Partial Circulant Matrices and Differential Privacy

Jalaj Upadhyay *

Abstract

This paper resolves an open problem raised by Blocki *et al.* (FOCS 2012), i.e., whether other variants of the Johnson-Lindenstrauss transform preserves differential privacy or not? We modify and generalize the construction of of Vybiral (Journal of Functional Analysis: 260(4)), and show that this general method for constructing the Johnson-Lindenstrauss transform also preserves differential privacy, requires only n Gaussian samples and n Bernoulli trial, preserves privacy and allows matrix-vector multiplication in $O(n \log r)$ time, where $r \ll n$ is the dimension of the projection space. In this respect, this work unconditionally improves the run time of Blocki *et al.* (FOCS 2012) without using the graph sparsification trick of Upadhyay (ASIACRYPT 2013). For the metric of measuring randomness, we stick to the norm used by earlier researchers who studied variants of the Johnson-Lindenstrauss transform and its applications, i.e., count the number of random samples made. In concise, we improve the sampling complexity by quadratic factor, and the run time of cut queries by an $O(n^{o(1)})$ factor and that of covariance queries by an $O(n^{0.38})$ factor. As a positive effect of our modification to the basic transform of Vybiral, we also achieve a tighter dimension bound.

Our proof of both the privacy and utility guarantee uses several new ideas. In order to improve the dimension bound, for the utility proof, we use some known results from the domain of statistical model selection. This makes our proof short and elegant, relying just on one basic concentration inequality. For the differential privacy, even though our mechanism closely resembles that of Blocki *et al.* (FOCS 2012) and Upadhyay (ASIACRYPT 2013), we cannot use their proof idea. This is because even the modified version of Vybiral's projection matrix has non-trivial correlation between any two rows of published matrices, and, therefore, we cannot invoke the composition theorem of Dwork, Rothblum and Vadhan (STOC 2009). We argue that our published matrix is not r-multivariate distribution; rather one matrix-variate distribution. We compute the distribution of our published matrix and then prove it preserves differential privacy.

Keywords. Partial Circulant Matrices, Differential privacy, Sampling Complexity.

1 Introduction

In a recent work, Blocki *et al.* [9] proved that the Johnson-Lindenstrauss transform with random i.d.d. Gaussian normal variable preserves differential privacy, a very robust guarantee of privacy on database query. They left the question open whether other variants of the Johnson-Lindenstrauss transform, more specifically the fast Johnson-Lindenstrauss transform, the randomness efficient Johnson-Lindenstrauss, and the sparse Johnson-Lindenstrauss transform, preserves privacy or not? The focus of this work is a general class of the Johnson-Lindenstrauss transform, of which the only known construction of Vybiral [50] is a part, which allows fast matrix-vector multiplication and uses only linear (in the intrinsic dimension) random samples. We prove that it also preserves differential privacy under a slightly stronger condition; thereby, answering one of the open questions raised by Blocki *et al.* [9]. The transform of Vybiral [50] is based on a general class of matrices called *partial circulant matrices*[1], and achieves a suboptimal dimension reduction. We note that unless

*David R. Cheriton School of Computer Science University of Waterloo. email: jalaj.upadhyay@uwaterloo.ca

[1]Partial circulant matrices are a class of matrices indexed by a n-dimensional vector and formed as follows: the first row is the n-dimensional vector and the rest of the rows are formed iteratively by shifting the entries of the previous rows one position left.

there is a significant improvement in the concentration properties of the partial circulant matrices, one cannot improve the dimension bound achieved by Vybiral [50]. We make a slight modification to their transform (more specifically, by composing a Walsh-Hadamard transform matrix) to get an almost optimal dimension reduction. We then consider the general class of matrices of which partial circulant matrices are special instance. We prove the utility and privacy bound for this general class of matrices. Using this proof, we also give a simpler proof for the recent construction of Upadhyay [49].

One of the reasons Blocki *et al.* [9] perceived the study of other variants of the Johnson-Lindenstrauss important is due to their algorithmic and practical implications [8]. As argued in a series of work by Ailon and Liberty [2, 3, 4] and Krahmer and Ward [37], the dimension of the projected space, run time of the transform, and the number of random samples used by the transform are the most important parameters for a Johnson-Lindenstrauss transform from an algorithmic as well as the applications and implementation point of view. On the other hand, major focus of research in differential privacy until now has been towards providing a tight utility and privacy tradeoff. We sought to bring the resource consideration in the domain of differential privacy as well. Similar questions were also raised by Dwork *et al.* [22] and Upadhyay [48], where the focus was on the design of efficient mechanisms for answering cut-queries on sparse graphs. This paper forward that study of privacy preserving mechanisms while also taking in account the amount of randomness used.

One could argue that the number of random bits used is a more natural notion for considering the randomness complexity. However, in this paper, we stick with the norm used by researchers interested in the Johnson-Lindenstrauss transform and its applications, i.e., measure the randomness complexity in terms of the number of random samples. This measure has been used in the domain of dimension reduction and its application like compressed sensing [13], machine learning [5], quantum algorithms [17], and numerical analysis [16, 46].

Ailon and Liberty [2, 3, 4] and Krahmer-Ward [37] have thoroughly motivated why the number of random samples is an important parameter with respect to other applications of the Johnson-Lindenstrauss transform. Apart from all those reasons, one of the other main reasons for this choice in the domain of differential privacy is that it gives a much cleaner picture and a good quantitative estimate on the actual random bits used and the (actual) run-time of the mechanism[2] –it is the sampling process which is implemented in practice and might cause several issues. We refer readers to Kapralov and Talwar [33] for various theoretical and Chaudhary et al. [15] and Mironov [40] for practical issues faced during sampling. Moreover, sampling complexity also shed some light on the efficiency of the mechanism.

OUR TECHNIQUES. In this paper, we investigate whether we could improve the sampling complexity while preserving differential privacy. We answer this question affirmatively. We do this by making suitable modifications to a known transform by Vybiral [50] while maintaining its sampling complexity and efficiency in terms of matrix-vector multiplication. Moreover, we also give a sharper analysis to get a tighter bound than achieved by Vybiral [50] in terms of the dimension of the projected space. In concise, with this modification, we achieve a quadratic improvement in the sampling complexity of all known mechanisms for answering cut queries and covariance queries, and $\mathrm{poly}\log n$ improvement in the run time over the mechanism based on graph sparsification [48]. We follow up with the basic techniques used in our proofs.

OVERVIEW OF THE CONSTRUCTION. We start by giving a brief exposition of the construction by Vybiral [50]. Vybiral [50] first pick n random Gaussian samples to form the first row and then construct the rest of the $r-1$ rows by shifting the vector left-wise relative to the previous row. This matrix falls in the class of matrices known as *circulant matrices* and satisfies the *Restricted Isometry Property* [13], which we define next. For any set $T \subseteq \{1, \cdots, r\}$, we say that an $r \times n$ matrix Φ satisfies the *Restricted Isometry Property* of

[2]This could be seen in parallel to the complexity measure used in generic attacks on hash functions as well as concrete attacks on hash functions where we just measure the number of hash computations done, and not the actual atomic operations required (see for example, the attack on SHA-1 [51] and MD5 [52] and generic attacks [34, 32]).

order k if there exists an $0 < \varepsilon < 1$ such that, for all set T with $|T| < k$,

$$\mathsf{Pr}[(1-\varepsilon)\|\mathbf{x}_T\|_2^2 \leq \|\Phi_T \mathbf{x}_T\|_2^2 \leq (1+\varepsilon)\|\mathbf{x}_T\|_2^2] \geq 1-\eta \tag{1}$$

holds, where Φ_T ($\mathbf{x}_T$, respectively) is the restriction of Φ ($\mathbf{x}$, respectively) to the indices in T.

Rauhut *et al.* [43] have shown that a partial circulant matrices formed as above satisfies the Restricted Isometry Property for values of $r \geq \max(\varepsilon^{-1}\sqrt{(k\log n)^3}, \varepsilon^{-2} k \log^4 n)$. Vybiral then multiply this matrix P to a diagonal matrix formed by a Rademacher sequence, where each element is ± 1 with probability $1/2$. By Theorem 1, this construction satisfies the Johnson-Lindenstrauss bound for suboptimal value of $r = O(\varepsilon^{-2}\log^2 m)$, where m is the number of vectors on which the transform is to be applied.

Theorem 1. *(Krahmer-Ward [37, Proposition 3.2])* Let ε be an arbitrary constant. Let Φ be a matrix of order k and dimension $r \times n$ that satisfies the relation $k \leq c_1 \delta_k^2 r/\log(n/r)$ and equation (1). Then the matrix ΦD, where D is an $n \times n$ diagonal matrix formed by Rademacher sequence, is a Johnson-Lindenstrauss transform with r rows.

Therefore, unless there is a significant improvement in the concentration result on partial circulant matrices, the dimension bound achieved by Vybiral [50] is hard to beat. The key observation here is that the diagonal Rademacher matrices does not produce a proper "mixing" of the entries of the partial circulant matrices. For this, we need to compose it with a matrix that allows fast matrix-vector multiplication, preserves the Euclidean norm of the input vectors, and does not introduce more randomness. Our key observation is that, instead of using only a diagonal matrix formed by Rademacher sequence, if we also compose a Walsh-Hadamard matrix, then we can achieve a good enough mixing to allow a better concentration result[3]. This in turn helps us to strengthen the Vybiral's bound [50].

To generalize this construction, the key point to note is that partial circulant matrices are nothing special. They are simply the first r rows of a fully circulant matrix, and, therefore, can be seen as a result of applying a truncated permutation matrix from the left of a fully circulant matrix. We use the idea of Rudelson and Vershynin [45] and sample any r rows of a circulant matrix. This increases our sampling complexity by an additive factor of r. Therefore, at the cost of gross oversimplification, an intuitive way to see the general class of the Johnson-Lindenstrauss transform is as a hybrid of the known constructions of projection matrix with Restricted Isometry Property and known constructions of the Johnson-Lindenstrauss transform.

TECHNIQUES USED FOR THE UTILITY PROOF. The general idea to prove approximation result of the Johnson-Lindenstrauss lemma 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. For example, in the simplified proof of the Johnson-Lindenstrauss transform, the above method gives the failure bound of at most $1 - 1/n$. The idea is to then repeat the experiment required number of times to get a bound closer to any constant desired. We cannot rely on repetition because it would increase the random samples required. Therefore, we have to give a tighter bound. For this, we rely on a result from statistical model selection.

We break our analysis in two parts. We first use the isometry of Ailon and Chazelle [1], which precondition the input vectors to get a vector with bounded co-ordinates and same 2-norm. Then, we use this promise to prove that when we multiply a partial circulant matrix with Gaussian entries from the right, then $\tilde{\mathbf{x}}$ preserves the Euclidean norm with high probability. For this, we use known concentration inequalities from the area of model selection. Unlike the earlier results on randomness-efficient fast Johnson-Lindenstrauss transform that uses matrices satisfying the Restricted Isometry Property of optimal order, the proof in this paper is elementary and relies on basic concentration inequalities. This gives us a transform with slight increase in matrix-vector multiplication time. We rectify this by using a preconditioning matrix and show that this matrix also preserves the isometry property getting a final run time of $O(n \log r)$.

[3] This composed matrix is the isometry matrix of Ailon and Chazelle [1]

Method	Cut-queries	Covariance-Queries	Run-time	# Random Samples
Randomized Response [11]	$O(\sqrt{sn\log\kappa}/\varepsilon)$	$\tilde{O}(\sqrt{d\log\kappa}/\varepsilon)$	$\Theta(n^2)$	$O(n^2)$
Exponential [12, 39]	$O(n\log n/\varepsilon)$	$O(n\log n/\varepsilon)$	Intractable	$O(n^2)$
Multiplicative Weight [30]	$\tilde{O}(\sqrt{\lvert\mathcal{E}(\mathcal{G})\rvert\log\kappa}/\varepsilon)$	$\tilde{O}(d\sqrt{n\log\kappa}/\varepsilon)$	$O(n^2)$	$O(n^2)$
Johnson-Lindenstrauss [9]	$O(s\sqrt{\log\kappa}/\varepsilon)$	$O(\varepsilon^{-2}\log\kappa)$	$O(rn^2)$	$O(rn)$
Graph Sparsification [48]	$O(s\sqrt{\log\kappa}/\varepsilon)$	–	$\tilde{O}(n^2)$	$O(n^2)$
This paper	$O(s\sqrt{\log\kappa}/\varepsilon)$	$O(\varepsilon^{-2}\log\kappa)$	$O(n^2\log r)$	$2n+r$

Table 1: Comparison between our mechanism and other mechanism.

TECHNIQUES USED FOR THE PRIVACY PROOF. The proof of differential privacy is far more involved. It is tempting to assume that multiplying our projection matrix (because of the form it has) results in r multivariate Gaussian and proof of [9] can be applied. However, there are subtle correlation between two rows (or two columns) of the projection matrix. Therefore, applying our projection matrix to a private matrix does not yield n independent multivariate distribution, but rather one matrix-variate distribution. We first compute the matrix-variate distribution of the published matrix and then prove it is differentially private. Our proof uses various characterization of positive-definite matrices and Hermittian matrices. For this proof, we need to prove the concentration result for a χ^2-distribution with n degrees of freedom[4].

Using the above technique for privacy proof, we also give a simpler proof for the construction given by Upadhyay [49]. We recall that the author used a different isometry matrix than that of Ailon and Chazelle [1] for projection that preserves differential privacy. We give our proof for the original construction. Our proof involves reducing the proof of differential privacy for their original construction to that used in this paper.

One can also implement our mechanisms *as distributed algorithms*, a desirable feature as argued by [6]. This is because our mechanism uses operations that have efficient distributed algorithms. For example, one could use Jacobi method for SVD [36] and Cannon's algorithm for multiplication [14].

We summarize our results and its comparison with previous works in Table 1. The second, and third column is the noise added by the respective mechanisms. In the table, κ is the number of queries and r is the dimension of the projected space in our transform.

RELATED WORK. The first formal definition of Differential Privacy was given by Dwork et al. [21] to address the privacy concern of any participants. The key idea used in Dwork et al. [21] is to add noise according to a Laplace distribution to the output of a query; the Gaussian variant was proven to preserve differential privacy by Dwork et al. [20] in a follow-up work. Since then, many sanitizer for preserving differential privacy has been proposed in the literature, including the Exponential mechanism [12, 39], the Multiplicative Update mechanism [25, 26, 28, 30], the Median mechanism [44], the Boosting mechanism [23], and the Random Projection mechanism [35]. All these mechanisms have a common theme: they perturb the output before responding to queries. Blocki et al. [9, 10] and Upadhyay [47, 48] took a complementary approach. They perturb the input by performing a random projection of the input and show that existing algorithms preserves differential privacy if the input is perturbed in a reversible manner.

2 Preliminaries, Notations, and Basic Definitions

NOTATION. We fix the letter n to denote the space of the input vectors, m to denote the number of vectors, and r to denote the subspace to which the vectors are projected. We use the symbol ε to denote the approximation parameter in the statement of JL transform. We use bold letter face for vectors, for example $\mathbf{x}$. We

[4]The chi-squared distribution with n degrees of freedom is the distribution of a sum of the squares of n independent standard normal random variables

use the notation $\langle a_1, \cdots, a_n \rangle$ to denote the individual entries of an n-dimensional vector. We use the symbol W to denote an $n \times n$ discrete Fourier transform. We reserve the symbol Π to denote a random permutation matrix and $A_{1..r}$ to denote the matrix formed by taking the first r rows of A, $A_{i:}$ to denote i-th row and $A_{:j}$ to denote the j-th column of matrix A. We use Dirac notation to represent vectors, i.e., $\langle \cdot |$ to represent row vector and $| \cdot \rangle$ to represent a column vector. We use bold faced capital letters, like $\mathbb{A}$, to denote n copies of matrix A stacked together row-wise. For a vector $\mathbf{x}$, we use the notation $\mathsf{Diag}(\mathbf{x})$ to represent a diagonal matrix with non-zero entries $\langle x_1, \cdots, x_n \rangle$.

PRIVACY AND UTILITY. In this work, we deal with privacy-preserving mechanisms for various queries: cut-queries on a graph and directional covariance queries on a matrix. We also deal with the problem of publishing low rank approximation of a matrix. We work with the natural relaxed notion of differential privacy, known as *approximate differential privacy*, which requires us to definie *neighboring data-sets*. Two data-sets are *neighboring* if they differ on at most one entry.

Definition 1. A randomized mechanism, $\mathcal{K}$, gives (ε, δ)-*differential privacy*, if for all neighboring 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 lemma in our analysis in Section 3.

Lemma 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.

STATISTICAL MODEL SELECTION AND PROBABILITY THEORY. The main ingredients in our utility proof are inequalities from model selection. We review some of its basics and probability theory that are required to understand our proof. 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.

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 3. 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 3 is often used to control the expectation of the supremum of a finite family of exponentially integrable variables. Pisier [42] proved the following fundamental lemma.

Lemma 4. (Pisier [42]) Let $\{\zeta_f\}_{f \in F}$ be a finite family of random variables and ψ be as in Lemma 3. 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 5. 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 [38]. In this paper, we use the following result by Birge and Massart [7] for bounding the utility.

Theorem 6. (Birge-Massart [7]) 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 5 and Talagrand inequality, the authors also proved the following corollary to Theorem 6.

Corollary 7. (Birge-Massart [7]) 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 $\mathsf{Pr}\left[\zeta \geq \sqrt{2a\tau} + b\tau\right] \leq \exp(-\tau)$.

Linear Algebra and Gaussian Distribution. Our analysis of privacy makes extensive use of linear algebra and statistical properties of Gaussian distribution. We give an exposition to the level required to understand this paper. Let A be an $n \times d$ matrix. The singular value decomposition (SVD) of A is $A = V\Lambda U^{\mathsf{T}}$, where U, V are unitary matrices and Λ is a diagonal matrix consisting of 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 use standard Walsh-Hadamard matrix and discrete Fourier transform matrix. A order m Walsh-Hadamard matrix is $2^m \times 2^m$ matrix formed recursively as follows: $W_0 = 1$ and $W_m = \frac{1}{\sqrt{2}}\begin{pmatrix} W_{m-1} & W_{m-1} \\ W_{m-1} & -W_{m-1} \end{pmatrix}$. We use the symbol $\mathcal{F}$ to denote discrete Fourier transform. We assume that the private matrix has a dimension $n \times d$ where n is a power of 2. This is without any loss of generality because we can simply append block matrix of 0 to make the number of rows a power of 2 while incurring at most a constant overhead.

Given a random variable, X, we denote by $X \sim \mathcal{N}(\mu, \sigma^2)$ the fact that X is distributed according to a Gaussian distribution with the probability density function, $\mathsf{PDF}_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. The Gaussian distribution is invariant under affine transformation. This is called *spherical symmetry* of Gaussian variable. The multivariate Gaussian distribution is a generalization of univariate Gaussian distribution. Given a m dimensional multivariate random variable, $X \sim \mathcal{N}(\mu, \Sigma)$ with mean $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma = \mathbb{E}[(X-\mu)(X-\mu)^{\mathsf{T}}]$, the PDF of a multivariate Gaussian is given by $\mathsf{PDF}_{\mathbf{X}}(\mathbf{x}) := \frac{1}{\sqrt{2\pi\tilde{\Delta}(\Sigma)}} \exp\left(-\frac{1}{2}\mathbf{x}^{\mathsf{T}}\Sigma^{\dagger}\mathbf{x}\right)$. 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.

3 Partial Circulant Matrices and Differential Privacy

In this section, we show that the transformation of Vybiral [50], after slight modification, also preserves differential privacy and gives a Johnson-Lindenstrauss bound with same randomness complexity but with improved dimension bound at the slight depreciation of run-time. We then give a slight modification to reduce the run time to $O(n \log r)$, where r is the dimension of the projected space. Our first variant is a matrix $\Phi = PWD$, where W is a Walsh-Hadamard transform and gives a run-time of $O(n \log n)$. To improve the run time to $O(n \log r)$, we replace the Walsh-Hadamard transform by discrete Fourier transform. Though this change is very nominal, it makes the utility proof more complicated as we have to deal with the complex vectors instead of real vector. The mechanisms for answering cut-queries and covariance queries follows by substituting our projection matrix instead of the random Gaussian matrix based Johnson-Lindenstrauss transform used in Blocki *et al.* [9]. We can also compute the matrix multiplication and linear regression in the streaming model by substituting our projection matrix in the mechanism of Upadhyay [47].

Description of the Matrix P. Let $\alpha := \langle \alpha_1, \cdots, \alpha_n \rangle$ be n i.d.d. Gaussian samples. Let P be a matrix formed by permuting these entries by multiplying it with a set of permutation matrices. More

Construction of Φ: Construct the matrices D and P as below.

1. D is an $n \times n$ diagonal matrix such that $\mathsf{Pr}[D_{ii} = +1] = \mathsf{Pr}[D_{ii} = -1] = 1/2$.
2. P is a matrix with entries picked from a Gaussian distribution; however, the rows are permutation of each other as defined in Section 3.

Compute $\Phi = PWD$, where W is the normalized $n \times n$ Walsh-Hadamard transform matrix.

Figure 1: The Random Projection Operator

concretely, let $\Pi_1, \cdots \Pi_r$ be the set of distinct permutations on $[1..n]$. Then the row i of P are formed by permuting the entries in α according to the permutation Π_i, i.e., $P = \begin{pmatrix} \Pi_1(\alpha) & \Pi_2(\alpha) & \cdots & \Pi_r(\alpha) \end{pmatrix}^{\mathsf{T}}$.

It is also easy to show that for the random choice of permutations, the matrix P is full rank with high probability. However, this would require sampling and defeat our purpose of constructing a sampling-efficient mechanism. We get around this problem by specifying some known permutations, like Affine transformation (e.g., shift operation), or combinatorial designs, like r random rows of Latin squares. Latin squares are combinatorial design in which a square $n \times n$ matrices have entries from 1 to n such that every row and column have all the entries.

We first note few salient feature of our modified transform. The matrix P alone cannot be used for the random projection because for some bad input vector x, the estimate of $\|Px\|_2$ can be really bad. For example, when x is along a single coordinate, then only the non-zero values of P along this coordinate will contribute to Px, giving a very bad variance bound. For this, we need to precondition the input with WD. In non-technical terms, it allows us to mimic a projection matrix with every entries picked i.d.d. Note that the assumption of almost uniformly distributed data-base has been made in earlier works, like [44].

We recall that our construction is same as applying the isometry of Ailon-Chazelle [1] to the construction of Rauhut *et al.* [43]. The partial circulant matrix was proven to satisfy the Restricted Isometry Property with $r \geq \max(\varepsilon^{-1}\sqrt{(k \log n)^3}, \varepsilon^{-2} k \log^4 n)$ by Rauhut *et al.* [43]. By the theorem of Krahmer and Ward [37] (Theorem 1), it also satisfies the Johnson-Lindenstrauss bound for suboptimal dimension reduction. However, with an extra W, we manage to achieve a tighter bound. We conclude this section by giving the formal description of the projection matrix (Figure 1) and a proof of the utility (Theorem 11) and privacy guarantee (Theorem 8).

Theorem 8. *Privacy Guarantee.* Let Φ be a $n \times r$ projection matrix constructed by transposing the construction in Figure 1. If the singular values of an $n \times d$ matrix A is at least $(16n \ln(1/\delta))/\varepsilon$. Then for any private input matrix A, $A^{\mathsf{T}}\Phi$ preserves (ε, δ)-differential privacy. Moreover, the computation requires $O(nd \log r)$ basic operations.

A remark about the above theorem is due here. Note that we have a factor of n instead of a factor of r as in Blocki *et al.* [9]. However, as mentioned by the authors, in order to answer all cut (or covariance) queries, one actually need $r \geq n$. In other words, in order to answer all the cut-queries and covariance queries, the mechanism does not perform dimension reduction, rather it increases the dimension. In that respect, differential privacy is distinct from all other applications of the Johnson-Lindenstrauss transform. However, we need a quadratic higher requirement from the singular value than in Blocki *et al.* [9]. This is not surprising as we expect to pay the price of faster and randomness efficient computation in some or the other way.

Proof. Before we give our proof, we argue why the proof of Blocki *et al.* [9] does not extend to our case. One of the reasons why the proof of Blocki *et al.* [9] does not generalize to any Johnson-Lindenstrauss transform is its strong dependency on the fact that each samples in a dense Gaussian matrices is picked i.d.d. More con-

cretely, each row of their published matrix is a multivariate Gaussian and preserves differential privacy. They then invoke the composition theorem of Dwork, Rothblum and Vadhan [23] to prove differential privacy of the entire published matrix. Unfortunately, we cannot invoke the composition theorem. The reason why Blocki *et al.* [9] could invoke the composition theorem and we cannot is due to the independence of every row of their projection matrix. On the other hand, our resulting projection matrix reuses random samples. Therefore, applying our projection matrix to a private matrix does not yield n independent multivariate distribution, but rather one matrix-variate distribution. We compute the resulting distribution and then prove that it preserves differential privacy. In this sense, our proof uses the same idea as used by Upadhyay [47]; however, there are some subtle manner in which the proof here differs from the earlier proof, which we point out at suitable points in the proof.

One alternate way to look at the matrix P as a matrix formed by sampling first r rows of a fully circulant matrix. Let $\mathbb{I}_k$ denote the $k \times k$ identity matrix. For the ease of the presentation of the proof, we assume that our matrix P is formed by shift operator that shifts entries leftward, i.e.,

$$P = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_2 & \cdots & \alpha_n & \alpha_1 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_r & \alpha_1 & \cdots & \alpha_{r-1} \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbb{I}_r & 0 \end{pmatrix}}_{n \text{ columns}} \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_2 & \cdots & \alpha_n & \alpha_1 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \alpha_1 & \cdots & \alpha_{n-1} \end{pmatrix}. \tag{2}$$

Therefore, for the rest of this proof, we just concentrate on fully-circulant matrix and call it P. It is not difficult to show that any other fixed permutation also yields the same distribution. Let denote by $\mathsf{vec}(P)$ the vector formed by the entries of P. Then, we can write the covariance matrix of $\mathsf{vec}(P)$ as follows:

$$\Lambda := \mathsf{COV}(\mathsf{vec}(P)) = \underbrace{\begin{pmatrix} \mathbb{I}_{n/2} & 0 & 0 & \mathbb{I}_{n-1} & 0 & \mathbb{I}_{n-2} & \cdots & 0 & \mathbb{I}_1 \\ 0 & \mathbb{I}_{n/2} & \mathbb{I}_1 & 0 & \mathbb{I}_2 & 0 & \cdots & \mathbb{I}_{n-1} & 0 \\ 0 & \mathbb{I}_1 & \mathbb{I}_{n/2} & 0 & \cdots & \cdots & \cdots & \vdots & \vdots \\ \mathbb{I}_{n-1} & 0 & 0 & \mathbb{I}_{n/2} & & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbb{I}_{n-2} & \vdots & \vdots & \vdots & \ddots & \cdots & \mathbb{I}_{n/2+1} & 0 \\ \mathbb{I}_2 & 0 & \cdots & \cdots & \cdots & \cdots & \ddots & 0 & \mathbb{I}_{n/2-1} \\ 0 & \mathbb{I}_{n-1} & 0 & \mathbb{I}_{n-2} & \cdots & \cdots & \cdots & \mathbb{I}_{n/2} & 0 \\ \mathbb{I}_1 & 0 & \mathbb{I}_2 & 0 & \cdots & & \cdots & 0 & \mathbb{I}_{n/2} \end{pmatrix}}_{n^2 \text{ columns}} n^2 \text{ rows} \tag{3}$$

We first note that WD play no role in the output distribution. This is because of the spherical symmetry of a vector of Gaussian distribution. Therefore, without any loss of generality, we can analyze the distribution $A^\mathsf{T}P$ instead of $A^\mathsf{T}\Phi$. Also, note that $\|A^\mathsf{T} - \tilde{A}^\mathsf{T}\|_2 = \|(A^\mathsf{T} - \tilde{A}^\mathsf{T})(WD)^\mathsf{T}\|_2$; therefore, proving privacy of A^T is the same as proving privacy for $(AWD)^\mathsf{T}$.

In order to compute the PDF of the matrix-variate distribution corresponding to the published matrix, we follow the standard technique. We look at the published matrix as a vector and analyze the corresponding multivariate distribution. Recall that the published matrix is not an n independent multi-variate distribution, rather it is one matrix-variate distribution and there are non-trivial covariance between the entries of two rows of the published matrix. In Lemma 16, we prove that a covariance matrix is a positive semi-definite matrix; therefore, we can write equation (3) succinctly in form of its Cholesky decomposition, say $\Lambda = LL^\mathsf{T}$. This is the first point where we diverge from the proof of Upadhyay [47].

First note that $\det(\Lambda) = 1$; therefore, if $sv_1, \cdots sv_n$ are the singular values of Λ, then $\prod_i sv_i^2 = 1$. In other words, if the singular value decomposition of $A = U\Sigma V^\mathsf{T}$, then that of $A^\mathsf{T}\Phi$ is $V\Sigma U'^\mathsf{T}$ for $U' = (LU)^\mathsf{T}$.

Therefore, the singular values of $A^{\mathsf{T}}\Phi$ remains that of A, the left singular vectors are the orthonormal columns of V and right singular vectors are that of U'. Moreover, $\|U'\| \leq 1$. Using the left spherical symmetry of Gaussian distribution, and since the Jacobian of the transformation $A^{\mathsf{T}}\Phi$ is $\sqrt{\det A^{\mathsf{T}}A}$, the resulting matrix variate distribution for $X \sim A^{\mathsf{T}}Y$ for Y picked from a distribution with mean vector 0 and covariance matrix Λ has the covariance matrix $\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A}$, where $\mathbb{A}$ is matrix formed by stacking n copies of A row-wise. Therefore,

$$\mathsf{PDF}_{A^{\mathsf{T}}\Phi}(X) = \frac{1}{\sqrt{\det(\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A})}} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(X^{\mathsf{T}}(\mathbb{A}^{\mathsf{T}}\Lambda\mathbb{A})^{-1}X\right)\right). \tag{4}$$

For the sake of simplicity, let us denote by $\mathbb{B} = L^{\mathsf{T}}\mathbb{A}$ such that the singular value decomposition of B is $U'\Sigma V^{\mathsf{T}}$. Let $\tilde{B}$ be the corresponding matrix for the neighbouring matrix A with singular value decomposition $\tilde{U}'\tilde{\Sigma}\tilde{V}^{\mathsf{T}}$. Then from equation (4), we can write the distribution of the published matrices corresponding to neighbouring matrices A and $\tilde{A}$ as follows.

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

In order to prove the differential privacy, we prove the following lemma.

Lemma 9. For a matrix A with all singular values greater than $\frac{16r\log(4/\delta)}{\varepsilon}$, the following holds

$$\sqrt{\frac{\det(\mathbb{B}^{\mathsf{T}}\mathbb{B})}{\det(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})}} \in \exp(\pm\varepsilon) \tag{5}$$

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

$$\mathsf{Pr}\left[\left|\mathrm{Tr}\left(X^{\mathsf{T}}\left((\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1} - (\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}\right)X\right)\right| \leq \varepsilon\right] \geq 1-\delta \tag{6}$$

Proof. The first part of the proof follows simply as in Blocki *et al.* [9]. More concretely, we have $\det(\mathbb{B}^{\mathsf{T}}\mathbb{B}) = \left(\prod_i \sigma_i^2\right)n$, where $\sigma_1 \geq \cdots \geq \sigma_d \geq \sigma_{\mathsf{min}}$ are the singular values of B. Let $\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_d \geq \sigma_{\mathsf{min}}$ be its singular value for $\tilde{B}$. Since the singular values of $B - \tilde{B}$ and $\tilde{B} - B$ are the same, $\sum_i(\sigma_i - \tilde{\sigma}_i) \leq 1$ using Linskii's theorem. Therefore,

$$\frac{\det(\mathbb{B}^{\mathsf{T}}\mathbb{B})}{\det(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})} = \left(\prod_i \frac{\tilde{\sigma}_i^2}{\sigma_i^2}\right) \leq \exp\left(\frac{\varepsilon}{8\log(2/\delta)}\right)\sum_i(\tilde{\sigma}_i - \sigma_i) \leq \exp(\varepsilon).$$

Similarly, we can bound $\frac{\det(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})}{\det(\mathbb{B}^{\mathsf{T}}\mathbb{B})} \leq \exp(\varepsilon)$.

PROOF OF EQUATION (6). In this part, we bound the following expression.

$$\left|\mathrm{Tr}\left(X^{\mathsf{T}}\left(\left(\mathbb{B}^{\mathsf{T}}\mathbb{B}\right)^{-1} - \left(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}}\right)^{-1}\right)X\right)\right|. \tag{7}$$

We can write $\tilde{A} = A + |v\rangle\langle e_i|$ for some i and a unit vector v. Let E be a matrix formed by n-copies of $|e_i\rangle\langle v|$ stacked together. The following is immediate.

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

Using the singular value decomposition of $B = U'\Sigma V^{\mathsf{T}}$ and $\tilde{B} = \tilde{U}'\tilde{\Sigma}\tilde{V}^{\mathsf{T}}$, we first analyze $\mathbb{B}^{\mathsf{T}}\mathbb{B}$.

$$\mathbb{B}^{\mathsf{T}}\mathbb{B} = \begin{pmatrix} B^{\mathsf{T}} & \cdots & B^{\mathsf{T}} \end{pmatrix} \begin{pmatrix} B \\ \vdots \\ B \end{pmatrix} = nB^{\mathsf{T}}B = V^{\mathsf{T}}(n\Sigma)V.$$

Similarly, $\mathbb{B}^{\mathsf{T}}E = nB^{\mathsf{T}}(|e_i\rangle\langle v|) = nV\Sigma U'^{\mathsf{T}}|e_i\rangle\langle v|$, $\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}} = n\tilde{B}^{\mathsf{T}}\tilde{B}$, and $\mathbb{E}^{\mathsf{T}}\tilde{\mathbb{B}} = n|v\rangle\langle e_i|\tilde{U}'\tilde{\Sigma}\tilde{V}^{\mathsf{T}}$. Since $X = A^{\mathsf{T}}\Phi$, we can further solve the above expression.

$$\begin{aligned}
&\mathrm{Tr}\left(X^{\mathsf{T}}\left(\left(\mathbb{B}^{\mathsf{T}}\mathbb{B}\right)^{-1} - \left(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}}\right)^{-1}\right)X\right) = \mathrm{Tr}\left(X^{\mathsf{T}}\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}E + E^{\mathsf{T}}\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)X\right) \\
&\quad = \mathrm{Tr}\left(\Phi^{\mathsf{T}}A\left((\mathbb{B}^{\mathsf{T}}\mathbb{B})^{-1}(\mathbb{B}^{\mathsf{T}}E + E^{\mathsf{T}}\tilde{\mathbb{B}})(\tilde{\mathbb{B}}^{\mathsf{T}}\tilde{\mathbb{B}})^{-1}\right)A^{\mathsf{T}}\Phi\right) \\
&\quad = \mathrm{Tr}\left(\Phi^{\mathsf{T}}A\left(V(n\Sigma)^{-1}U'^{\mathsf{T}}|e_i\rangle\langle v|\tilde{V}(n\tilde{\Sigma})^{-2}\tilde{V}^{\mathsf{T}} + V(n\Sigma)^{-2}V^{\mathsf{T}}|v\rangle\langle e_i|\tilde{U}'(n\tilde{\Sigma})^{-1}\tilde{V}^{\mathsf{T}}\right)A^{\mathsf{T}}\Phi\right) \\
&\quad = \mathrm{Tr}\left(\Phi^{\mathsf{T}}\left(n^{-1}UU'^{\mathsf{T}}|e_i\rangle\langle v|\tilde{V}(n\tilde{\Sigma})^{-2}\tilde{V}^{\mathsf{T}} + V(n\Sigma)^{-2}V^{\mathsf{T}}|v\rangle\langle e_i|\tilde{U}'(n\tilde{\Sigma})^{-1}\tilde{V}^{\mathsf{T}}V\Sigma U^{\mathsf{T}}\right)\Phi\right) \\
&\quad = \mathrm{Tr}\left(\Phi\Phi^{\mathsf{T}}\left(n^{-1}UU'^{\mathsf{T}}|e_i\rangle\langle v|\tilde{V}(n\tilde{\Sigma})^{-2}\tilde{V}^{\mathsf{T}} + V(n\Sigma)^{-2}V^{\mathsf{T}}|v\rangle\langle e_i|\tilde{U}'(n\tilde{\Sigma})^{-1}\tilde{V}^{\mathsf{T}}V\Sigma U^{\mathsf{T}}\right)\right) \\
&\quad \leq \underbrace{\mathrm{Tr}\left(\Phi\Phi^{\mathsf{T}}\right)}_{S}\mathrm{Tr}\left(\underbrace{n^{-1}UU'^{\mathsf{T}}|e_i\rangle}_{B_1}\underbrace{\langle v|\tilde{V}(n\tilde{\Sigma})^{-2}\tilde{V}^{\mathsf{T}}}_{B_2} + \underbrace{V(n\Sigma)^{-2}V^{\mathsf{T}}|v\rangle}_{B_3}\underbrace{\langle e_i|\tilde{U}'(n\tilde{\Sigma})^{-1}\tilde{V}^{\mathsf{T}}V\Sigma U^{\mathsf{T}}}_{B_4}\right),
\end{aligned}$$

where the last inequality follows from the fact that $\mathrm{Tr}(XY) \leq \mathrm{Tr}(X)\mathrm{Tr}(Y)$ for Hermittian matrices X and Y. In fact, the two matrices in question are positive semi-definite. We bound each of the above trace terms. To bound S, we recall the fundamental relation between discrete Fourier transform and circulant matrices. Recall that P is made by circulating the Gaussian vectors $\langle \alpha_1, \cdots, \alpha_n \rangle$ to form a $r \times n$ matrix. Then

$$P = \mathcal{F}_n \mathsf{Diag}(\sqrt{n}\mathcal{F}_n\alpha)\mathcal{F}_n^{-1}. \tag{8}$$

Therefore, to bound the trace of $\Phi\Phi^{\mathsf{T}}$ or equivalently PP^{T}, we have to bound the following.

$$PP^{\mathsf{T}} = \mathcal{F}_n \mathsf{Diag}(\sqrt{n}\mathcal{F}_n\alpha)\mathcal{F}_n^{-1}\left[\mathcal{F}_n \mathsf{Diag}(\sqrt{n}\mathcal{F}_n\alpha)\mathcal{F}_n^{-1}\right]^{\mathsf{T}} = \mathcal{F}_n \mathsf{Diag}(n|\mathcal{F}_n\alpha|^2)\mathcal{F}_n^{-1}. \tag{9}$$

Since α and $Z\alpha$ are equidistributed when the rows of Z are orthonormal, we need the following lemma to bound S.

Lemma 10. 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. □

Now e_i and v are unit vectors and the singular values of $A, \tilde{A}$ is at least σ_{min}. We can easily bound B_1, B_2, B_3, and B_4 by $1, 1/\sigma_{\mathsf{min}}, 1/\sigma_{\mathsf{min}}+1/\sigma_{\mathsf{min}}^2$, and $1+1/\sigma_{\mathsf{min}}$, respectively. Using Lemma 10,

$$\mathsf{Pr}\left[(7)\leq\frac{4(1+\eta)n^2}{n\sigma_{\mathsf{min}}^2}\left(\frac{1}{\sigma_{\mathsf{min}}}+\frac{1}{\sigma_{\mathsf{min}}^2}\right)\ln(4/\delta_0)\leq 5\varepsilon\right]\geq 1-\delta.$$

Rescaling the value of ε, the lemma follows. □

It is straightforward to see that Lemma 9 implies Theorem 8. □

Theorem 11. *Utility Guarantee.* Let Φ be as in Figure 1. Then for any set of m vectors $\mathbf{x}\in\mathbb{R}^n$, the following holds with the probability at least $2/3$,

$$(1-\eta)\sqrt{r}\|\mathbf{x}\|_2^2\leq\|\Phi\mathbf{x}\|_2^2\leq(1+\eta)\sqrt{r}\|\mathbf{x}\|_2^2 \tag{10}$$

Proof. The usual idea in proving the concentration bound of Johnson Lindenstrauss transform 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 first proof the result when we use a Walsh-Hadamard matrix instead of discrete-Fourier transform.

We break the analysis in two parts. We first use the fact that WD is an isometry [1], i.e., for any vector $\mathbf{x}$ of unit length, $WD\mathbf{x}$ 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 of the characterization of partial circulant matrices.

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 follows simply from the following result by Ailon and Chazelle [1].

Theorem 12. *(Ailon-Chazelle [1], Wolff [53, Proposition 4.2])* Let $\mathbf{x}\in\mathbb{R}^n$ and $t>0$. Let W and D be as defined in our construction. Then, for any $\eta>0$, we have $\mathsf{Pr}\left[\|WD\mathbf{x}\|_\infty\geq\sqrt{2e/n}\log(2n/\eta)\|\mathbf{x}\|_2\right]\leq\eta$.

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 Φ to get the concentration result makes the proof very lengthy. We follow the approach taken by Upadhyay [49]. This is possible because of the nice representation of a partial circulant matrices by a discrete Fourier transform

matrix. We first get around the problem of dealing with the permutation matrices by making a substitution, i.e., for the matrix Φ and any vector $\mathbf{x} \in \mathbb{R}^n$,

$$\|\Phi \mathbf{x}\|_2 = \sum_{j=1}^{r} \left| \sum_{i=1}^{n} \alpha_{(i-j) \mod n} \langle x_i, x_i \rangle \right|^2 = \|Z\alpha\|_2,$$

with entries of Z are $z_{i,j} = (\Pi_{1..r})_{i:}(\mathsf{Diag}(\tilde{\mathbf{x}}))_{:j}$. The rest of the proof is very similar to Upadhyay [49]; we include it for the sake of completion.

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.

$$\begin{aligned}
\mathsf{Pr}_\alpha \left[\|\Phi \mathbf{x}\|_2^2 \geq (1+\varepsilon) \right] &= \mathsf{Pr}_\alpha \left[\|WZ\alpha\|_2^2 \geq (1+\varepsilon)r \right] = \mathsf{Pr}_\alpha \left[\|Z\alpha\|_2^2 \geq (1+\varepsilon)r \right] \\
&= \mathsf{Pr}_\alpha \left[\|U\Sigma V^{\mathsf{T}}\alpha\|_2^2 \geq (1+\varepsilon)r \right] = \mathsf{Pr}_\gamma \left[\|U\Sigma\gamma\|_2^2 \geq (1+\varepsilon)r \right] \\
&= \mathsf{Pr}_\gamma \left[\|\Sigma\gamma\|_2^2 \geq (1+\varepsilon)r \right]. \qquad (11)
\end{aligned}$$

In other words, if we can prove the concentration bound on $\sum \sigma_i^2 |\gamma_i|^2$, we are done. We use the Corollary 7 to Theorem 6, for which we need to find the function ψ corresponding to our case. For this, we work in two step process. Let $0 < \lambda < 1/2a$. We first give following two simple propositions.

Proposition 13. 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

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

□

Proposition 14. 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

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

Proof. 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 13 to a vector of Gaussian variables results in Proposition 14. More concretely, from the linearity of expectation, we have

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

□

Since $\Sigma = \mathsf{Diag}(\sigma_1, \cdots, \sigma_r)$ and $Z = U\Sigma V^T$. Then, since Gaussian distribution is invariant if we multiply with a matrix with orthonormal rows on the right, we can restate Proposition 14 as

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

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 7, we have

$$\mathsf{Pr}_\gamma\left[\sum_{i=1}^{r}\sigma_i^2\left(\gamma_i^2-1\right) \geq 2\|Z\|_\infty^2\tau + 2\|Z\|_2^2\sqrt{\tau}\right] \leq \exp(-\tau). \tag{12}$$

We need to estimate $\|Z\|_\infty$ and $\|Z\|_2$. This is where the guarantee on $\|\tilde{\mathbf{x}}\|_\infty$ is useful. Using Theorem 12, with probability $19/20$, we have

$$\max_{\|\mathbf{x}\|_2=1}\|WD\mathbf{x}\|_\infty = \sqrt{2n^{-1/2}\log(40n)}.$$

Consequently, from the symmetry of the matrices, we have

$$\|Z\|_\infty^2 = \max_{\mathbf{x}\in\mathbb{R}^n,\|\mathbf{x}\|_2=1}\|Z\mathbf{x}\|_2^2 \leq n\|WD\mathbf{x}\|_\infty^2 = n\|\tilde{\mathbf{x}}\|_\infty^2 = 2\log(40n). \tag{13}$$

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

$$\|Z\|_2^2 \leq \|Z\|_F\cdot\|Z\|_\infty = \sqrt{2r\log(40n)}. \tag{14}$$

Since $\sum_j \sigma_j^2 = r$, by setting $\tau = cr\varepsilon^2/\log(40n)$ for a small constant c, and using equations (11), (12), (13), and (14), we have

$$\mathsf{Pr}_\alpha[\|\Phi\mathbf{x}\|_2^2 \geq (1+\varepsilon)] = \mathsf{Pr}_\gamma\left[\sum_{j=1}^{r}\sigma_j^2(|\gamma_j|-1) \geq \varepsilon r\right] < \exp\left(-\frac{r\varepsilon^2}{\log(40n)}\right). \tag{15}$$

For $(15) < 1/10m$, 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

$$\mathsf{Pr}_\alpha[\|\Phi\mathbf{x}\|_2^2 \leq (1-\varepsilon)] = \mathsf{Pr}_\gamma\left[\sum_{j=1}^{r}\sigma_j^2(|\gamma_j|-1) \leq -\varepsilon r\right] < \frac{1}{6m}. \tag{16}$$

Combining equation (15) and equation (16) and using union bound over all $x \in \mathcal{S}$, the result follows. □

As we mentioned in Section 1, our proof extends to give a simpler privacy proof for the projection matrix given by Upadhyay [49]. We first recall their construction.

CONSTRUCTION 2. Let D be a diagonal Bernoulli matrix and M be a diagonal Gaussian matrix. Let Π and Π' be a permutation matrix. Then Upadhyay [49] showed that $\Pi_{1..r}M\Pi W D$ satisfies the Johnson-Lindenstrauss bound, where $\Pi_{1..r}$ is a matrix restricted to the first r rows of a permutation matrix Π. We follow up with the details as to how we can mould the proof of Theorem 8 to the case of Construction 2.

Theorem 15. Let Φ be a $n \times r$ projection matrix constructed by transposing the matrix of Construction 2. If the singular values of an $n \times d$ matrix A is at least $(16n\ln(1/\delta))/\varepsilon$. Then for any private input matrix A, $A^\mathsf{T}\Phi$ preserves (ε,δ)-differential privacy. Moreover, the computation requires $O(nd\log r)$ basic operations.

Proof. Our proof reduces the problem of proving privacy for Construction 2 to that for Theorem 8. First note that for two neighbouring matrices A and A', $\|A-\tilde{A}\| = \|WD(A-\tilde{A})\| \leq 1$. Also, note that $A^\mathsf{T}\Pi_{1..r}$ and $\tilde{A}\Pi_{1..r}$ differs by at most one row by a unit entry depending on whether $\Pi_{1..r}$ picks that row or not. Therefore, $\|(A^\mathsf{T}-\tilde{A}^\mathsf{T})\Pi_{1..r}\| \leq 1$ given that $\|A-\tilde{A}\| \leq 1$. Moreover, for discrete Fourier transform $\mathcal{F}_n$, we also have $\|(A^\mathsf{T}-\tilde{A}^\mathsf{T})\Pi_{1..r}\mathcal{F}_n\| \leq 1$ given that $\|(A^\mathsf{T}-\tilde{A}^\mathsf{T})\Pi_{1..r}\| \leq 1$ because $\|\mathcal{F}_n\| \leq 1$. Let

$B = A^{\mathsf{T}}\Pi_{1..r}\mathcal{F}_n$ and $\tilde{B} = \tilde{A}\Pi_{1..r}\mathcal{F}_n$. Therefore, proving Theorem 15 reduces to that of proving that for $B^{\mathsf{T}}\mathcal{F}_n^{\mathsf{T}}M\Pi WD$. Also, from Lemma 2, we have that proving privacy of $B^{\mathsf{T}}\mathcal{F}_n^{\mathsf{T}}M\Pi WD$ is equivalent to proving privacy for $B^{\mathsf{T}}\mathcal{F}_n^{\mathsf{T}}M\mathcal{F}_n$. Now recall that M is $\mathsf{Diag}(\alpha)$ for Gaussian vector α. Therefore, M is distributed equivalent to $\mathsf{Diag}(\mathcal{F}_n\alpha)$. In other words, proving privacy for $B^{\mathsf{T}}\mathcal{F}_n^{\mathsf{T}}M\mathcal{F}_n$ is equivalent to proving privacy for $B^{\mathsf{T}}\mathcal{F}_n^{\mathsf{T}}\mathsf{Diag}(\mathcal{F}_n\alpha)\mathcal{F}_n$. Using equation (8), Theorem 15 follows. □

4 Applications

5 Conclusion and Future Works

In this paper, we introduced a new linear operator that allows us to achieve the same guarantee in terms of run time, utility, and differential privacy as in [9, 48], which is by far the best compared to other mechanisms (see [9, 48] or Table 1 for more details), with a quadratic improvement in the sampling complexity. Building on the work of [27, 29], we use simple linear algebra to give an improved mechanism for finding low rank approximation that sanitizes the private matrix only once.

This work leaves several open questions. Of particular interest is whether sparse JL transform preserves differential privacy or not. Few constructions of sparse-JL transforms, like [18], use linear sampling to achieve sparsity. Since bottle-neck of our mechanisms i is matrix multiplications, sparsity would improve the run-time of the mechanisms. An interesting problem relates to the problem of error amplification. The question is whether we can introduce some error-correction techniques to the problem? Any positive result in this direction would help reduce the additive error.

In the context of this work, one major open problem is to find a non-trivial lower bounds on the sampling complexity and multiplicative noise. There are tight lower bounds known for the sampling complexity of the samplers [24] and for additive noise in a differentially-private mechanisms [19, 31, 33]. We believe we could use some ideas from these lower bound results to give lower bounds on the sampling complexity and multiplicative noise of differentially private mechanisms. Any such lower bounds, even in the non-interactive setting, would help in our understanding of the gap, if any, between traditional privacy and differential privacy.

References

[1] Nir Ailon and Bernard Chazelle. The Fast Johnson–Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009. 3, 4, 7, 11

[2] Nir Ailon and Edo Liberty. Fast dimension reduction using Rademacher series on dual BCH codes. In *SODA*, pages 1–9, 2008. 2

[3] Nir Ailon and Edo Liberty. Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes. *Discrete & Computational Geometry*, 42(4):615–630, 2009. 2

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

[5] Richard G Baraniuk and Michael B Wakin. Random projections of smooth manifolds. *Foundations of computational mathematics*, 9(1):51–77, 2009. 2

[6] 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

[7] Lucien Birgé and Pascal Massart. *From Model Selection to Adaptive Estimation*, pages 55–87. Springer New York, 1997. 5, 6

[8] Jeremiah Blocki. Personal communication. 2013. 2

[9] Jeremiah Blocki, Avrim Blum, Anupam Datta, and Or Sheffet. The Johnson-Lindenstrauss Transform Itself Preserves Differential Privacy. In *FOCS*, pages 410–419. IEEE Computer Society, 2012. 1, 2, 4, 6, 7, 8, 9, 14

[10] 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. 4

[11] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the sulq framework. In Chen Li, editor, *PODS*, pages 128–138. ACM, 2005. 4

[12] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to noninteractive database privacy. *J. ACM*, 60(2):12, 2013. 4

[13] 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. 2

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

[15] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *NIPS*, pages 998–1006, 2012. 2

[16] Kenneth L. Clarkson and David P. Woodruff. Numerical linear algebra in the streaming model. In Mitzenmacher [41], pages 205–214. 2

[17] Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and Limits of Nonlocal Strategies. In *IEEE Conference on Computational Complexity*, pages 236–249. IEEE Computer Society, 2004. 2

[18] Anirban Dasgupta, Ravi Kumar, and Tamás Sarlós. A sparse Johnson: Lindenstrauss transform. In *STOC*, pages 341–350, 2010. 14

[19] Anindya De. Lower bounds in differential privacy. In *Theory of Cryptography*, pages 321–338. Springer, 2012. 14

[20] 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. 4

[21] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In *TCC*, pages 265–284, 2006. 4

[22] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil P. Vadhan. On the complexity of differentially private data release: efficient algorithms and hardness results. In Mitzenmacher [41], pages 381–390. 2

[23] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. Boosting and Differential Privacy. In *FOCS*, pages 51–60, 2010. 4, 8

[24] Oded Goldreich. A sample of samplers: A computational perspective on sampling. *def*, 1:2n, 1997. 14

[25] 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. 4

[26] Anupam Gupta, Aaron Roth, and Jonathan Ullman. Iterative constructions and private data release. In *TCC*, pages 339–356, 2012. 4

[27] 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. 14

[28] Moritz Hardt, Katrina Ligett, and Frank McSherry. A Simple and Practical Algorithm for Differentially Private Data Release. In *NIPS*, pages 2348–2356, 2012. 4

[29] Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. In *STOC*, pages 1255–1268, 2012. 14

[30] Moritz Hardt and Guy N. Rothblum. A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis. In *FOCS*, pages 61–70, 2010. 4

[31] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *STOC*, pages 705–714, 2010. 14

[32] Antoine Joux. Multicollisions in iterated hash functions. application to cascaded constructions. In *Advances in Cryptology–CRYPTO 2004*, pages 306–316. Springer, 2004. 2

[33] Michael Kapralov and Kunal Talwar. On differentially private low rank approximation. In *SODA*, volume 5, page 1. SIAM, 2013. 2, 14

[34] John Kelsey and Bruce Schneier. Second preimages on n-bit hash functions for much less than 2 n work. In *Advances in Cryptology–EUROCRYPT 2005*, pages 474–490. Springer, 2005. 2

[35] Krishnaram Kenthapadi, Aleksandra Korolova, Ilya Mironov, and Nina Mishra. Privacy via the johnson-lindenstrauss transform. *arXiv preprint arXiv:1204.2606*, 2012. 4

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

[37] 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. 2, 3, 7

[38] Pascal Massart and Jean Picard. *Concentration inequalities and model selection*, volume 1896. Springer, 2007. 5

[39] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 94–103. IEEE, 2007. 4

[40] Ilya Mironov. On significance of the least significant bits for differential privacy. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 650–661. ACM, 2012. 2

[41] Michael Mitzenmacher, editor. *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*. ACM, 2009. 15

[42] 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. 5

[43] Holger Rauhut, Justin K. Romberg, and Joel A. Tropp. Restricted Isometries for Partial Random Circulant Matrices. *CoRR*, abs/1010.1847, 2010. 3, 7

[44] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *STOC*, pages 765–774, 2010. 4, 7

[45] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008. 3

[46] 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

[47] J. Upadhyay. Differentially Private Linear Algebra in the Streaming Model. *ArXiv e-prints*, September 2014. 4, 6, 8

[48] Jalaj Upadhyay. Random Projections, Graph Sparsification, and Differential Privacy. In *ASIACRYPT (1)*, pages 276–295, 2013. 2, 4, 14

[49] Jalaj Upadhyay. Random Projection, Restricted Isometry Property, and More. *Submitted to ITCS*, 2015. 2, 4, 11, 12, 13

[50] Jan Vybíral. A variant of the Johnson–Lindenstrauss lemma for circulant matrices. *Journal of Functional Analysis*, 260(4):1096–1105, 2 2011. 1, 2, 3, 6

[51] Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full sha-1. In *Advances in Cryptology–CRYPTO 2005*, pages 17–36. Springer, 2005. 2

[52] Xiaoyun Wang and Hongbo Yu. How to break md5 and other hash functions. In *Advances in Cryptology–EUROCRYPT 2005*, pages 19–35. Springer, 2005. 2

[53] P. Wolff. On randomness reduction in the Johnson-Lindenstrauss lemma. *ArXiv e-prints*, February 2012. 11

A Deferred Proof

Lemma 16. 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

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

Now the quantity under the summation is of form $\sum\sum x_ix_jz_iz_j = (\mathbf{x}^{\mathsf{T}}\mathbf{z}) \geq 0$. 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, $\Sigma \in S^n_{++}$. □