Analysis of a Privacy-preserving PCA Algorithm using Random Matrix Theory

Lu Wei*, Anand D. Sarwate†, Jukka Corander‡, Alfred Hero§, and Vahid Tarokh¶

*Department of Electrical and Computer Engineering, University of Michigan-Dearborn, USA (luwe@umich.edu)
†Department of Electrical and Computer Engineering, Rutgers University, USA (anand.sarwate@rutgers.edu)
‡Department of Biostatistics, University of Oslo, Norway (jukka.corander@medisin.uio.no)
§Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, USA (hero@eecs.umich.edu)
¶School of Engineering and Applied Sciences, Harvard University, USA (vahid@seas.harvard.edu)

***Abstract*—To generate useful summarizations of data while maintaining privacy of sensitive information is a challenging task, especially in the big data era. The privacy-preserving principal component algorithm proposed in [1] is a promising approach when a low rank data summarization is desired. However, the analysis in [1] is limited to the case of a single principal component, which makes use of bounds on the vector-valued Bingham distribution in the unit sphere. By exploring the non-commutative structure of data matrices in the full Stiefel manifold, we extend the analysis to an arbitrary number of principal components. Our results are obtained by analyzing the asymptotic behavior of the matrix-variate Bingham distribution using tools from random matrix theory.**

I. Introduction

With the advent of large-scale data collection, being able to generate useful summarizations of the data while maintaining privacy is a challenging task. When data is collected about individuals, privacy becomes a crucial consideration. In particular, machine learning and data mining algorithms are sensitive to the risk of disclosing private information. Dimension reduction plays an important role in the preprocessing stage of various machine learning algorithms. Reducing the extrinsic dimension (say d) of data to its intrinsic dimension (say k, with $k < d$) identifies good low-dimensional representations that facilitate learning, classification, and data mining. One of the classical and most useful approaches for dimension reduction is Principal Component Analysis (PCA), which uses the singular value decomposition to find a low-dimensional subspace of a data matrix.

The recently proposed differentially private PCA algorithm in [1] offers a framework for exploring the tradeoff between privacy and utility of low rank data representations. This algorithm guarantees differential privacy [2], which is a strong, cryptographically-motivated definition of privacy that has recently received significant research attention. Algorithms for differentially private PCA differ in, among other properties, the privacy guarantees they make, i.e. ϵ-differential privacy versus (ϵ, δ)-differential privacy, where the latter guarantee is weaker. The first dimension-reduction algorithm for PCA was the SULQ method [3], which guaranteed (ϵ, δ)-differential privacy by adding a Gaussian noise to the data covariance matrix. In [4], an improved algorithm was proposed with optimal utility guarantees for a general intrinsic dimension k by adding symmetric Gaussian noise was proposed. Empirically and theoretically, the algorithm of [4] outperforms most (ϵ, δ) methods [5], [6]. The authors in [5] recently studied an (ϵ, δ)-private approach for approximating the covariance with applications to regression that involved adding a Wishart noise. In an independent work, [6] proposed an ϵ-private Wishart-noise addition method and demonstrated that this method has favorable empirical performance among ϵ-private methods.

As an alternative to adding noise, a different approach to private PCA is to produce a randomized estimate of the PCA subspace [1], [7]–[9]. In this direction, [1] proposed an ϵ-private method for choosing the PCA subspace by the exponential mechanism [10], which corresponds to sampling from the matrix Bingham distribution [11]. The proposed algorithm was studied in the special case of one principal component, i.e. $k = 1$, where the performance analysis for an arbitrary k was left open in [1]. In this paper, we continue the study of the privacy-preserving PCA algorithm proposed in [1]. We extend the analysis to an arbitrary number of principal components. Firstly, an exact integral representation of the performance metric is obtained, which involves the normalization constant of the matrix-variate Bingham distribution. Using random matrix theory, the asymptotic normalization of the Bingham distribution is derived, which leads to the limiting behavior of the performance metric. The key idea is to interpret the normalization constant as the moment generating function of linear spectral statistics of a certain random matrix ensemble. Closed-form expressions for asymptotic performance are obtained by expanding the generating function in terms of moments, which we calculate explicitly. The derived asymptotic results take simple forms and are reasonably accurate in regimes of practical interest.

II. Privacy-preserving PCA

The differentially private PCA algorithm of [1] consists of sampling from the matrix Bingham distribution. In their model, the algorithm takes as input a data matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$

containing d-dimensional data vectors (columns) corresponding to n individuals, where each column $\mathbf{x}_i$ satisfies $\|\mathbf{x}_i\| \leq 1$. As a result, the Frobenius norm of the data covariance matrix[1]

$$\mathbf{A} = \frac{1}{n}\mathbf{X}\mathbf{X}' \tag{1}$$

is at most one, i.e. $||\mathbf{A}||_{\mathrm{F}} \leq 1$. In order to find a low-rank approximation to the data covariance $\mathbf{A}$, we perform eigenvalue decomposition $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, where $\mathbf{\Lambda}$ is a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_d)$ of the ordered eigenvalues $0 \leq \lambda_d \leq \cdots \leq \lambda_1 < \infty$ of $\mathbf{A}$. Assume that the intrinsic dimension is k with $k < d$ and denote by $\mathbf{V}_k$ the subspace of $\mathbf{V}$ corresponding to the eigenvectors of the top k eigenvalues.

For a privacy-preserving matrix $\widehat{\mathbf{V}}_k$, which can be considered as a randomly perturbed version of $\mathbf{V}_k$, a natural measure on the quality of $\widehat{\mathbf{V}}_k$ in approximating $\mathbf{V}_k$ is

$$\mathrm{tr}\left(\widehat{\mathbf{V}}_k'\mathbf{A}\widehat{\mathbf{V}}_k\right). \tag{2}$$

This function is maximized when $\widehat{\mathbf{V}}_k = \mathbf{V}_k$ for any given k. The random matrix $\widehat{\mathbf{V}}_k$ belongs to the Stiefel manifold, i.e. $\widehat{\mathbf{V}}_k \in V_{d,k}$ with $\widehat{\mathbf{V}}_k'\widehat{\mathbf{V}}_k = \mathbf{I}_k$. By the same line of argument of the case $k = 1$ in [1, Th. 6], it follows that sampling from the matrix Bingham distribution is differentially private [10] for any k. Moreover, it holds that

$$\phi_{\mathrm{L}} := \sum_{i=d-k+1}^{d} \lambda_i \leq \mathrm{tr}\left(\widehat{\mathbf{V}}_k'\mathbf{A}\widehat{\mathbf{V}}_k\right) \leq \sum_{i=1}^{k} \lambda_i := \phi_{\mathrm{U}}, \tag{3}$$

where the lower and upper bounds are achieved when $\widehat{\mathbf{V}}_k$ consist of k eigenvectors corresponding to the last k and first k eigenvalues of $\mathbf{A}$, respectively. As a result, the parametrization of level sets

$$\mathcal{S}_\tau = \left\{\widehat{\mathbf{V}}_k : \mathrm{tr}\left(\widehat{\mathbf{V}}_k'\mathbf{A}\widehat{\mathbf{V}}_k\right) < \tau\left(\phi_{\mathrm{U}} - \phi_{\mathrm{L}}\right) + \phi_{\mathrm{L}}\right\} \tag{4}$$

implies that $\tau \in [0, 1]$. By the well-known exponential mechanism [10], the probability density of the quality function is the matrix variate Bingham distribution[2] [1], [11]

$$Q\left(\mathbf{U}, \mathbf{A}\right) = \frac{1}{D(\mathbf{A})}\mathrm{etr}\left(\mathbf{U}'\mathbf{A}\mathbf{U}\right), \qquad \mathbf{U}'\mathbf{U} = \mathbf{I}_k, \tag{5}$$

where the normalization constant $D(\mathbf{A})$ is obtained as [12]

$$D(\mathbf{A}) = \int_{\mathbf{U}\in V_{d,k}} \mathrm{etr}\left(\mathbf{U}'\mathbf{A}\mathbf{U}\right)(\mathrm{d}\mathbf{U}) = {}_1F_1\left(\frac{k}{2}, \frac{d}{2}, \mathbf{A}\right). \tag{6}$$

Here, ${}_1F_1\left(k/2, d/2, \mathbf{A}\right)$ denotes the confluent hypergeometric function of matrix argument.

The matrix Bingham distribution is a natural generalization of the vector Bingham distribution ($k = 1$) for modeling data from unit sphere (a.k.a. directional statistics) to Stiefel manifold (a.k.a. orientation statistics). The matrix Bingham distribution, while appealing from a structural view, presents several computational and analytic challenges. The Stiefel manifold is a not a convex set, which precludes using standard arguments used to characterize densities on the manifold. While the shape of the distribution has an easy interpretation, the density itself is difficult to compute. Specifically, efficiently computing the density's normalizing constant is known to be difficult [13], [14]. Sampling from the distribution is also challenging. The only available sampler uses Gibbs sampling [15], which is without an analytically justified burn-in time. Moreover, such a sampler only provides approximate sampling.

We say that an algorithm provides a (τ, η)-close approximation to $\mathbf{V}_k$ if

$$\mathbb{P}\left(\bar{\mathcal{S}}_\tau\right) \geq 1 - \eta, \tag{7}$$

where

$$\mathbb{P}\left(\mathcal{S}_\tau\right) = \int_{\mathcal{S}_\tau} Q\left(\widehat{\mathbf{V}}_k, \frac{n\epsilon}{2}\mathbf{A}\right)(\mathrm{d}\widehat{\mathbf{V}}_k) \tag{8}$$

and $\bar{\mathcal{S}}_\tau$ is the complement of $\mathcal{S}_\tau$. The presence of the factor $n\epsilon/2$ is to cancel the normalization $1/n$ in the data covariance (1) as well as to compensate the change of $\exp(2\epsilon)$ in the density as a result of the change of a single user's data by the exponential mechanism [10, Eq. (5)]

To obtain explicit expressions of (8), one can derive approximations to the hypergeometric function (6) in the case $\|\mathbf{A}\|_{\mathrm{F}} \leq 1$. More importantly, the central question posed by [1] is to characterize the sample complexity. Namely, to determine the number of data points n needed for the algorithm to be a (τ, η)-close and ϵ-differentially private approximation to the top-k PCA subspace. However, the proposed sample complexity lower bound in [1, Th. 7] is limited to the special case $k = 1$. In the next section, we extend the analysis in [1] to $k > 1$ by deriving a sample complexity lower bound.

III. Performance Analysis via Random Matrix Theory

This section presents the main technical contributions of this paper. Firstly, we reduce the matrix integral (8) that captures the utility of the produced subspace to a scalar integral. Then, we derive an asymptotic normalization constant of the matrix-variate Bingham distribution. Finally, the proposed asymptotic result is utilized to construct a sample complexity lower bound valid for an arbitrary number of principal components. The proofs of the results in this section are omitted due to the space limitation.

A. An Exact Representation

The first result is an exact scalar integral representation of the matrix-variate integral (8).

Proposition 1. *The integral over Stiefel manifold* (8) *admits the following one-dimensional integral representation*

$$\mathbb{P}(\mathcal{S}_\tau) = \frac{\left(D(\beta\mathbf{\Lambda})\right)^{-1}}{2\pi}\int_{-\infty}^{\infty}\frac{\imath}{\nu}\mathrm{e}^{\imath\nu\frac{\phi_{\mathrm{L}}}{\phi_{\mathrm{U}}-\phi_{\mathrm{L}}}}\left(1 - \mathrm{e}^{\imath\tau\nu}\right)D(\mathbf{\Theta})\,\mathrm{d}\nu, \tag{9}$$

where $\beta = n\epsilon/2$ and the matrix-variate hypergeometric function $D(\cdot)$ is defined in (6). Here, $\mathbf{\Theta}$ is a diagonal matrix

[1] $(\cdot)^\dagger$ denotes the matrix transpose operation.
[2] $\mathrm{etr}(\cdot) = \mathrm{e}^{\mathrm{tr}(\cdot)}$ denotes exponential of trace.

with the i-th diagonal entry

$$\theta_i = \left(\beta - \frac{\imath\nu}{\phi_{\mathrm{U}} - \phi_{\mathrm{L}}}\right)\lambda_i, \qquad i = 1, \ldots, d. \tag{10}$$

In principle, Proposition 1 allows us to compute the exact probability $\mathbb{P}(\mathcal{S}_\tau)$ via, e.g. the software package developed in [16] for hypergeometric functions of matrix argument. In practice, an exact evaluation of the matrix-variate hypergeometric functions in (9) that involve a slow converging infinite sum over partitions is rather difficult [14], [16]. This motivates the search for simple and accurate closed-form approximations to the hypergeometric functions when the eigenvalues of the matrix arguments are small.

B. Approximating the Matrix-variate Hypergeometric Function

We define two asymptotic regimes for our analysis. The asymptotics we consider are as d (and possibly k) tend to infinity. Such scaling results are common in the random matrix theory literature and allow us to use those tools to analyze the Bingham distribution in high dimensions. We further show empirically in Section IV that our results are reasonably tight for more modest d and k. The asymptotic regime $\mathfrak{D}_1$ corresponds to a scenario when all eigenvalues of $\mathbf{A}$ approach each other as $d \to \infty$. An example of $\mathfrak{D}_1$ is the case where $|\lambda_i - \lambda_j| \leq 1/d$, $\forall i, j$, $i \neq j$. The asymptotic regime $\mathfrak{D}_2$ corresponds to the case where $\mathfrak{D}_1$ holds and in addition the target dimension $k \to \infty$ as well.

It can be seen from (9) that the key to analysis of the asymptotic behavior of $\mathbb{P}(\mathcal{S}_\tau)$ is specification of an asymptotic form for the normalization constant (6) of matrix Bingham distribution. Such a result is given in the following lemma.

Lemma 1. *In the regime $\mathfrak{D}_1$ as defined above, we have*

$$ {}_1F_1\left(\frac{k}{2}, \frac{d}{2}, \mathbf{A}\right) \overset{\mathfrak{D}_1}{=} \exp\left(\frac{k\sum_{i=1}^d \lambda_i}{d} - \frac{k(d-k)\left(\sum_{i=1}^d \lambda_i\right)^2}{d^2(d+2)(d-1)} + \frac{k(d-k)\sum_{i=1}^d \lambda_i^2}{d(d+2)(d-1)}\right) \tag{11}$$

for the confluent hypergeometric function of matrix argument.

The asymptotic approximation (11) to the normalization constant of matrix-variate Bingham distribution seems new. In the special case $k = 1$, an expansion of the distribution in the regime when all λ_i approach infinity was derived in [17]. A Laplace approximation for the distribution was proposed in [14], which, however, involves numerically solving a non-trivial algebraic equation for every k and d.

Based on Lemma 1, in the following we construct an accurate estimate of $\mathbb{P}(\mathcal{S}_\tau)$.

Proposition 2. *Under the asymptotic regime $\mathfrak{D}_1$, we have*

$$\mathbb{P}(\mathcal{S}_\tau) \overset{\mathfrak{D}_1}{=} \frac{1}{2}\left(\mathsf{erf}\left(\frac{\tau - b}{a}\right) + \mathsf{erf}\left(\frac{b}{a}\right)\right), \tag{12}$$

where

$$a = \frac{2}{\phi_{\mathrm{U}} - \phi_{\mathrm{L}}}\sqrt{\frac{k(d-k)}{d^2(d+2)(d-1)}\sum_{1 \leq i < j \leq d}(\lambda_i - \lambda_j)^2}, \tag{13}$$

$$b = \frac{1}{\phi_{\mathrm{U}} - \phi_{\mathrm{L}}}\Bigg(-\phi_{\mathrm{L}} + \frac{k}{d}\sum_{i=1}^d \lambda_i + \frac{2\beta k(d-k)}{d^2(d+2)(d-1)}\sum_{1 \leq i < j \leq d}(\lambda_i - \lambda_j)^2\Bigg), \tag{14}$$

and

$$\mathsf{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t \tag{15}$$

is Gauss error function.

Though Proposition 2 greatly simplifies the evaluation of the probability $\mathbb{P}(\mathcal{S}_\tau)$, analytically solving β (thus the sample complexity n) for a given $\mathbb{P}(\mathcal{S}_\tau)$ is still not possible. This is due to the presence of β in both error functions (12). To address this issue, some further asymptotic analysis is needed.

C. Asymptotic Sample Complexity of Private PCA

With the above preparations we are ready to state the main result of this paper. The key observation is that in the asymptotic regime $\mathfrak{D}_2$, the term $\mathsf{erf}\,(b/a)$ in (12) approaches a constant. This enables us to obtain the sample complexity as a function of $\mathbb{P}(\mathcal{S}_\tau)$.

Corollary 1. *Under the asymptotic regime $\mathfrak{D}_2$, the number of samples needed to satisfy the (τ, η)-close approximation (7) is given by*

$$n \overset{\mathfrak{D}_2}{\geq} \frac{1}{g}\left((\phi_{\mathrm{U}} - \phi_{\mathrm{L}})\left(\tau - a \times \mathsf{erf}^{-1}(2\eta - 1)\right) - \frac{k}{d}\sum_{i=1}^d \lambda_i + \phi_{\mathrm{L}}\right), \tag{16}$$

where a is as defined in (13). Here,

$$g = \frac{\epsilon k(d-k)}{d^2(d+2)(d-1)}\sum_{1 \leq i < j \leq d}(\lambda_i - \lambda_j)^2, \tag{17}$$

and $\mathsf{erf}^{-1}(\cdot)$ denotes the inverse Gauss error function.

Corollary 1 extends Theorem 7 in [1] from a single to an arbitrary number of principal components. The result shows the tradeoff between the sample complexity and the other parameters. The anticipated tradeoff between the privacy parameter ϵ and the approximation quality τ, η is also captured in (16) via a non-trivial relation. The derived lower bound (16) is a function of all the eigenvalue gaps $\lambda_i - \lambda_j$, $1 \leq i < j \leq d$. On the contrary, the corresponding result in [1] depends only on the largest eigenvalue gap $\lambda_1 - \lambda_2$. Intuitively, the ability of the proposed result (16) to capture all the eigenvalue gaps may explain the reason for the high approximation accuracy shown in the next section.

IV. Numerical Results and Discussions

Implementing the considered privacy-preserving PCA algorithm amounts to sampling from the matrix Bingham distribution (5) for a given privacy parameter ϵ. The sampling is typically implemented by a Gibbs sampler [15]. Gibbs sampling is an MCMC technique in which samples are generated by a Markov chain whose stationary distribution is the Bingham distribution (5). Using the MCMC procedure, implementation of the algorithm (for an arbitrary k) has been extensively discussed in Section 6 of the original work [1]. Rather than reproducing their results, in this section we focus on illustrating the accuracy of our derived results as well as studying the sample complexity for random data matrices.

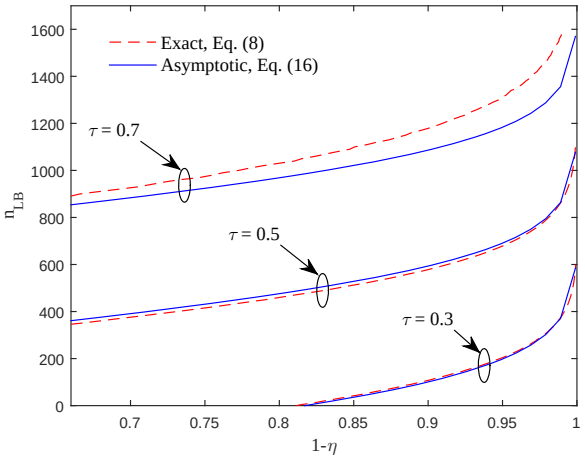

Fig. 1. Finite size accuracy of Corollary 1: data matrices with $d = 8$, $k = 3$, $\epsilon = 0.1$, and $\mathbf{\Lambda} = \text{diag}(0.32, 0.21, 0.17, 0.03, 0.02, 0.01, 0.001)$.

In Figure 1, we plot the lower bound of the number of samples n_{LB} needed to satisfy the (τ, η)-close approximate PCA as a function of the accuracy parameter $1 - \eta$. We consider different values of the utility parameter τ, where the privacy parameter is set at $\epsilon = 0.1$. We focus on the more interesting regime when $1 - \eta$ is high. The analytical curves are obtained by the asymptotic approximation (16) and the exact curves are drawn by numerical integration of (8) with the normalization constant (6) computed via the software package in [16]. As observed in Figure 1, the proposed approximation (16) already attains good accuracy for not-so-large k and d, although it is formally valid when both k and d approach infinity.

In Figure 1, the data matrix is assumed to be fixed. For utility guarantee in practice, it is often important to formulate a statistical guarantee rather than a deterministic one valid only for a specific data matrix. This requires the underlying data matrices being assumed random. As a result, the induced sample complexity becomes a random variable. The knowledge of the statistics of the sample complexity could then be used to set up a statistical guarantee. As a first step to understanding its statistical behavior, we plot in Figure 2 the average sample complexity by Monte Carlo simulations. We assume the entries

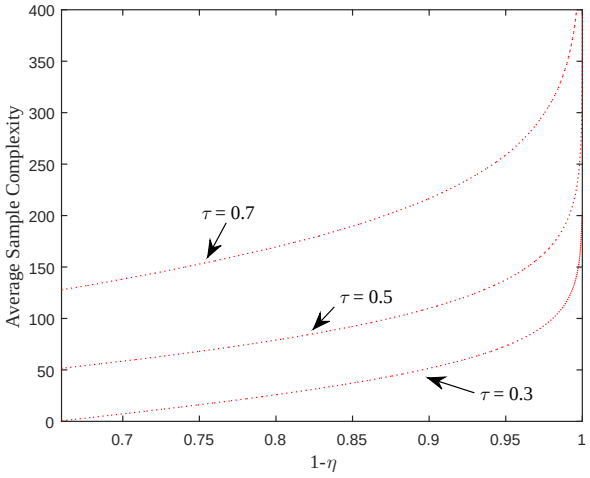

Fig. 2. Simulation of average sample complexity: Gaussian data matrices with $d = 5$, $\epsilon = 0.1$, and eigenvalues of population covariance matrix $(0.5, 0.4, 0.01, 0.001, 0.0001)$.

of the data matrix $\mathbf{X}$ follow an i.i.d. standard Gaussian distribution, which is a reasonable choice when one has no prior information of the distribution of the data matrix. The eigenvalues of the population covariance matrix $\mathbb{E}\left[\mathbf{X}\mathbf{X}'\right]/n$ are chosen to be $(0.5, 0.4, 0.01, 0.001, 0.0001)$ with $d = 5$ and the privacy parameter is set at $\epsilon = 0.2$. Each curve is drawn by averaging over 10^4 realizations of the data matrices. Each realization is normalized by $\sqrt{\rho}/\sqrt{\text{tr}\left(\mathbf{X}\mathbf{X}'\right)}$, where ρ equals the sum of eigenvalues of the population covariance matrix. The normalization ensures that the condition $||\mathbf{A}||_{\text{F}} \leq 1$ is satisfied. It is observed that the shape of the average sample complexity curves is similar to the deterministic case in Figure 1. On the other hand, an analytical approximation to these simulated curves remains an interesting open problem.

V. Conclusion

In this paper, we studied the performance of a private PCA algorithm for an arbitrary number of principal components. In particular, we derived a sample complexity bound that extends a recent result that is valid for a single principal component. The proposed result takes a simple form and is accurate in the regimes of practical interests as demonstrated by empirical simulations. The main tools leading to the analytical results are matrix-variate hypergeometric functions and asymptotic analysis via random matrix theory.

Acknowledgment

L. Wei, A. Hero, and V. Tarokh's work was partially supported by ARO grant W911NF-15-1-0479. A. Sarwate's work was partially supported by NSF grant CCF-1453432, NIH grant 1R01DA040487-01A1, and DARPA and the US Navy under contract N66001-15-C-4070.

REFERENCES

[1] K. Chaudhuri, A. D. Sarwate, and K. Sinha, "A near-optimal algorithm for differentially-private principal components," *J. Mach. Learn. Res.,* vol. 14, pp. 2905-2943, 2013.

[2] C. Dwork, , F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," In Proceedings of Theory of Cryptography Conference (TCC'06), 2006.

[3] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: the SuLQ framework," In Proceedings of ACM Symposium on Principles of Database Systems (PODS'05), 2005.

[4] C. Dwork, K. Talwar, A. Thakurta, and L. Zhang, "Analyze Gauss: Optimal bounds for privacy-preserving principal component analysis," In Proceedings of ACM Symposium on Theory of Computing (STOC'14), 2014.

[5] O. Sheffet, "Private approximations of the 2nd-moment matrix using existing techniques in linear regression," arXiv: 1507.00056, 2015.

[6] H. Imtiaz and A. Sarwate, "Symmetric matrix perturbation for differentially-private principal component analysis," In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16), 2016.

[7] M. Kapralov and K. Talwar, "On differentially private low rank approximation," In Proceedings of Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'13), 2013.

[8] M. Hardt and A. Roth, "Beyond worst-case analysis in private singular vector computation," In Proceedings of ACM Symposium on Theory of Computing (STOC'13), 2013.

[9] M. Hardt and E. Price, "The noisy power method: A meta algorithm with applications," In Proceedings of Conference on Neural Information Processing Systems (NIPS'14), 2014.

[10] F. McSherry and K. Talwar, "Mechanism design via differential privacy," In Proceedings of Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 2007.

[11] Y. Chikuse, *Statistics on Special Manifolds.* Lecture Notes in Statistics 174, New York: Springer, 2003.

[12] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Arguments.* Singapore: World Scientific, 1997.

[13] P. Koev and A. Edelman, "The efficient evaluation of the hypergeometric function of a matrix argument," *Math. Comp.,* vol. 75, no. 254, pp. 833-846, 2006.

[14] R. Butler and A. Wood, "Laplace approximations for hypergeometric functions with matrix argument," *Ann. Statist.,* vol. 30, pp. 1155-1177, 2002.

[15] P. Hoff, "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data," *J. Comp. Graph. Stat.,* vol. 18, pp. 438-456, 2009.

[16] P. Koev and A. Edelman, "Hypergeometric function of a matrix argument," http://www-math.mit.edu/∼plamen/software/mhgref.html, 2008.

[17] C. Bingham, T. Chang, and D. St. P. Richards, "Approximating the matrix Fisher and Bingham distributions: Applications to spherical regression and procrustes analysis," *J. Multivar. Anal.*, vol. 41, pp. 314-337, 1992.