma 2.5. A matrix A has $\|A\|_2 \leq 1$ if and only if $\begin{pmatrix} \mathbb{I} & A \\ A^{\mathsf{T}} & \mathbb{I} \end{pmatrix}$ is positive semi-definite.

Lemma 2.6. Let $A = (a_{ij})$ be a symmetric $n \times n$ matrix. If $|a_{ii}| \geq |a_{ij}|$ for all $1 \leq i, j \leq n$, then A is a positive semi-definite matrix.

Lemma 2.7. Let A and B be Hermittian matrices with only $0, 1$ entries. Then $\text{Tr}(AB) \leq \text{Tr}(A)\text{Tr}(B)$. Moreover, if $\|A - B\| \leq 1$, then $\text{Tr}(A^\mathsf{T}A) - \text{Tr}(B^\mathsf{T}B) \leq 2$.

Lemma 2.8. 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. $\text{Tr}(CXB) = (\mathsf{vec}(C^\mathsf{T}))^\mathsf{T}(\mathbb{I}_q \otimes X)\mathsf{vec}(B)$,
3. $\text{Tr}(DX^\mathsf{T}EXB) = (\mathsf{vec}(X))^\mathsf{T}(D^\mathsf{T}B^\mathsf{T} \otimes E)(\mathsf{vec}(X)) = (\mathsf{vec}(X))^\mathsf{T}(BD \otimes E^\mathsf{T})\mathsf{vec}(X)$.

GAUSSIAN DISTRIBUTION. 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 $\mathsf{PDF}_\mathbf{X}(\mathbf{x}) := \frac{1}{\sqrt{(2\pi)^{\mathsf{rank}(\Sigma)}\tilde{\Delta}(\Sigma)}} \exp\left(-\frac{1}{2}\mathbf{x}^\mathsf{T}\Sigma^\dagger\mathbf{x}\right)$ with mean $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma = \mathbb{E}[(X - \mu)(X-\mu)^\mathsf{T}]$. It is easy to see from the description of the PDF that, in order to define the PDF corresponding to a multivariate Gaussian distribution, Σ has to have full rank. 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 $\mathbf{y} \sim \mathcal{N}(A\mu + \mathbf{b}, A\Sigma A^\mathsf{T})$.

RANDOM MATRIX THEORY. Our analysis for utility guarantee relies heavily on the theory of random matrices. We enumerate few of the key lemmata that we use in our analysis of low-rank approximation. For more details, we refer the readers to excellent books on random matrix theory [58] and multivariate statistics [48].

Lemma 2.9. Let Ω be a Gaussian matrix. Then for any fixed matrices A and B, $\mathbb{E}[\|A\Omega B\|_2] \leq \|A\|_2\|B\|_F + \|A\|_F\|B\|_2$.

Lemma 2.10. 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 2.11. Let Ω be a random $n \times (k+p)$ Gaussian matrix whose entries are picked from the distribution $\mathcal{N}(0, 1)$. Then $\mathbb{E}[\text{Tr}((\Omega^\mathsf{T}\Omega)^{-1})] = k/(p-1)$.

Theorem 2.12. *(Johnson-Lindenstrauss lemma)* Fix any $\alpha < 1/2$ and let m be a positive integer. Let Ω be a $k \times n$ matrix whose entries are 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 $\mathbb{R}^n$

$$\forall x \in S, \mathsf{Pr}_M\left[\|\Omega x\|_2 \in (1 \pm \alpha)\|x\|_2\right] \geq 2/3.$$

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. We review some of the basic probability theory used in this paper.

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 2.13. 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 2.13 is often used to control the expectation of the supremum of a finite family of exponentially integrable variables. Pisier [50] proved the following fundamental lemma.

Lemma 2.14. (Pisier [50]) Let $\{\zeta_f\}_{f\in F}$ be a finite family of random variables and ψ be as in Lemma 2.13. 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 2.15. 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 [44]. In this paper, we use the following result by Birge and Massart [5] for this purpose.

Theorem 2.16. (Birge-Massart [5]) 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 2.15 and Talagrand inequality, the authors also proved the following corollary to Theorem 2.16.

Corollary 2.17. (Birge-Massart [5]) 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 $\mathsf{Pr}\left[\zeta \geq \sqrt{2a\tau} + b\tau\right] \leq \exp(-\tau)$.

3 Space Efficient Differentially Private Sketch Generation

We study differential privacy in the well known streaming model of computation [3]. We present it at the level required to understand this paper. A more formal definition appears in Alon *et al.* [3]. It has three entities: a stream generator (database owner) $\mathcal{S}$, a (database) curator $\mathcal{K}$, and a query maker $\mathcal{Q}$. $\mathcal{S}$ starts the process at time $t = 0$. The curator initializes its data structure to $\mathcal{D}_0$. From $t = 0$, the curator is allowed only one-pass over the input matrix, i.e., it can copy any entry of the data-base to its memory space during exactly one time epoch. It updates its data structure to $\mathcal{D}_t$ using $\mathcal{D}_{t-1}$ and newly accessed data-points of the matrix. At certain time, $t = t_q$, the query maker $\mathcal{Q}$ generates a query function q. The curator responds with the response, $q(\mathcal{D}_{t_q})$.

On input a streamed vector v, flag, parameters ε, δ, and an $r \times O(n)$ random Gaussian matrix Ω, the mechanisms does the following:

Variant 1: If flag$= 0$, compute $Y_v = \langle v|\Omega^{\mathsf{T}}$; else compute $Y_v = \Omega|v\rangle$. In the end, return Y_v.

Variant 2: If flag$= 0$, compute $Y_v = \langle v|\Omega^{\mathsf{T}}\Omega$; else compute $Y_v = \Omega^{\mathsf{T}}\Omega|v\rangle$. In the end, return Y_v.

Figure 1: Private Sketch Generation (PSG) Algorithm

The streaming model has a resource bound on the curator. A curator is only allowed to use time *polynomial in the size of the data base* to construct the data structure, and memory *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 2.1 with the two neighbouring streams differing in at most one entry.*

In all our mechanisms, the curator uses sketch of the private matrix as the data-structure. The sketch of the matrix is generated row (or column) wise using one of the variants presented in Figure 1.

The first variant has some resemblance to the mechanism of Blocki *et al.* [6] and Upadhyay [59] in the sense that we multiply an appropriate Gaussian matrix, while the second variant can be seen as its extension in the sense that two successive application of Gaussian matrix in a defined form also preserve privacy. However, the analogy ends here. For example, [6, 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 value before invoking PSG (see Section 5). As argued by Blocki *et al.* [6], their mechanism does not give a guarantee that 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. First of all, their mechanism is not a single pass 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 [52, 57]: first to subtract the mean of the entries of the matrix and second to compute the projection on the altered SVD). 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. Secondly, they project the entries of the columns of the private matrix to a higher dimensional space; here, we perform embedding to a lower dimensional subspace in the similar vein as other applications of JL-transform.

Theorem 3.1. If the singular values of the streamed matrix to the first variant of the PSG algorithm are at least $\sigma_{\mathsf{min}} := \frac{4\sqrt{r\log(2/\delta)\log(r/\delta)}}{\varepsilon}$ and for second variant are at least $\sigma_{\mathsf{min}} := \frac{4r\log(r/\delta)}{\varepsilon}$. Then PSG using $r \times O(n)$ Gaussian matrix preserves (ε, δ)-differential privacy.

The proof of the first variant follows the idea of Blocki *et al.* [6] 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 $\Phi = \sum_{i=1}^{r}|a_i\rangle\langle a_i|$ for n-variate Gaussians, $a_1, \cdots, a_r$. This is the technical part of the proof; rest of the proof follows by evaluating the pdf for neighbouring matrices. We give a detail proof for variant 1 in Appendix B.

Proof. We now prove that the second variant preserves privacy if the singular values of the streamed matrix follows the hypothesis of the theorem. 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 just like the transition from identity to arbitrary positive definite covariance matrices in the multivariate Gaussian distribution. Let $\alpha_1, \cdots, \alpha_r$ be i.i.d. $\mathcal{N}(0, \mathbb{I})$ be r multivariate Gaussian distribution,

i.e., $\alpha_{ij} \sim \mathcal{N}(0,1)$ for $1 \leq i \leq r, 1 \leq j \leq n$. The distribution we are interested in is $\Phi = \sum_{i=1}^{r} |\alpha_i\rangle\langle\alpha_i|$. 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 covariance matrix of the random variables is $\mathbb{I}$.

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

$$\begin{aligned}\mathsf{PDF}(\Phi; \mathbb{I}) = \mathsf{PDF}(\langle\alpha_1, \alpha_1\rangle; \mathbb{I})\mathsf{PDF}((\langle\alpha_2, \alpha_1\rangle, \langle\alpha_2, \alpha_2\rangle)|\langle\alpha_1, \alpha_1\rangle; \mathbb{I}) \cdots \\ \mathsf{PDF}((\langle\alpha_r, \alpha_1\rangle, \cdots \langle\alpha_r, \alpha_r\rangle)|\Phi_{[r-1]}; \mathbb{I}). \end{aligned} \tag{2}$$

There are $r(r+1)/2$ distinct entries, $\langle\alpha_1, \alpha_1\rangle, (\langle\alpha_2, \alpha_1\rangle, \langle\alpha_2, \alpha_2\rangle), \cdots, (\langle\alpha_r, \alpha_1\rangle, \cdots \langle\alpha_r, \alpha_r\rangle)$. 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 $(\langle\alpha_i, \alpha_1\rangle, \cdots, \langle\alpha_i, \alpha_{i-1}\rangle)$. Let $\beta_{i-1}^{\mathsf{T}} = (\langle\alpha_i, \alpha_1\rangle, \cdots, \langle\alpha_i, \alpha_{i-1}\rangle)$. Then we use the fact that there is a transformation of Jacobian one from $\left(\langle\alpha_i, \alpha_1\rangle, \cdots, \langle\alpha_i, \alpha_{(i-1)}\rangle, \langle\alpha_i, \alpha_i\rangle - \beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\beta_{i-1}\right)$ to $\left(\langle\alpha_i, \alpha_1\rangle, \cdots, \langle\alpha_i, \alpha_{(i)}\rangle\right)$ to compute each term in the chain rule [48, 51].

Distribution of β_{i-1}. We first prove that β_{i-1} is $(i-1)$ variate Gaussian distribution. Since $\Sigma = \mathbb{I}$, and $\alpha_{11}, \cdots, \alpha_{1n}, \cdots, \alpha_{r1}, \cdots, \alpha_{rn}$ are i.i.d. $\mathcal{N}(0,1)$, from the elementary property of linear functions of normal variables, conditional on α_{kj} for $1 \leq k \leq i-1$ and $1 \leq j \leq n$, β_{i-1} is $(i-1)$ variate Gaussian distribution with

$$\Phi_{[i]} = \begin{pmatrix} \langle\alpha_1, \alpha_1\rangle & \cdots & \langle\alpha_1, \alpha_i\rangle \\ \vdots & \ddots & \vdots \\ \langle\alpha_i, \alpha_1\rangle & \cdots & \langle\alpha_i, \alpha_i\rangle \end{pmatrix}$$

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

$$\mathsf{COV}(\beta_{i-1}, \alpha_{ij}) = \left[\mathsf{COV}\left(\langle\alpha_i, \alpha_1\rangle\alpha_{ij}\right), \cdots, \mathsf{COV}\left(\langle\alpha_i, \alpha_{(i-1)}\rangle\alpha_{ij}\right)\right] = (\alpha_{1j}, \cdots, \alpha_{i-1,j})^{\mathsf{T}}$$

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

$$\mathsf{COV}\left[\beta_{i-1}, \alpha_{ij} - \beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\alpha_1 j, \cdots \alpha_{i-1,j})^{\mathsf{T}}|\alpha_{kj}\right] = 0 \qquad \forall 1 \leq k \leq i-1, \tag{3}$$

as the left hand side equals $(\alpha_1 j, \cdots \alpha_{i-1,j})^{\mathsf{T}} - \Phi_{[i-1]}\Phi_{[i-1]}^{-1}(\alpha_1 j, \cdots \alpha_{i-1,j})^{\mathsf{T}}$.

Therefore, β_{i-1} is independent of $\sum_{j=1}^{k}\left(\alpha_{ij} - (\alpha_{1j}, \cdots, \alpha_{i-1,j})^{\mathsf{T}}\Phi_{[i-1]}^{-1}(\alpha_{1j,\cdots,\alpha_{i-1,j}})\right)^2$. Rao [51] proved that

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

the standard χ^2 distribution.

Computing every term in the chain rule. From the fact that β_{i-1} is a $(i-1)$-variate Gaussian distribution, equation (3), equation (4), and the identity

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

we first calculate the joint pdf of

$$\begin{aligned}
&\Big(\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i,\alpha_{(i-1)}\rangle,\langle\alpha_i,\alpha_i\rangle-\langle\beta_{i-1}|\Phi_{[i-1]}^{-1}|\beta_{i-1}\rangle\Big)^{\mathsf{T}}\\
&\quad=\frac{\exp\left(-\frac{1}{2}\left(\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i,\alpha_{(i-1)}\rangle\right)^{\mathsf{T}}\Phi_{[i-1]}^{-1}\left(\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i\alpha_{(i-1)}\rangle\right)\right)}{(2\pi)^{(i-1)/2}\Delta(\Phi_{[i-1}]^{1/2}}\\
&\qquad\times\frac{\exp\left(-\frac{\langle\alpha_i,\alpha_i\rangle-\beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\beta_{i-1}}{2}\right)\left(\langle\alpha_i,\alpha_i\rangle-\beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\beta_{i-1}\right)^{(n-i+1)/2-1}}{2^{(n-i+1)/2}\Gamma((n-i+1)/2)}\\
&\quad=\frac{\exp\left(-\frac{\beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\beta_{i-1}}{2}\right)\Delta(\Phi_i)^{(n-i-1)/2}}{2^{m/2}\pi^{(i-1)/2}\Gamma((n-i+1)/2)\Delta(\Phi_{[i-1]})^{(n-i)/2}}\\
&\quad=\mathsf{PDF}((\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i,\alpha_{i-1}\rangle,\langle\alpha_i,\alpha_i\rangle)|\Phi_{[i-1]}),
\end{aligned}\tag{5}$$

where the last step used the fact that there is a one-to-one transformation of Jacobian 1 to $\left(\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i,\alpha_{(i)}\rangle\right)$ from $\left(\langle\alpha_i,\alpha_1\rangle,\cdots,\langle\alpha_i,\alpha_{(i-1)}\rangle,\langle\alpha_i,\alpha_i\rangle-\beta_{i-1}^{\mathsf{T}}\Phi_{[i-1]}^{-1}\beta_{i-1}\right)$.

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 [51]) 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 $\alpha_1,\cdots,\alpha_r$, i.e., $\sum_{i=1}^r A|\alpha_i\rangle\langle\alpha_i|$, $\sum_{i=1}^r|\alpha_i\rangle\langle\alpha_i|A^{\mathsf{T}}$, and $\sum_{i=1}^r|\alpha_i\rangle A^{\mathsf{T}}\langle\alpha_i|$ have the same distribution. More concretely, for $\sum_{i=1}^r A|\alpha_i\rangle\langle\alpha_i|$, the distribution is

$$\mathsf{PDF}(\Phi;A^{\mathsf{T}}A)=\mathsf{PDF}(\Phi;\mathbb{I})\frac{\mathsf{PDF}(\langle\alpha_1,\cdots,\alpha_r\rangle;A^{\mathsf{T}}A)}{\mathsf{PDF}(\langle\alpha_1,\cdots,\alpha_r\rangle;I)}=\frac{\exp(-\mathrm{Tr}((A^{\mathsf{T}}A)^{-1}\Phi)/2)\Delta(\Phi)^{(n-r-1)/2}}{2^{rn/2}\pi^{r(r-1)/4}\Delta(A^{\mathsf{T}}A)^{r/2}\prod_{i=1}^r\Gamma((n-i+1)/2)}.$$

We can now prove the privacy guarantee. Let $\delta_0=\delta/r$. Let A and $\tilde{A}$ be the matrix such that $A-\tilde{A}=E=ve^{\mathsf{T}}$. 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.* [6], it is straightforward to see that combination of the following proves differential privacy of the published matrix:

$$\sqrt{\frac{\tilde{\Delta}(A^{\mathsf{T}}A)}{\tilde{\Delta}(\bar{A}^{\mathsf{T}}\bar{A})}}\in\exp(\pm\varepsilon/r)\quad\text{and}\quad\Pr\left[|\mathrm{Tr}\left(\left((A^{\mathsf{T}}A)^{-1}-(\tilde{A}^{\mathsf{T}}\tilde{A})^{-1}\right)\Phi\right)|\le\varepsilon\right]\ge1-\delta.$$

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.* [6]. For the second bound required for the privacy, we first bound the following

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

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 $E = ve_i^\mathsf{T}$ for some i, we can further solve the above expression.

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

Fix a j. We bound the following.

$$|\text{Tr}\left(\langle\alpha_j|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\alpha_j\rangle\right) + \text{Tr}\left(\langle\alpha_j|V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\alpha_j\rangle\right)|. \tag{6}$$

We now look at each term in the above expression. $\langle\alpha_j|V\Sigma^{-2}V^\mathsf{T}|v\rangle$ is distributed as $\mathcal{N}(0, \|V\Sigma^{-2}V^\mathsf{T}|v\rangle\|^2)$, $\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}|\alpha_j\rangle$ as $\mathcal{N}(0, \|\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}\|^2)$, $\langle\alpha_j|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle$ as $\mathcal{N}(0, \|V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\|^2)$, and $\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}|\alpha_j\rangle$ 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 $1/\sigma_{\mathsf{min}}, 1/\sigma_{\mathsf{min}} + 1/\sigma_{\mathsf{min}}^2, 1$, and $1 + 1/\sigma_{\mathsf{min}}$, respectively.

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

$$\mathsf{Pr}\left[(6) \leq 2\left(\frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2}\right)\ln(4/\delta_0) \leq \varepsilon\right] \geq 1 - \delta_0.$$

Taking union bound, we have with probability $1 - \delta$, $-\varepsilon \leq \text{Tr}\left(\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)\Phi\right) \leq \varepsilon$. Adjusting and renaming the value of ε, we have the result. □

4 Update Time Efficient Private Sketch Generation

In this section, we use the projection matrix of Upadhyay [60] to get the improvement over naive private sketch generation mechanism as far as update time is concerned. More concretely, the projection matrix used in Figure 1 would be a product of a smoothening matrix and a projection matrix. We use the smoothening matrix $W^{(n)}D$ as used in many of the earlier works, where D is an $n \times n$ diagonal matrix with $d_{ii} \sim \mathcal{N}(0, 1)$. Our construction differs in the manner by which we construct the projection matrix. We define it in more detail next. Let $\alpha := \langle\alpha_1, \cdots, \alpha_n\rangle$ be n i.d.d. Gaussian samples and $M = \mathsf{Diag}(\alpha)$. Let Π and Π' be random permutation matrices. Then P is the matrix $\Pi_{1..r}M\Pi'$. The matrix P alone cannot be used for

the random projection because, for some bad input vector $\mathbf{x}$, the estimate of $\|P\mathbf{x}\|_2$ can be really bad. For example, when $\mathbf{x}$ is along a single coordinate, then only the non-zero values of P along this coordinate will contribute to $P\mathbf{x}$, giving a very bad variance bound. For this, we need to precondition the input with $W^{(n)}D$. In non-technical terms, along with permutation matrices, it allows us to mimic a projection matrix with every entries picked i.d.d. The final projection matrix is $\Omega = \frac{1}{\sqrt{r}}PW^{(n)}D$ and the redefined mechanism for private sketch generation uses $\frac{1}{\sqrt{r}}PW^{(n)}D$ instead of Ω in Figure 1. The theorems and their proof appears in Upadhyay [60]. We state them here for the sake of completion.

Theorem 4.1. Let $\varepsilon \in (0,1)$ be a constant. Let Ω be an $r \times n$ matrix as constructed above, where $r = c\varepsilon^{-2}\log m \log n$ for a large global constant c. Then for any set $\mathcal{S}$ of m vectors $\mathbf{x} \in \mathbb{R}^n$, the following holds with probability at least $2/3$,

$$(1-\varepsilon)\|\mathbf{x}\|_2^2 \leq \|\Omega\mathbf{x}\|_2^2 \leq (1+\varepsilon)\|\mathbf{x}\|_2^2. \tag{7}$$

Proof. The usual idea in proving the concentration bound of the above form is to first bound the expectation of the random variables corresponding to the output of an application of the transformation matrix, and then use the standard concentration bound to prove the result. We use the same idea.

We break the analysis in two parts. We first prove that $W^{(n)}D$ is an isometry using the idea of Ailon and Chazelle [1], i.e., for any vector $\mathbf{x}$ of unit length, $W^{(n)}D\mathbf{x}$ has bounded co-ordinates. Then, we use this promise to prove the following: when we multiply a diagonal Gaussian matrix from the right to this smoothen vector and sample r rows, then this preserves the Euclidean norm with high probability. We are done because every entries of M is picked independently and multiplying Π with $\tilde{\mathbf{x}}$ only permutes the co-ordinates of $\tilde{\mathbf{x}}$.

Since the transformation is linear, without loss of generality, we can assume $\mathbf{x}$ is a unit vector. Fix a $\mathbf{x} \in \mathcal{S}$. The first step is proving the equivalent isometry bound of Ailon and Chazelle [1], i.e.,

Theorem 4.2. Let $\mathbf{x} \in \mathbb{R}^n$. Let $W^{(n)}$ be Walsh-Hadamard matrix and D be a diagonal Gaussian matrix with each entries picked i.d.d. Then, for any $\eta > 0$, we have $\mathsf{Pr}\left[\|W^{(n)}D\mathbf{x}\|_\infty \geq \sqrt{2/n}\log(2n/\eta)\,\|\mathbf{x}\|_2\right] \leq \eta$.

Note that in case of Ailon-Chazelle [1], D is a diagonal matrix formed by Rademacher sequence. Intuitively, the above theorem holds because these sequence in the limiting case is a normal distribution.

Proof. Fix a row j of the vector $\tilde{\mathbf{x}} = W^{(n)}D\mathbf{x}$ where D is diagonal Gaussian matrix. Note that $\tilde{x}_j = \sum_i \beta_{ji}x_i$ for $\beta_{ji} \sim \mathcal{N}(0,1/n)$. The latter is because sign change preserve the Gaussian distribution (there is a positive or negative sign due to Hadamard transform, but we can neglect it without effecting the bound– it only changes whether β_{ji} is picked from positive or negative side of the 0-mean Gaussian). Now, note that since every entry of D is picked from $\mathcal{N}(0,1/n)$ if we move the normalization constant to D, we have the Cramer transform $\psi_{\zeta_i}(\lambda) = \lambda^2/2nx_i^2$ for $\zeta_i = \beta_{ji}x_i$. This implies that the inverse Cramer transform is $\psi^*_{\zeta_i}(t) = t^2nx_i^2/2$.

The following set of equations are now immediate.

$$\begin{aligned}
\mathsf{Pr}[|\tilde{x}_j| \geq t] &= 2\mathsf{Pr}[\tilde{x}_j \geq t] = 2\mathsf{Pr}[\langle\beta_j,\mathbf{x}\rangle \geq t] \\
&= 2\mathsf{Pr}[\exp(tn\langle\beta_j,\mathbf{x}\rangle > \exp(t^2n))] \\
&= 2\prod_i \mathsf{Pr}\left[\exp\left(tn\beta_{ji}x_i\right) > \exp(t^2n)\right] \\
&\leq 2\exp\left(\|\mathbf{x}\|_2^2t^2n/2\right).
\end{aligned}$$

Now noting that the transform is linear; therefore, without any loss of generality, we can assume that $\mathbf{x}$ is a unit vector, we see that this is exactly the isometry bound of Ailon and Chazelle [1]. The result thus follows.

□

Step 2 (Bounding the expectation). The second step is to use the guarantee that $\|\tilde{\mathbf{x}}\|_\infty = O(n^{-1/2}\sqrt{\log m})$ to get the desired expectation bound. The naive method to work with the permutation in the matrix Ω to get the concentration result makes the proof very lengthy. We follow the approach of Vybiral [61]. 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 $\mathbf{x} \in \mathbb{R}^n$,

$$\|\Omega\mathbf{x}\|_2 = \|P\tilde{\mathbf{x}}\|_2 = \|Z\alpha\|_2,$$

with entries of Z are $z_{i,j} = (\Pi_{1..r})_{i:}(\mathsf{Diag}(\Pi'\tilde{\mathbf{x}}))_{:j}$.

Let $U\Sigma V^{\mathsf{T}}$ be the SVD of Z, $\gamma = V^{\mathsf{T}}\alpha$. Let $\sigma := \langle \sigma_1, \cdots, \sigma_r \rangle$ be the singular values of Z and $\langle \gamma_1, \cdots, \gamma_r \rangle$ be the co-ordinates of $V^{\mathsf{T}}\alpha$. Making this substitution, we have the following equalities.

$$\begin{aligned}
\mathsf{Pr}_\alpha\left[\|\Omega\mathbf{x}\|_2^2 \geq (1+\varepsilon)\right] &= \mathsf{Pr}_\alpha\left[\|WZ\alpha\|_2^2 \geq (1+\varepsilon)r\right] = \mathsf{Pr}_\alpha\left[\|Z\alpha\|_2^2 \geq (1+\varepsilon)r\right] \\
&= \mathsf{Pr}_\alpha\left[\|U\Sigma V^{\mathsf{T}}\alpha\|_2^2 \geq (1+\varepsilon)r\right] = \mathsf{Pr}_\gamma\left[\|U\Sigma\gamma\|_2^2 \geq (1+\varepsilon)r\right] \\
&= \mathsf{Pr}_\gamma\left[\|\Sigma\gamma\|_2^2 \geq (1+\varepsilon)r\right]. \qquad (8)
\end{aligned}$$

In other words, if we can prove the concentration bound on $\sum \sigma_i^2|\gamma_i|^2$, we are done. We use Theorem 2.16, which requires the function ψ corresponding to our case. For this, we work in two step process. Let $0 < \lambda < 1/2a$. We first give the following simple proposition.

Proposition 4.3. Let $Y \sim \mathcal{N}(0,1)$ and $S := \log(\mathbb{E}[\exp(a\lambda(Y^2-1))])$. Then $S \leq \frac{a^2\lambda^2}{1-2a\lambda}$.

Proof. $S := \log(\mathbb{E}_Y[\exp(\lambda a(Y^2-1))])$, where $Y \sim \mathcal{N}(0,1)$. A simple calculation shows that when $Y \sim \mathcal{N}(0,1)$, then

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

□

Let $Y_1, \cdots, Y_r$ be random variables picked using the distribution $\mathcal{N}(0,1)$. From the linearity of expectation, a simple extension of Proposition 4.3 to a vector of Gaussian variables results in the following set of inequalities.

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

Therefore, we have the following bound.

Proposition 4.4. Let $Y_1, \cdots Y_r$ be picked from $\mathcal{N}(0,1)$ and $\sigma = \langle \sigma_1, \cdots, \sigma_r \rangle$ be an r dimensional vector. Let λ be an arbitrary constant such that $0 < \lambda < 1/2\|\sigma\|_\infty$. Then

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

Since $\Sigma = \mathsf{Diag}(\sigma_1, \cdots, \sigma_r)$ and $Z = U\Sigma V^T$. Then we can state Proposition 4.4 in the following equivalent form

$$\sum_{i=1}^{r} \log\left(\mathbb{E}_{\gamma_i}\left[\exp(\lambda\sigma_i^2(\gamma_i^2-1))\right]\right) \leq \frac{\lambda^2\|Z\|_2^4}{1-2\|Z\|_\infty^2\lambda} = \frac{2\lambda^2\|Z\|_2^4}{2(1-2\|Z\|_\infty^2\lambda)}.$$

The right hand side has the form $\psi(u) = \frac{a\lambda^2}{2(1-b\lambda)}$ for $a = 2\|Z\|_2^4$ and $b = 2\|Z\|_\infty^2$. Using Corollary 2.17, we have

$$\mathsf{Pr}_\gamma \left[\sum_{i=1}^{r} \sigma_i^2 \left(\gamma_i^2 - 1\right) \geq 2\|Z\|_\infty^2 \tau + 2\|Z\|_2^2 \sqrt{\tau} \right] \leq \exp(-\tau). \tag{9}$$

We need to estimate $\|Z\|_\infty$ and $\|Z\|_2$. This is where the guarantee on $\|\tilde{\mathbf{x}}\|_\infty$ is useful. Using Theorem 4.2, with probability $19/20$, we have

$$\max \|WD\mathbf{x}\|_\infty = \sqrt{\frac{2}{n} \log (40n)}.$$

Consequently, from the symmetry of the matrices, we have

$$\|Z\|_\infty^2 = \max_{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\|_2 = 1} \|Z\mathbf{x}\|_2^2 \leq n\|WD\mathbf{x}\|_\infty^2 = n\|\tilde{\mathbf{x}}\|_\infty^2 = 2\log (40n). \tag{10}$$

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

$$\|Z\|_2^2 \leq \|Z\|_F \cdot \|Z\|_\infty = \sqrt{2r \log (40n)}. \tag{11}$$

Since $\sum_j \sigma_j^2 = r$, by setting $\tau = cr\varepsilon^2 / \log(40n)$ for a small constant c, and using equations (8), (9), (10), and (11), we have

$$\mathsf{Pr}_\alpha[\|\Omega\mathbf{x}\|_2^2 \geq (1+\varepsilon)] = \mathsf{Pr}_\gamma \left[\sum_{j=1}^{r} \sigma_j^2 (|\gamma_j| - 1) \geq \varepsilon r \right] < \exp\left(-\frac{r\varepsilon^2}{\log(2n/\eta)} \right). \tag{12}$$

For $(12) < 1/6m$, 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

$$\mathsf{Pr}_\alpha[\|\Omega\mathbf{x}\|_2^2 \leq (1-\varepsilon)] = \mathsf{Pr}_\gamma \left[\sum_{j=1}^{r} \sigma_j^2 (|\gamma_j| - 1) \leq -\varepsilon r \right] < \frac{1}{6m}. \tag{13}$$

Combining equation (12) and equation (13) and union bound over all $x \in \mathcal{S}$, the result follows. □

Theorem 4.5. If the singular values of the input matrix A is at least $\sqrt{\frac{16n \log(2/\delta)}{\varepsilon}}$, then publishing $A^\mathsf{T}\Omega$, where $\Omega = PWD$ with P and W as in Section 4 and D is a diagonal Gaussian matrix, is (ε, δ)-differentially private.

Proof. We first show that if the PDF is proportional to

$$\frac{\exp\left(-\frac{1}{2}\mathrm{Tr}\left((A^\mathsf{T}A)^{-1}X^\mathsf{T}(A^\mathsf{T}A)^{-1}X\right)\right)}{\det(A^\mathsf{T}A)^{2n}},$$

then $A^\mathsf{T}\Omega$ preserves differential privacy. We later prove that the distribution of $A^\mathsf{T}\Omega$ has the required form. Let $\delta_0 = \delta/2n$ and $\varepsilon_0 = \varepsilon/4n$. Let A and $\tilde{A}$ be the matrix such that $A - \tilde{A} = E = ve^\mathsf{T}$. The published matrices corresponding to the two neighboring matrices have the following probability density function

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

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

where C is the normalization factor that depends on σ and π. As in Blocki *et al.* [6], we break down our proof to two parts, which are stated succinctly in the form of following lemma.

Lemma 4.6. For a matrix A with all singular values greater than $\sqrt{\frac{32r\log(4/\delta)}{\varepsilon}}$, the following holds

$$\frac{\det(A^\mathsf{T}A)}{\det(\tilde{A}^\mathsf{T}\tilde{A})} \in \exp(\pm\varepsilon/2n). \tag{14}$$

In addition, if $X = A^\mathsf{T}\Omega$, then

$$\mathsf{Pr}\left[|\mathrm{Tr}\left((A^\mathsf{T}A)^{-1}X^\mathsf{T}(A^\mathsf{T}A)^{-1}X - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right)| \leq \varepsilon\right] \geq 1-\delta. \tag{15}$$

Proof. It is straightforward to see that combination of two statements in above lemma proves differential privacy of the published matrix. We next prove the lemma.

Proof of equation (21). The first part of the proof follows simply as in Blocki *et al.* [6]. 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 $\bar{A}$. 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,

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

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

Proof of equation (15). For the second part, we first note that $A^\mathsf{T}A$, $\tilde{A}^\mathsf{T}\tilde{A}$, and $X^\mathsf{T}(A^\mathsf{T}A)^{-1}X - X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X$ are all Hermitian. From Lemma 2.7, we have

$$\begin{aligned}
&\mathrm{Tr}\left((A^\mathsf{T}A)^{-1}X^\mathsf{T}(A^\mathsf{T}A)^{-1}X - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right) \\
&\leq \mathrm{Tr}\left((A^\mathsf{T}A)^{-1}\right)\left(\mathrm{Tr}\left(X^\mathsf{T}(A^\mathsf{T}A)^{-1}X - X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right)\right) + \frac{\mathrm{Tr}((A^\mathsf{T}A)^{-1})\mathrm{Tr}\left(X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right)}{2} \\
&\leq \frac{1}{\sigma_{\mathsf{min}}^2}\left(\mathrm{Tr}\left(X^\mathsf{T}(A^\mathsf{T}A)^{-1}X - X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right)\right) + \frac{1}{2\sigma_{\mathsf{min}}^2}\mathrm{Tr}\left(X^\mathsf{T}(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X\right).
\end{aligned} \tag{16}$$

The second inequality follows because $\mathrm{Tr}(\tilde{A}^\mathsf{T}\tilde{A}) \leq \mathrm{Tr}(A^\mathsf{T}A) + 2$, which implies that $-\mathrm{Tr}(\tilde{A}^\mathsf{T}\tilde{A})^{-1} \leq -1/(\mathrm{Tr}(A^\mathsf{T}A) + 2)$. Since $\mathrm{Tr}(A^\mathsf{T}A) \geq 2$ for all $n \geq 2$, we have $-\mathrm{Tr}(\tilde{A}^\mathsf{T}\tilde{A})^{-1} \leq -1/2\mathrm{Tr}(A^\mathsf{T}A)$.

Therefore, if we prove that the absolute value of equation (16) is bounded by ε with probability $1-\delta$, we are done. We first bound the first part of the term as follows.

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

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, and assuming we sample $X = A^\mathsf{T}\Omega$, we can further solve the above expression.

$$\begin{aligned}
X^\mathsf{T}\left((A^\mathsf{T}A)^{-1} - (\tilde{A}^\mathsf{T}\tilde{A})^{-1}\right)X &= X^\mathsf{T}((A^\mathsf{T}A)^{-1}(A^\mathsf{T}E + E^\mathsf{T}\tilde{A})(\tilde{A}^\mathsf{T}\tilde{A})^{-1}X) \\
&= \Omega^\mathsf{T}\left(V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T} + V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}\right)\Omega \\
&= \Omega^\mathsf{T}\underbrace{V\Sigma^{-1}U^\mathsf{T}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^\mathsf{T}}_{S_1}\Omega + \Omega^\mathsf{T}\underbrace{V\Sigma^{-2}V^\mathsf{T}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^\mathsf{T}}_{S_2}\Omega.
\end{aligned}$$

We need to bound the absolute value of the trace of $S = \left(\Omega^{\mathsf{T}} S_1 \Omega + \Omega^{\mathsf{T}} S_2 \Omega\right)$. Now, $\Omega = |\alpha\rangle\langle\beta|$; therefore,

$$\begin{aligned}
\mathrm{Tr}(S) &= \mathrm{Tr}((|\alpha\rangle\langle\beta|)^{\mathsf{T}} V\Sigma^{-1}U^{\mathsf{T}}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}}|\alpha\rangle\langle\beta| + (|\alpha\rangle\langle\beta|)^{\mathsf{T}} V\Sigma^{-2}V^{\mathsf{T}}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}}|\alpha\rangle\langle\beta|) \\
&= \mathrm{Tr}(|\beta\rangle\langle\alpha|V\Sigma^{-1}U^{\mathsf{T}}|e_i\rangle\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}}|\alpha\rangle\langle\beta| + |\beta\rangle\langle\alpha|V\Sigma^{-2}V^{\mathsf{T}}|v\rangle\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}}|\alpha\rangle\langle\beta|) \\
&= \mathrm{Tr}(\langle\beta,\beta\rangle(\underbrace{\langle\alpha|V\Sigma^{-1}U^{\mathsf{T}}|e_i\rangle}_{B_1}\underbrace{\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}}|\alpha\rangle}_{B_2} + \underbrace{\langle\alpha|V\Sigma^{-2}V^{\mathsf{T}}|v\rangle}_{B_3}\underbrace{\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}}|\alpha\rangle}_{B_4})).
\end{aligned}$$

In other words, we have to bound

$$|\mathrm{Tr}(S)| = |\mathrm{Tr}\left(\langle\beta,\beta\rangle\left(B_1B_2 + B_3B_4\right)\right)| \leq |\mathrm{Tr}(\langle\beta,\beta\rangle)\left(\mathrm{Tr}(B_1)\mathrm{Tr}(B_2) + \mathrm{Tr}(B_3)\mathrm{Tr}(B_4)\right)|. \tag{17}$$

We now bound each terms in equation (17). To bound the first term, $\mathrm{Tr}(\langle\beta,\beta\rangle)$, we use the following lemma about Gaussian distribution.

Lemma 4.7. Let $\beta_1, \cdots, \beta_n$ be n i.d.d. $\mathcal{N}(0,1)$ random variables. Then,

$$\mathsf{Pr}\left[\sum_{i=1}^{n}\beta_i^2 > 2(1+\eta)n\right] \leq 2^{-\eta n/2}.$$

Proof. First, from the definition of normal distribution, we know that $\mathsf{Pr}[\beta_i = t] = \frac{1}{\sqrt{2\pi}}\exp\left(-t^2/2\right)$. Then consider the random variable $Z_i = \exp(\beta_i^2/4)$. 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,

$$\begin{aligned}
\mathsf{Pr}_{\beta_1,\cdots\beta_n}[\beta_1^2 + \cdots \beta_n^2 > \lambda] &= \mathsf{Pr}_{\beta_1,\cdots,\beta_n}\left[\frac{\beta_1^2 + \cdots \beta_n^2}{4} > \frac{\lambda}{4}\right] \\
&= \mathsf{Pr}_{\beta_1,\cdots,\beta_n}\left[\exp\left(\frac{\beta_1^2 + \cdots \beta_n^2}{4}\right) > \exp\left(\frac{\lambda}{4}\right)\right] \\
&\leq \exp(-\lambda/4)\mathbb{E}_{\beta_1,\cdots,\beta_n}\left[\exp\left(\frac{\beta_1^2 + \cdots \beta_n^2}{4}\right)\right].
\end{aligned}$$

Since all β_i are i.d.d., the above expression is bounded as

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

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

We now look at each of the terms B_1, B_2, B_3, and B_4 in equation (17). $\langle\alpha|V\Sigma^{-2}V^{\mathsf{T}}|v\rangle$ is distributed as $\mathcal{N}(0, \|V\Sigma^{-2}V^{\mathsf{T}}|v\rangle\|^2)$, $\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}}|\alpha\rangle$ as $\mathcal{N}(0, \|\langle v|\tilde{V}\Lambda^{-2}\tilde{V}^{\mathsf{T}}\|^2)$, $\langle\alpha|V\Sigma^{-1}U^{\mathsf{T}}|e_i\rangle$ as $\mathcal{N}(0, \|V\Sigma^{-1}U^{\mathsf{T}}|e_i\rangle\|^2)$, and $\langle e_i|\tilde{U}\Lambda^{-1}\tilde{V}^{\mathsf{T}}|\alpha\rangle$ 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 $1/\sigma_{\mathsf{min}}, 1/\sigma_{\mathsf{min}} + 1/\sigma_{\mathsf{min}}^2, 1$, and $1 + 1/\sigma_{\mathsf{min}}$, respectively. It is easy to see that the second term in equation (16) is bounded by $\mathrm{Tr}(\langle\beta,\beta\rangle\langle\alpha|UV^{\mathsf{T}}|\alpha\rangle)/2\sigma_{\mathsf{min}}^2 \leq \mathrm{Tr}(\langle\beta,\beta\rangle)/2\sigma_{\mathsf{min}}^2$ using the same arithmetic as above. Note that

$$\frac{4(1+\eta)n}{\sigma_{\mathsf{min}}^2}\left(\frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2}\right)\ln(4/\delta_0) \leq \frac{4(1+\eta)n}{\sigma_{\mathsf{min}}}\left(\frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2}\right)\ln(4/\delta_0) \leq 4\varepsilon.$$

Therefore, we have

$$\mathsf{Pr}\left[(16) \leq 4\varepsilon\right] \geq 1 - \delta.$$

Rescaling the value of ε, the lemma follows. □

For the rest of this section, we show that the PDF is proportional to the form used in the analysis above.

Computing the PDF. Since we are going to divide the PDFs corresponding to the output distribution of the two neighbouring matrices for the proof of differential privacy, we do not care about the normalization factor in the rest of the proof. Our method to compute the PDF is generic three step process to compute any matrix-variate distribution. In the first step, we compute the joint probability distribution of individual entries of Ω. In the second step, we use the idea that every row of Ω is i.d.d. to get the multivariate distribution corresponding to individual row. Finally, we use the fact that two rows of the matrix Ω has different α_i to give the matrix variate distribution[1].

Let

$$a(x) := \frac{1}{\sqrt{2\pi\sigma_a}} \exp\left(-\frac{x-\mu_a}{2\sigma_a}\right) \qquad \text{and} \qquad b(x) := \frac{1}{\sqrt{2\pi\sigma_b}} \exp\left(-\frac{x-\mu_b}{2\sigma_b}\right),$$

then their product has the term in exponent,

$$\begin{aligned} e &= \left(-\frac{(x-\mu_a)}{2\sigma_a} + -\frac{(x-\mu_b)}{2\sigma_b}\right) \\ &= \frac{(\sigma_a^2+\sigma_b^2)x^2 - 2(\mu_a\sigma_b^2 + \mu_b\sigma_b^2)x + \mu_a^2\sigma_b^2 + \mu_b^2\sigma_a^2}{2\sigma_a^2+\sigma_b^2} \\ &= \frac{x^2 - 2\frac{\mu_a\sigma_b^2+\mu_b\sigma_b^2}{2\sigma_a^2+\sigma_b^2}x + \frac{\mu_a^2\sigma_b^2+\mu_b^2\sigma_a^2}{2\sigma_a^2\sigma_b^2}}{2\frac{\sigma_a^2\sigma_b^2}{\sigma_a^2+\sigma_b^2}} \end{aligned}$$

Note that e has a quadratic form, but we need to complete the square to get it in a desired form. For the sake of brevity, let us denote by

$$\sigma := \sqrt{\frac{\sigma_a^2\sigma_b^2}{\sigma_a^2+\sigma_b^2}} \qquad \text{and} \qquad \mu := \frac{\mu_a\sigma_b^2 + \mu_b\sigma_b^2}{\sigma_a^2+\sigma_b^2}.$$

Then we can rewrite

$$e := \frac{(x-\mu)^2}{2\sigma^2} + \frac{(\mu_a-\mu_b)^2}{2(\sigma_a^2+\sigma_b^2)}$$

Completing the squares, we have

$$a(x)b(x) := \frac{1}{N}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\exp\left(-\frac{(\mu_a-\mu_b)^2}{2(\sigma_a^2+\sigma_b^2)}\right) = \frac{1}{N}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{18}$$

The second equality follows because $\mu_a = \mu_b = 0$ and $\sigma_a = \sigma_b = 1$. Therefore, the distribution of $a(x)b(x)$ is proportional to $\exp(-x^2/4)$. Now consider a n-variate distribution which has the above product

[1]The entries of Ω are $\Pi'_{1..r}\Omega'$ where $\Omega'_{ij} = \alpha_i(\Pi(W\beta)_j)$. The first observation is that neither of the two permutations effects the final result. This is because, as all β_j are i.d.d. Gaussian, $\Pi W\beta$ and β has the same distribution (spherical symmetry of random Gaussian vector). Therefore, we do not include Π to avoid cluttering of too many expressions. Further, we do not include $\Pi'_{1..r}$ in our computation because differential privacy is preserved under any post-processing.

form. From the independence of these variables (any entry in the i-th row vector has β_j for $1 \leq j \leq n$ i.d.d. and scaled according to α_i). Note that this is not a χ^2-distribution, which is the sum of squares of entries picked from a normal distribution, i.e., the sum in Lemma 4.7.

We now perform the second step, i.e., the transition from univariate to multivariate distribution. The technique is quiet standard to derive the multivariate version of any univariate distribution. Let $\mathbf{z} = (z_1, \cdots, z_n)$ such that Z_i is distributed proportional to equation (18). Then we can write, $\mathsf{PDF}_{\mathbf{Z}}(\mathbf{z})$ is proportional to

$$\prod \exp\left(\frac{-z_i^2}{2}\right) = \exp\left(\frac{-\langle \mathbf{z}, \mathbf{z}\rangle}{2}\right)$$

Using the fact that $\mathbb{E}[\mathbf{z}] = 0$ and $\mathsf{COV}[\mathbf{z}] = \mathbb{I}_n$, we can compute the t moment generating function as below.

$$M_{\mathbf{Z}}(t) = \mathbb{E}[\exp(\langle t, \mathbf{z}\rangle)] = \mathbb{E}\left[\prod_{i=1}^{n} \exp(t_i z_i)\right] = \exp\left(\frac{1}{2}\sum t_i^2\right) = \exp(\langle t, t\rangle/2).$$

Now, this is fine for n independent univariate distribution and is used as a stepping stone for the more complicated multivariate distribution. Let Σ be the covariance matrix of the multivariate distribution. The multivariate distribution is defined only for positive definite matrix, so we assume that Σ is a positive definite matrix. Let μ be the mean of the distribution. We want to find the probability distribution function of a multivariate distribution where the mean is μ and variance is Σ. This can be done by using the substitution $\mathbf{x} = \sqrt{\Sigma}\mathbf{z}$. Now lets compute the moment generating function.

$$M_{\mathbf{x}}(t) = \mathbb{E}[\exp(\langle t, \mathbf{x}\rangle)] = \mathbb{E}[\exp(\langle t|\Sigma^{1/2}|\mathbf{z}\rangle + \langle t, \mathbf{z}\rangle)] = \mathbb{E}[\exp(\langle \Sigma^{1/2}t, \mathbf{z}\rangle)] = \mathbb{E}[(\langle t|\Sigma|t\rangle)/2].$$

Now, from the transformation method, $\mathsf{PDF}_{\mathbf{X}}(\mathbf{x}) = \mathsf{PDF}_{\mathbf{Z}}(\mathbf{z})\det(\Sigma)^{-1/2}$, where $\det(\Sigma)^{-1/2}$ is the Jacobian. Therefore, we have $\mathsf{PDF}_{\mathbf{X}}(\mathbf{x})$ is proportional to

$$\frac{1}{\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathsf{T}}\Sigma^{-1}\mathbf{x}\right).$$

In particular, when $\Sigma = A^{\mathsf{T}}A$, the PDF is proportional to the following

$$\det(A^{\mathsf{T}}A)^{1/2} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathsf{T}}(A^{\mathsf{T}}A)^{-1}\mathbf{x}\right). \tag{19}$$

We next show that $\Sigma \in S^n_{++}$, i.e., Σ has to semi-definite matrix. We use this in order to get the closed form expression for matrix-variate distribution. We start by showing the standard result that Σ has to be positive semi-definite.

Proposition 4.8. Suppose that Σ is the covariance matrix corresponding to some random vector $\mathbf{x}$. Then Σ is symmetric positive semi-definite.

Proof. For any vector $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x}^{\mathsf{T}}\Sigma\mathbf{x} = \sum\sum(\Sigma_{ij}x_i x_j) = \sum\sum(\mathsf{COV}[x_i, x_j])\, x_i x_j = \mathbb{E}\left[\sum\sum(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])x_i x_j\right].$$

Now the quantity under the summation is of form $\sum\sum x_i x_j z_i z_j = (\mathbf{x}^{\mathsf{T}}\mathbf{z}) \geq 0$. Therefore, the quantity inside the expectation is always non-negative; therefore, the expectation is non-negative. This proves the proposition. Now, for the definition of PDF for the above multivariate distribution, Σ^{-1} should exists; therefore, $\Sigma \in S^n_{++}$. □

Now, we move on to compute the closed-formed expression for the matrix-variate distribution we are working with using Proposition 4.8. Recall that Ω is picked from a distribution of matrix whose each entries are $(-1)^{(i-1)(j-1)}\alpha_i\beta_j$ for $1 \leq i, j \leq n$ for $\alpha_i, \beta_j \sim \mathcal{N}(0,1)$. This implies that the row variance and the column variance of the matrix are both $A^\mathsf{T}A$ (due to independence of α_i across rows and β_j across columns and since both of them are picked from the same normal distribution independent of each other). Therefore, we can directly use the result in matrix variate distribution that extends a multi-variate distribution to matrix-variate distribution. More specifically, using equation (19), we get the matrix-variate distribution proportional to

$$\frac{1}{\det(A^\mathsf{T}A \otimes A^\mathsf{T}A)} \exp\left(-\frac{1}{2}\mathrm{Tr}\left((A^\mathsf{T}A \otimes A^\mathsf{T}A)^{-1}\Omega^\mathsf{T}\Omega\right)\right).$$

Now, using the following identities for positive semidefinite matrices stated in Lemma 2.8,

$$\det(A^\mathsf{T}A \otimes A^\mathsf{T}A) = \det(A^\mathsf{T}A)^n \times \det(A^\mathsf{T}A)^n$$
$$\mathrm{Tr}\left((A^\mathsf{T}A \otimes A^\mathsf{T}A)^{-1}\Omega^\mathsf{T}\Omega\right) = \mathrm{Tr}\left((A^\mathsf{T}A)^{-1}\Omega^\mathsf{T}(A^\mathsf{T}A)^{-1}\Omega\right),$$

as required, we have the probability distribution is proportional to

$$\frac{\exp\left(-\frac{1}{2}\mathrm{Tr}\left((A^\mathsf{T}A)^{-1}\Omega^\mathsf{T}(A^\mathsf{T}A)^{-1}\Omega\right)\right)}{\det(A^\mathsf{T}A)^{2n}}$$

as required. □

5 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 steps 2 and 3 in Figure 2 and step 2 in Figure 3), and stop updating the data structure once d rows (or columns) are streamed. This simplifies the presentation as well as the analysis. By the argument of Hardt and Roth [34], this leads to a depreciation of the privacy guarantee by at most half. The utility proof does not change for matrix multiplication and linear regression if we use A or A'. However, the analysis for utility gets more complicated in the case of LRA because right and left singular vectors are 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 matrix product and linear regression. For example, if we naively use the streamed matrix in PSG and then use the product estimates, 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, the identity term of the published matrix causes the additive error. We use the same idea for linear regression as well. To prove the privacy of linear regression and matrix product, we use the first variant, while for low-rank approximation, we need both the variants defined in Figure 1. The details are in the respective sections.

We note that this is not the only method that could be used for the first two problems. However, an alternative would require some changes in step 2: compute the SVD using one of the standard algorithms [52, 57] and then add scaled e_t to the matrix. Surprisingly, there are streaming algorithms for computing the SVD of a matrix [52, 57]– by say, setting the target rank in the algorithm of [52] to be $\mathsf{rank}(A)$. However, this makes the mechanisms more complicated with added utility loss. For the sake of coherence and simplicity, we do not explore this idea any further.

Initialization: On input parameters $\alpha, \beta, \varepsilon, \delta$, set $r = O(\log(1/\beta)/\alpha^2)$. Pick an $r \times 2(n+d)$ matrix as in Section 4, Ω. Set $s = \frac{16r\ln(2/\delta)}{\epsilon}\ln(16r/\delta)$. Set Y_{A_0} and Y_{B_0} to be all zero matrix.
On input the row r of a $d \times n$ matrix A and column c of an $n \times d$ matrix B, the mechanism does the following:

1. Compute $\|A_{r:}\|_F$ and $\|B_{:c}\|_F$ and update the values of $\|A_t\|_F$ and $\|B_t\|_F$.
2. Set $\tilde{A}_{r:} = s\begin{pmatrix} e_r & \langle 0^{n+d}| & A_{r:}\end{pmatrix}$ and $\tilde{B}_{:c} = s\begin{pmatrix} e_c & \langle 0^{n+d}| & B_{:c}\end{pmatrix}$.
3. Invoke variant 1 of PSG with inputs $(\tilde{A}_{r:}, \mathsf{flag} = 0, r, \varepsilon, \Omega)$ and $(\tilde{B}_{:c}, \mathsf{flag} = 1, r, \varepsilon, \Omega)$. Append the rows and columns returned to update the sketch Y_{A_r} and Y_{B_c}.

On query to compute the matrix product at time t, compute the product using $Y_{A_t}/s, Y_{B_t}/s$.

Figure 2: The Mechanism for Matrix Product

5.1 Matrix Multiplication

In this section, we present our mechanism to compute the matrix product (see Figure 2).

Theorem 5.1. Let Ω be an $r \times 2(n+d)$ random matrix as described in Section 4. Then, the mechanism in Figure 2 preserves (ε, δ)-differential privacy while using $O(d\alpha^{-2}\kappa\log(1/\beta))$ bits of space to compute the matrix product of two conforming matrices A and B with an additive error, $\tau \leq \sqrt{n}\alpha$.

Proof. It is easy to see that the space required to store the sketch is at most $O(d^2\kappa\log(1/\beta)/\alpha)$ bits. Similarly, for storing the norms and other things in the data structure, we use at most rd entries; therefore, it does not change the space complexity in terms of $O(\cdot)$ bound.

The privacy guarantee follows because the mechanism maintains the singular values of input matrix above the threshold of Theorem 3.1 ($\tilde{A}_i\tilde{A}_i^\mathsf{T} \succeq U\,(32r\ln(2/\delta)/\varepsilon)\ln(16r/\delta)^2\mathbb{I}U^\mathsf{T} = \sigma_{\mathsf{min}}^2\mathbb{I}$ for both $i = \{1, 2\}$). The proof of utility readily follows from the proof of Clarkson-Woodruff [15]. To prove the utility guarantee of our mechanism, we need a variance bound on $\|A\Omega^\mathsf{T}\Omega B - AB\|_F^2$. Using this, the rest of the proof is just arranging the terms, bounding $\|\Omega^\mathsf{T}\Omega - \mathbb{I}\|_2$, and using sub-additivity of norm. The following lemma corresponding to our projection matrix in Section 4 follows immediately from using the proof of the corresponding lemma of Sarlos [55, Corollary 4].

Lemma 5.2. Let Ω be a $k \times d$ matrix as constructed in Section 4 with every entries picked from the distribution $\mathcal{N}(0, 1)$, then for a set of m vectors, $v_1, \cdots, v_m \in \mathbb{R}^n$, with probability at least $1 - 2\exp(-k\alpha^2/8)$, for any pair v_i, v_j, we have $|\langle\Omega v_i, \Omega v_j\rangle - \langle v_i, v_j\rangle| \leq \alpha\|v_i\| \cdot \|v_j\|$.

The proof of the utility of mechanism in Figure 2 follows readily from Lemma 5.2. We need to upper bound the quantity $\|AB - \tilde{A}\Omega^\mathsf{T}\Omega\tilde{B}/s^2\|_F$. Now

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

Therefore,

$$\|AB - \tilde{A}\Omega^\mathsf{T}\Omega\tilde{B}/s^2\|_F = \|AB - A\Omega^\mathsf{T}\Omega B - \mathbb{I}\Omega^\mathsf{T}\Omega\mathbb{I}\|_F \leq \|AB - A\Omega^\mathsf{T}\Omega B\|_F + \|\mathbb{I} - \Omega^\mathsf{T}\Omega\|_F. \quad (20)$$

To bound the first term, let random variable X_{ij} denote $(AB)_{ij} - (A\Omega^\mathsf{T}\Omega B)_{ij}$. Then, with probability at least $1 - 2\exp(-k\alpha^2/8)$, we have $|X_{ij}| \leq \alpha\|A_{i:}\|_2 \cdot \|B_{:j}\|_2$, and using Lemma 5.2,

$$\|AB - A\Omega^\mathsf{T}\Omega B\|_F{}^2 = \sum|X_{ij}|^2 \leq \sum\alpha^2\|A_{i:}\|_2^2\|B_{:j}\|_2^2 \leq \alpha^2\|A\|_F^2\|B\|_F^2. \quad (21)$$

For the second term, we need to bound the variance on unitaries. For any unitaries U_1 and U_2, the analysis of Kane and Neilson [39] gives $\|U_1^\mathsf{T}\Omega^\mathsf{T}\Omega U_2 - U_1U_2\|_2 \leq \alpha$. Plugging this and equation (21)

Initialization: On input parameters $\alpha, \beta, \varepsilon, \delta$, set $r = O(d\log(1/\beta)/\alpha)$. Pick an $r \times 2(n+d)$ matrix, Ω, as constructed in Section 4. Set $s = \frac{32\log(1/\beta)\ln(2/\delta)}{\alpha\varepsilon}\ln(16r/\delta)$ and Y_{A_0} and Y_{B_0} to be all zero matrix.
On input the column c of an $n \times d$ matrix A, the mechanism does the following:

1. Set $\tilde{A}_{:c} = s\left(e_c \quad \langle 0^{n+d}| \quad A_{:c}\right)$.
2. Invoke variant 1 of PSG with inputs $(\tilde{A}_{:c}, \mathsf{flag} = 0, r, \varepsilon, \Omega)$. Append the column returned to get an updated sketch Y_{A_c}.
3. On being queried with a set of query vectors $B = \{b_1, \cdots, b_m\}$, invoke PSG as in step 1 with input matrix $(B, \mathsf{flag} = 1, r, \varepsilon, \Omega)$ to get the sketch Y_B by appending the new column.

Compute an X satisfying $\min_X \|Y_{A_t}X - Y_B\|$.

Figure 3: The Mechanism for Linear Regression

in equation (20), using the fact that $\|X\|_2 \leq \|X\|_F \leq \sqrt{n}\|X\|_2$ for any $n \times n$ matrix X, and adjusting the value of α, the result follows.

□

5.2 Linear Regression

The mechanism for linear regression follows from the observation that PSG is a low distortion embedding of the vectors. Therefore, if one solves the linear regression problem in the lower dimensional space, it should translate to the higher dimension. The technical part is to prove that this intuition is in fact correct.

Theorem 5.3. Let Ω be an $r \times 2(n+d)$ matrix, where $r = O(d\log(1/\beta)/\alpha)$. Then for every constant $c > 32$, with probability at least $1-\beta$, if $\tilde{X}$ is the solution of $\min_X \|\Omega(AX - B)\|^2$ for an $n \times d$ input matrix, then the mechanism in Figure 3 uses $O(d^2\alpha^{-1}\kappa\log(1/\beta))$ bits and solves the linear regression problem in an (ε, δ)-differentially private manner with additive error $\tau \leq O(\sqrt{n}\alpha)$.

This is better than the bound achieved by Jain, Kothari, and Thakurta [37] for $\alpha < 1$, which is a more practical choice of parameter. Jain, Kothari, and Thakurta [37] gave a private learning algorithm for linear regression and achieve 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[2].Our approach to prove the utility bound for linear regression follows from straightforward application of the result of Kane and Nelson [39] to the proof of Clarkson and Woodruff [15]: first bound $\|U^{\mathsf{T}}(A\tilde{X} - X^*)\|_F$, where X^* is the right value of the linear regression, and $\tilde{X}$ is the value of regression for the sketch of A and B using Lemma 5.4, and then apply Pythagorus theorem and an observation that A and U have the same column-space.

Lemma 5.4. *(Sarlos [55], 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 $\|U\Omega^{\mathsf{T}}\Omega U^{\mathsf{T}} - \mathbb{I}\|_2 \leq \alpha$.

For the sake of completion, we give the detailed proof in Appendix A.

5.3 Low Rank Approximation

The mechanism of Hardt and Roth [34] is a privacy-preserving version of the prototype mentioned in [31]. They improve the worst case lower bound under a *low coherence* assumption. The main prototype in [31] is the following: construct a low-dimensional subspace that captures the action of the matrix (*range-finding*) and restrict the matrix to that subspace to compute the required factorization (*projection*). More concretely,

[2]It should be noted though that the setting of private learning algorithm is stricter than streaming model and might be one of the reason why we achieve such significant improvement.

On input parameters $\alpha, \beta, \varepsilon, \delta$, the target rank k, set $w = \frac{16k\ln(2/\delta)}{\varepsilon}$. 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 $Y_A = \begin{pmatrix} w\mathbb{I} & A \end{pmatrix} \Omega$ as in Figure 2 and Figure 3. 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

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.

We show that the two-step prototype mentioned in [31] can be replaced by a two-step algorithm in which the input matrix is explicitly needed only in the first step. Let $\hat{U}\Lambda\hat{V}^{\mathsf{T}}$ be the factorization of the low rank approximation of A we aim to achieve. The crucial observation 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. Since we are using Ω twice, we have to rely on privacy of the second variant of PSG algorithm. 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 $\|A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| = \|A - \Psi\Psi^{\mathsf{T}}A + \Psi\Psi^{\mathsf{T}}A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| \leq \|A - \Psi\Psi^{\mathsf{T}}A\| + \|\Psi\Psi^{\mathsf{T}}A - \Psi\Psi^{\mathsf{T}}A\Psi\Psi^{\mathsf{T}}\| \leq 2\eta$. In order to simplify the presentation, we state our mechanism for symmetric matrices in Figure 4 (note that [35, 40] also made this assumption).

In Figure 4, we give a mechanism for symmetric matrices. To construct a mechanism for non-symmetric matrices, we construct two sketches Y_t and $\tilde{Y}_t$ corresponding to A_1 and A_2 using a single pass over A and using two Gaussian matrices Ω and $\tilde{\Omega}$ of appropriate dimension, where $A_1 = \begin{pmatrix} s\mathbb{I} & A \end{pmatrix}$ and $A_2 = \begin{pmatrix} A^{\mathsf{T}} & s\mathbb{I} \end{pmatrix}$ for appropriate dimension identity matrices in each of A_1 and A_2. Basically, we do the following using a single pass over the matrix A:

$$\begin{pmatrix} Y \\ \tilde{Y} \end{pmatrix} = \begin{pmatrix} s\mathbb{I}_n & A \\ A^{\mathsf{T}} & \end{pmatrix} \begin{pmatrix} \Omega \\ \tilde{\Omega} \end{pmatrix},$$

where $\mathbb{I}_n$ is an $n \times n$ identity matrix. Since $\begin{pmatrix} Y & \tilde{Y} \end{pmatrix}^{\mathsf{T}}$ corresponds to a symmetric matrix, we can use the steps used in the projection stage in Figure 4. We state our result for the general case.

Theorem 5.5. 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 $O(k(n+d)\alpha^{-2}\log(nd))$ bits while preserving (ε, δ)-differential privacy such that

$$\|A - \tilde{A}\|_F \leq \left(1 + \frac{k}{p-1}\right)^{1/2} \min_{rk(A')<k} \|A - A'\|_F + \frac{2k\sqrt{(n+d)\ln(2/\delta)}}{\varepsilon}, \quad \text{and} \tag{22}$$

$$\|A - \tilde{A}\|_2 \leq \left(1 + \frac{k}{p-1}\right)^{1/2} \lambda_{k+1} + \frac{e\sqrt{(k+p)\sum_{j>k}\lambda_j^2}}{p} + \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}. \tag{23}$$

We need both the variants of PSG algorithm for the privacy proof–we use the first variant for range finding step 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 3.1. Since rest of the computations are deterministic, privacy follows from Lemma 2.2, Theorems 2.1 and 3.1, 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 $\tilde{A}$ after the run of the algorithm. Thus, B must satisfy $B\Psi^\mathsf{T}\Omega \approx \Psi^\mathsf{T}Y$. This is what step 2 does. Therefore, $\tilde{A}_t = \Pi_Y A_t$ for the projection operator Π_Y with the same range as Ψ. The rest of the utility proof relies heavily on the fact that $(\Omega\Omega^\mathsf{T})^{-1}$ exists, left singular vectors does not play any essential role in the concentration bound, and the rotational invariance of multivariate Gaussian distribution. 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. We note that our proof is very different from the proof of Sarlos [55] which is for two-pass algorithm and Clarkson-Woodruff [15], and more in line with the proof of Halko *et al.* [31]. The proof of Sarlos [55] and Clarkson-Woodruff [15] uses the error bound computed in the estimate of matrix multiplication, while that used here is based on perturbation theory. The details follows.

Proof. The space complexity is easy to follow. The privacy guarantee follows from Theorem 3.1 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 3.1, and the distribution of $\Psi^\mathsf{T}\Omega = \Psi^\mathsf{T}\Pi_Y\Omega$ is the same as that of the second variant as we reuse Ω. The latter statement can be shown using the proof of Bura and Pfeiffer [10]. They showed that if a random matrix has a normal distribution in limits, then its left singular matrix has a normal limit distribution. Using their idea for normally distributed matrix, the left singular matrix has a normal distribution[3]. A simplified argument could be the following. Since the entries of $\Omega \in \mathbb{R}^{n\times r}$ 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., $v = \mathsf{vec}(V)$ represent rn vector with entries $v_{i+nj} = V_{ij}$. Then $\mathsf{vec}(G\Omega R^\mathsf{T}) = (G \otimes R)\mathsf{vec}(\Omega)$ using Lemma 2.8. 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 $G\Omega R^\mathsf{T}$ (owing to the invariance under rotation). Consequently the distribution of each singular vector is also spherically distributed. It is easy to see that Ψ is the left singular matrix of Y; therefore, from the above argument, it is also spherically distributed.

Now it follows from the proof of the second variant (Theorem 3.1) that it does not matter if we multiply A (or A^T) from left (or right, respectively) of vectors $\alpha_1, \cdots, \alpha_r$, i.e., $\sum_{i=1}^r A|\alpha_i\rangle\langle\alpha_i|$, $\sum_{i=1}^r |\alpha_i\rangle\langle\alpha_i|A^\mathsf{T}$, and $\sum_{i=1}^r |\alpha_i\rangle A^\mathsf{T}\langle\alpha_i|$ 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 preprocessing preserves differential privacy, we can now complete the proof by invoking Theorem 3.1. The privacy guarantee due to Theorem 3.1 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 eigen-values 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 3.1.

For the utility guarantee, we need to show that the mechanism does what [31, Section 1.2] prototype algorithm achieves. The main prototype in [31] and that used by [34] is the following: construct a low-dimensional subspace that captures the action of the matrix (*range-finding*), restrict the matrix to that subspace, and compute a standard 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 $\Pi_Y = Y(Y^\mathsf{T}Y)^{-1}Y^\mathsf{T}$; projection then computes a rank k matrix $B = \Pi_Y A$. Note that $\Pi_Y = \Psi\Psi^\mathsf{T}$ for

[3]One gets a different covariance matrix, scaled by an orthonormal matrix, from that of $A^\mathsf{T}\Omega$, but since in our proof, we are only concerned with the singular values of the covariance matrices, this does not effect our analysis.

some orthonormal basis Ψ.

The utility guarantee needs a slight tweak from [31] though overall structure remains the same. Mainly, we need to worry about the computation of the minimal residual method and the requirement on the input to PSG algorithm. We include it here to make sure that we achieve the bound as claimed.

Keeping the most general case in mind, we first show that the left singular value has hardly any role to play in bounding the perturbation. We assume that we perform SVD – the case for eigenvalue decomposition follows similarly. 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 this matrix is more manageable for computation and any upper bound on the approximation of this matrix is an upper bound on the approximation of the original matrix. The actual bound on the approximation of the private matrix as computed in the main text 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 $\tilde{A} = \Pi_Y A$; therefore, we need to bound $\|(\mathbb{I} - \Pi_Y)A\|$, where $\|\cdot\|$ in this section refers to both the Frobenius as well as the spectral norm. From the Hölder's inequality on the second moment, we have

$$\mathbb{E}[\|(\mathbb{I} - \Pi_Y)A\|_F] \leq \left(\mathbb{E}\left[\|(\mathbb{I} - \Pi_Y)A\|_F^2\right]\right)^{1/2}. \tag{24}$$

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 other 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 a $n \times k$ matrix; therefore, $Y = A\Omega$ and $Y = U\begin{pmatrix} \Lambda_1 V_1^{\mathsf{T}}\Omega & \Lambda_2 V_2^{\mathsf{T}}\Omega \end{pmatrix}^{\mathsf{T}}$. 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.

$$\|(\mathbb{I} - \Pi_Y)A\| = \|U^{\mathsf{T}}(\mathbb{I} - \Pi_Y)A\| = \|U^{\mathsf{T}}(\mathbb{I} - \Pi_Y)UA'\| = \|\mathbb{I}A' - U^{\mathsf{T}}\Pi_Y UA'\|. \tag{25}$$

Now note that for an orthogonal projector corresponding to a matrix Y is uniquely defined by $\mathsf{range}(Y)$, the range of Y. Therefore,

$$\mathsf{range}(U^{\mathsf{T}}\Pi_Y U) = U^{\mathsf{T}}\mathsf{range}(\Pi_Y) = \mathsf{range}(U^{\mathsf{T}}\Pi_Y). \tag{26}$$

This gives us that $\|A' - U^{\mathsf{T}}\Pi_Y UA'\| = \|(\mathbb{I} - \Pi_{U^{\mathsf{T}}Y})A'\|$. 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. Our next idea is to 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 that 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 rest of the 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:

$$\|(\mathbb{I} - \Pi_{U^\intercal Y})A'\| \leq \|(\mathbb{I} - \Pi_C)A'\|. \tag{27}$$

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

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

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

$$\|(\mathbb{I} - \Pi_C)A'\| \leq \|S^{\mathsf{T}}SA'\| + \|A'\| \tag{28}$$

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

$$\|(\mathbb{I} - \Pi_C)A'\| \leq \sqrt{\|\Lambda_2'\| + \|\Lambda_2' V_2^{\mathsf{T}}\Omega(V_1^{\mathsf{T}}\Omega)^{-1}\|} \tag{29}$$

for both the required norms. Using equation (25), equation (27) and equation (28), this gives us a bound on the approximation of matrix A'. Till this point, our analysis closely follows the ideas of [31], 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.

5.3.1 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 (29) 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 (29), and combining equation (24) and the fact that $(\Omega\Omega^{\mathsf{T}})^{-1}$ has a well defined trace $k/(p-1)$ [48] and the relation $\|\Lambda'\|_F = \sqrt{\|\Lambda^2 + w^2\mathbb{I}\|_F} \leq \|\Lambda\|_F + w\|\mathbb{I}\|_F$.

5.3.2 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 (29). In particular, using Lemma 2.9 and 2.10, and Holder's inequality, the statement of the theorem for 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 2.10 followed by the sub-additivity of norms. Making these substitution and on simplification, we get the bound stated in equation (23) of Theorem 5.5.

$$(29) \leq \left(1 + \frac{k}{p-1}\right)^{1/2}\|\Lambda_2\|_2 + \frac{e\sqrt{(k+p)\sum_{j>k}\lambda_j^2}}{p} + \frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}.$$

□

We compare our results with the best possible results in the non-private setting. Eckart and Young [24] have shown that the quantity $\min_{rk(A')<k}\|A - A'\|_F$ in the first term of equation (22) is optimal. Likewise, Mirsky [47] proved that λ_{k+1} is the minimum spectral error for k-rank approximation. The second term in equation (23) 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,

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

therefore, the error always lies within some polynomial factor of λ_{k+1}, modulo some additive error.

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 $\delta \neq 0$.

5.3.3 Comparison with Earlier Works

In the recent past, 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 [34]: This work uses two passes over the input matrix; therefore, it does not fall in our one-pass streaming model of computation. They also work in the case of event level privacy–two datasets are neighbouring if they differ in at most one entry of Euclidean norm at most one. This makes their coherence conditions and notion of neighbouring data sets rotationally invariant. We use the same concept of neighbouring data sets; however, as argued by Blocki *et al.* [6], we achieve a better utility bound in the range finding step. Intuitively, this could be seen due to the absence of additive Gaussian noise.

Hardt and Roth [34] achieved an error bound of $\sqrt{kn}\log(k/\delta)/\varepsilon + \sqrt{\mu\|A\|_F(n/d)^{1/2}\log(k/\delta)/\varepsilon}$. Without any coherence assumption, their error bound depends on $\|A\|_F$, which can be as large as $\sqrt{nd}$ for binary matrices. On the other hand, we achieve a bound that is independent of $\|A\|_F$.

Hardt-Roth [35]: In some sense, this paper is based on Krylov subspace iteration combined with powering method of Halko *et al.* [31]. The coherence definition used in this paper depends on the maximum value of the left or right singular vectors. This makes their coherence condition rotationally variant. Moreover, they also make an assumption regarding the singular value separation, i.e., the first and the k-th singular value has a non-trivial separation. They define two data-sets as neighbouring in the same manner as in Hardt and Roth [35]. They give LRA in spectral norm, and therefore, their work has application in problems like principal component analysis. 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.* [31]; therefore, it cannot be implemented in a streaming fashion. Moreover, the error bound is $O(k^2\varepsilon^{-1}\sqrt{(\mathsf{rank}(A)\mu + k\log n)\log(1/\delta)}\log n)$ compared to $\frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}$ (equation (23)).

Kapralov-Talwar [40]: The only assumption this paper makes is that of singular value separation of the same form as in Hardt-Roth [35]. 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 error bound, they achieve a bound of $O(dk^3/(\varepsilon\gamma^2\delta^2))$compared to $\frac{2\sqrt{k(n+d)\ln(2/\delta)}}{\varepsilon}$ (equation (23)).

Hardt [32]: In this recent work, Hardt [32] 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 [35]. They define two data-sets as neighbouring in the same manner as in Hardt and Roth [35]. 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; 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 (23)) 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 [32].

Dwork *et al.* [23]: Dwork *et al.* [23] 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 (note that, we do not make any such assumption). However, they assume that two data-sets are neighbouring if they differ by at most one row unlike our notion which is the same as Hardt and Roth [34, 35]. Therefore, we do not see any natural way to compare. They give a bound that assumes a lower bound of $k\sqrt{n}\log^2(m/\delta)/\varepsilon^2$ on the optimal value, for $\delta < 1/\operatorname{poly}(n)$. 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 (23)). 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.

An Alternate Range Finding Step. We mention an alternative, though we do not pursue it any further, to the range finding step in Figure 4 for the sake of completion of the argument given in Section 3. One can also compute $Y_A = U(\Sigma^2 + w^2\mathbb{I}_{r\times r})^{1/2}V^{\mathsf{T}}\Omega$ using standard single-pass streaming algorithms on SVD computation [52, 57]– by say, setting the target rank in [52] to be $\mathsf{rank}(A)$. The projection step and our proof still remains the same. However, we lose on utility due to the SVD computation.

6 Discussions and Open Problems

In this paper, extending on a few of the earlier results [20, 23, 37], we explored differential privacy on streamed data. Ours is the first work that gives positive results for private analogue of sketch based streaming algorithms. Previously, there were known negative results for private analogue of sketch based approach for specific statistical queries [20]. On the backdrop of these two conflicting results, one important open problem is to characterize the class of queries that allow private sketch based mechanism. Another direction of future research is to study privacy of streamed data under different notion of neighbouring data. Another really interesting problem is to get a tighter space lower bound for answering these queries privately in the streaming model. In this paper, we only considered row-wise or column-wise stream. An interesting question is to consider turnstile stream. Another open problem, also suggested by Blocki *et al.* [6], is to get a better error bound by, maybe, using error correcting codes. It is also interesting to verify whether other variants of Johnson-Lindenstrauss, more specifically sparse and randomness efficient variants [2, 39, 43], also preserves privacy or not. Our results in Section 3 and Section 4 also raise a question as to whether there is any space-time tradeoff while performing these tasks privately. A particularly interesting problem is to characterize the class of query function that has a differentially private mechanism in the streaming model. We have a fair bit of understanding of the class of query functions when the whole database is provided to the curator. Our mechanism as provided is strictly non-interactive. An interesting domain of future research is what types of privacy in streaming model of computation one can achieve with interaction. Iterative multiplicative weight mechanisms seems to be the most natural candidate because of how it constructs the data-structure.

References

[1] Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009. 15

[2] Nir Ailon and Edo Liberty. An Almost Optimal Unrestricted Fast Johnson-Lindenstrauss Transform. *ACM Transactions on Algorithms*, 9(3):21, 2013. 31

[3] 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. 2, 7, 10

[4] 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

[5] Lucien Birgé and Pascal Massart. *From Model Selection to Adaptive Estimation*, pages 55–87. Springer New York, 1997. 10

[6] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss Transform Itself Preserves Differential Privacy. In *FOCS*, pages 410–419, 2012. 2, 5, 6, 7, 11, 13, 14, 17, 18, 29, 31, 35, 36

[7] 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. 2, 5, 7

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

[9] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *J. ACM*, 60(2):12, 2013. 7

[10] 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. 26

[11] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006. 6

[12] Emmanuel J. Candès and Terence Tao. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006. 6

[13] Lynn E Cannon. A cellular computer to implement the kalman filter algorithm. Technical report, DTIC Document, 1969. 5

[14] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *NIPS*, pages 998–1006, 2012. 7

[15] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In *STOC*, pages 205–214, 2009. 2, 3, 4, 5, 6, 23, 24, 26, 35

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

[17] Scott Deerwester. Improving information retrieval with latent semantic indexing. 1988. 4

[18] 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. 7

[19] 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. 7

[20] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In Schulman [56], pages 715–724. 2, 6, 31

[21] Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N Rothblum, and Sergey Yekhanin. Pan-Private Streaming Algorithms. In *ICS*, pages 66–80, 2010. 1, 7

[22] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and Differential Privacy. In *FOCS*, pages 51–60. IEEE Computer Society, 2010. 7, 8

[23] 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, 7, 30, 31

[24] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. 4, 29

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

[26] Philippe Flajolet and Nigel Martin. Probabilistic Counting. In *Foundations of Computer Science, 2007. FOCS'07*, pages 76–82, 1983. 2, 7

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

[28] 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. 4

[29] 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. 7

[30] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative Constructions and Private Data Release. In *TCC*, pages 339–356, 2012. 7

[31] 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, 24, 25, 26, 27, 28, 30

[32] Moritz Hardt. Robust subspace iteration and privacy-preserving spectral analysis. In *Allerton*, pages 1624–1626. IEEE, 2013. 2, 5, 6, 30

[33] 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. 7

[34] Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. In *STOC*, pages 1255–1268, 2012. 2, 3, 4, 5, 6, 7, 8, 22, 24, 26, 29, 30

[35] Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *STOC*, pages 331–340, 2013. 2, 3, 5, 6, 7, 8, 25, 30

[36] 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. 4

[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. 7, 24, 31

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

[39] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss Transforms. *J. ACM*, 61(1):4, 2014. 7, 23, 24, 31

[40] Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *SODA*, volume 5, page 1. SIAM, 2013. 2, 3, 5, 6, 7, 25, 29, 30

[41] Krishnaram Kenthapadi, Aleksandra Korolova, Ilya Mironov, and Nina Mishra. Privacy via the johnson-lindenstrauss transform. *arXiv preprint arXiv:1204.2606*, 2012. 7

[42] Erricos John Kontoghiorghes. *Handbook of parallel computing and statistics*. CRC Press, 2010. 5

[43] Felix Krahmer and Rachel Ward. New and Improved Johnson-Lindenstrauss Embeddings via the Restricted Isometry Property. *SIAM J. Math. Analysis*, 43(3):1269–1281, 2011. 5, 31

[44] Pascal Massart and Jean Picard. *Concentration inequalities and model selection*, volume 1896. Springer, 2007. 10

[45] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS'07*, pages 94–103. IEEE, 2007. 7

[46] Gerald M Meier et al. The international economics of development. Theory and policy. *The international economics of development. Theory and policy.*, 1968. 3

[47] Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *The quarterly journal of mathematics*, 11(1):50–59, 1960. 4, 29

[48] Robb J Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009. 9, 12, 29

[49] James Ian Munro and Mike Paterson Selection and Sorting with Limited Storage. In *Foundations of Computer Science, 1978. FOCS'78*, pages 253–258, IEEE, 1978. 2, 7

[50] 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. 10

[51] C Radhakrishna Rao. *Linear statistical inference and its applications*, volume 22. John Wiley & Sons, 2009. 12, 13

[52] Radim Rehurek. Subspace tracking for latent semantic analysis. In *Advances in Information Retrieval*, pages 289–300. Springer, 2011. 11, 22, 31

[53] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In Schulman [56], pages 765–774. 7

[54] 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. 5

[55] 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. 2, 4, 5, 6, 23, 24, 26

[56] 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. 32, 34

[57] Volker Strumpen, Henry Hoffmann, and Anant Agarwal. A Stream Algorithm for the SVD. 2003. 11, 22, 31

[58] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012. 9

[59] Jalaj Upadhyay. Random Projections, Graph Sparsification, and Differential Privacy. In *ASIACRYPT (1)*, pages 276–295, 2013. 2, 5, 7, 11

[60] Jalaj Upadhyay. Random Projection Matrix, Restricted Isometry Property, and More. *Submitted to ITCS*, 2015. 5, 14, 15

[61] Jan Vybíral. A variant of the Johnson–Lindenstrauss lemma for circulant matrices. *Journal of Functional Analysis*, 260(4):1096–1105, 2 2011. 16

A Utilty Proof of Section 5.2

We give the formal proof of Theorem 5.3, which is similar to Clarkson and Woodruff [15] modulo the analysis to consider the lift of singular value. We include it for the sake of completion. Let $\tilde{A} = U\tilde{\Sigma}V^{\mathsf{T}}$. 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 $\|U^{\mathsf{T}}\tilde{A}(\tilde{X} - X^*)\|$. For this, we first prove that $U^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega\tilde{A}(\tilde{X} - X^*)$ has a small norm. We have

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

This is because $U^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega(\tilde{A}\tilde{X} - B) = \tilde{A}^{\mathsf{T}}\Omega^{\mathsf{T}}\Omega(\tilde{A}\tilde{X} - B) = 0$. Therefore, from Theorem 5.1 with $\alpha' = \sqrt{\alpha/d}$ (since we chose r in Theorem 5.1 which differs by a factor of α and $1/d$ with respect to that in Theorem 5.3), we have

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

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

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

Using Lemma 5.4 and rearranging the terms, we get $\|U^{\mathsf{T}}\tilde{A}(\tilde{X} - X^*)\|_F \leq 2\sqrt{\alpha}\|B - AX^*\|_F + \sqrt{\tau}$. The utility proof is now immediate by observing that the column-space of A and U are the same, and the Pythagorus theorem on the norms. More concretely, with probability at least $1 - 2\beta$

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

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

B Differential Privacy of Variant 1

We give the proof Blocki *et al.* [6] for the sake of completion. We denote by $\bar{A}$ the matrix that differs from A by at most one entry. In other word, if A and $\bar{A}$ differs in row i, then there exists a unit vector v such that $A - \bar{A} = E = ve_i^{\mathsf{T}}$. Let $U\Sigma V^{\mathsf{T}}$ ($\bar{U}\Lambda\bar{V}^{\mathsf{T}}$, respectively) be the SVD of A ($\bar{A}$, respectively).

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

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

We prove the result for a row of the published matrix; the theorem follows from Theorem 2.1. It is straightforward to see that combination of the following proves differential privacy for a row of published matrix:

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

where

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

The first part of the proof follows simply as in Blocki *et al.* [6]. More concretely, we have $\tilde{\Delta}(A^\mathsf{T}A) = \prod_i \sigma_i^2$, where $\sigma_1 \ge \cdots \ge \sigma_d \ge \sigma_{\mathsf{min}}$ are the singular values of A. Let $\lambda_1 \ge \cdots \ge \lambda_d \ge \sigma_{\mathsf{min}}$ be its singular value for $\bar{A}$. Since the singular values of $A - \bar{A}$ and $\bar{A} - A$ are the same, $\sum_i(\sigma_i - \lambda_i) \le 1$ using Linskii's theorem. Therefore,

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

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 $\mathbb{E}[\Omega_i] = 0$ and $\mathsf{COV}(\Omega_i) = \mathbb{I}$. Let $\mathbf{y} = \Pi_i(\alpha)$. Then

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

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

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

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

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

Now since $\|\Lambda\|_2, \|\Sigma\|_2 \ge w$, plugging in the SVD of A and $A - A' = ve_i^\mathsf{T}$, 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.

$$\|(U\Sigma V^\mathsf{T})\Omega^\mathsf{T}(V\Sigma^{-1}U^\mathsf{T})e_i\|_2 \le \|\Omega^\mathsf{T}\|_2 \le 1, \qquad \|(U\Sigma V^\mathsf{T})\Omega^\mathsf{T}(V\Sigma^{-2}V^\mathsf{T})v\|_2 \le \|\Omega^\mathsf{T}\|_2\|\Sigma^{-1}\|_2 \le \frac{1}{\sigma_{\mathsf{min}}},$$
$$\|v^\mathsf{T}(\bar{V}\Lambda^{-2}\bar{V}^\mathsf{T})\Omega(\bar{V}\Lambda\bar{U}^\mathsf{T} - ve_i^\mathsf{T})\|_2 \le \|v^\mathsf{T}(\bar{V}\Lambda^{-2}\bar{V}^\mathsf{T})\Omega\bar{V}\Lambda\bar{U}^\mathsf{T}\|_2 + \|v^\mathsf{T}(\bar{V}\Lambda^{-2}\bar{V}^\mathsf{T})\Omega ve_i^\mathsf{T}\|_2 \le \frac{1}{\sigma_{\mathsf{min}}} + \frac{1}{\sigma_{\mathsf{min}}^2},$$
$$\|e_i^\mathsf{T}(\bar{U}\Lambda^{-1}\bar{V}^\mathsf{T})\Omega(\bar{V}\Lambda\bar{U}^\mathsf{T} - ve_i^\mathsf{T})\|_2 \le \|e_i^\mathsf{T}(\bar{U}\Lambda^{-1}\bar{V}^\mathsf{T})\Omega(\bar{V}\Lambda\bar{U}^\mathsf{T}\|_2 + \|e_i^\mathsf{T}(\bar{U}\Lambda^{-1}\bar{V}^\mathsf{T})\Omega ve_i^\mathsf{T}\|_2 \le 1 + \frac{1}{\sigma_{\mathsf{min}}},$$

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.