Principal Component Analysis in the Stochastic Differential Privacy Model

Fanhua Shang[1,2] **Zhihui Zhang**[1] **Tao Xu**[1] **Yuanyuan Liu**[1] **Hongying Liu**[*1]

[1]Key Lab. of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, China.
[2]Peng Cheng Laboratory, Shenzhen, China.

Abstract

In this paper, we study the differentially private Principal Component Analysis (PCA) problem in stochastic optimization settings. We first propose a new stochastic gradient perturbation PCA mechanism (DP-SPCA) for the calculation of the right singular subspace to achieve (ϵ, δ)-differential privacy. For achieving a better utility guarantee and performance, we then present a new differential privacy stochastic variance reduction mechanism (DP-VRPCA) with gradient perturbation for PCA. To the best of our knowledge, this is the first work of stochastic gradient perturbation for (ϵ, δ)-differentially private PCA. We also compare the proposed algorithms with existing state-of-the-art methods, and experiments on real-world datasets and on classification tasks confirm the improved theoretical guarantees of our algorithms.

1 INTRODUCTION

Dimensionality reduction is an essential tool for understanding complex datasets that appear in modern machine learning and data mining applications. Even though a single data point can be represented by hundreds or even thousands of features, the phenomenon of interest is that real-world data are usually highly redundant with a low intrinsic dimensionality. By reducing the data in a high dimensional space to a lower dimension space, we can discover important structural relationships between features. Many efficient methods use the transformed features for learning tasks such as classification and regression, and significantly reduce the space required to store the data. Other benefits include removal of noise and extraction of correlations. Principal components analysis (PCA) is one of the most classical methods for dimensionality reduction in statistics, machine learning, and signal processing, and it can be used in social science [Costello and Osborne, 2005], financial econometrics [Ait-Sahalia and Xiu, 2017], medicine [Barber et al., 1975] and genomics [Lu and Xu, 2013].

Given a dataset represented by a matrix $X \in \mathcal{R}^{n\times d}$, PCA is used to calculate a good low-rank approximation of the second moment matrix $A = \frac{1}{n}X^\top X$ of a set of data points in $\mathcal{R}^d$. The rank k of the approximation is chosen to be the intrinsic dimension of the data. This procedure is the process of computing the principal components (i.e., a k-dimensional subspace of $\mathcal{R}^d$).

1.1 DIFFERENTIALLY PRIVATE PCA

Many modern machine learning applications are performed on large amount of personal information about individuals, and thus these datasets contain sensitive information about user behaviors. Therefore, it is important to design efficient algorithms to discover important structural relationships in the data while considering the sensitive nature of the data. Differential privacy (DP) [Dwork et al., 2006] is a commonly recognized criterion that provides provable protection of identity, and many researchers have used it to develop privacy preserving and learning algorithms such as [Chaudhuri and Monteleoni, 2008, Chaudhuri et al., 2011, Bojarski et al., 2014]. A simple idea to hide personal information is to add some special types of noise to the original model. After that, the attacker has two outputs with slightly different inputs, and can not tell whether the output changes are due to artificial noise or input differences.

In this paper, we study the problem of differentially private PCA. More specifically, for a given matrix X that contains some sensitive information, our goal is to compute a low-dimensional subspace that captures the covariance of X as much as possible without compromising the privacy of any individual. There is a large number of literature on studying PCA with differential privacy such as [Dwork et al., 2014, Hardt and Roth, 2013, Hardt and Roth, 2011, Hardt

*Corresponding author.

Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI 2021), PMLR 161:1110–1119.

Table 1: Comparison of existing (ϵ, δ)-DP and $(\epsilon, 0)$-DP algorithms, and the proposed algorithms, where each component function is G-$Lipschitz$. Note that $\mu \in (0, 1)$, $\alpha = \log(1/\delta)/((1-\mu)\epsilon) + 1$, k is the number of the principal components, and σ_1, σ_k and σ_{k+1} denote the first, k-th and $(k+1)$-th eigenvalue of the covariance matrix A, respectively.

Algorithm	**Utility bound**	**Noise magnitude**	**Privacy**
Gaussian [Dwork et al., 2014]	$O(\frac{\sqrt{nk\ln(1.25/\delta)}}{\epsilon(\sigma_k^2-\sigma_{k+1}^2)})$	$O(\frac{\ln(1.25/\delta)}{\epsilon^2})$	(ϵ, δ)
Wishart [Jiang et al., 2015]	$O(\frac{\sqrt{k}d\log d}{n\epsilon(\sigma_k^2-\sigma_{k+1}^2)})$	$O(\frac{d\log d}{n\epsilon})$	$(\epsilon, 0)$
Local Gaussian [Wang and Xu, 2020]	$O(\frac{kd}{n\epsilon^2})$	$O(\frac{\sigma_1^2\sigma_{k+1}^2}{\sigma_k^2-\sigma_{k+1}^2})$	(ϵ, δ)
DP-SPCA (ours)	$O(\frac{\sqrt{kd}}{n\sqrt{\epsilon}})$	$O(\frac{m\alpha G^2}{n^2\epsilon\mu})$	(ϵ, δ)
DP-VRPCA (ours)	$O(\frac{\sqrt{k}d\log n}{n^2\epsilon^2})$	$O(\frac{G^2Tm}{n^2\epsilon^2})$	(ϵ, δ)

and Price, 2014, Blum et al., 2005, Chaudhuri et al., 2012, Kapralov and Talwar, 2012, Balcan et al., 2016, Jiang et al., 2015, Wang and Xu, 2020], which output a noisy projection matrix for dimensionality reduction while preserving the privacy of any single data point.

There are several general ways to construct differentially private PCA algorithms for differential privacy approximations. Roughly speaking, a great number of methods need to add noise, which can be roughly classified into three categories. The first type of methods are input perturbation methods by directly adding noise to the data before performing the required computation. The second type of methods are to perturb the objective function, which are called objective perturbation methods. And the third type of methods are to perturb the output of a non-DP algorithm. Most of existing methods use the well-known power method based on [Golub and Loan, 2013] to get the variance error bound for the top private left or right singular vectors of X in two ways of differential privacy. The first way uses input perturbation by directly adding noise to the second moment matrix A and then uses the power method to compute the eigenvector decomposition of a perturbed covariance matrix, e.g., [Chaudhuri et al., 2012, Jiang et al., 2015]. An alternative way is to add noise to the calculation process of the power method to achieve objective perturbation, e.g., [Dwork et al., 2014, Hardt and Roth, 2013, Hardt and Roth, 2011, Balcan et al., 2016].

However, each iteration of the power method requires multiplying the covariance matrix A by one or multiple vectors. When the sample size n and the dimension d are large, the total running time depends on the condition number κ of the matrix A and the eigengap $\lambda = (\sigma_k - \sigma_{k+1})/\sigma_k$ of the covariance matrix between its k-th and $(k+1)$-th eigenvalues (i.e., σ_k and σ_{k+1}), which means many passes are required over the data if λ is small and/or κ is large. This paper will propose two efficient stochastic gradient descent methods to compute the private right singular subspace with greater privacy protection, which can avoid repeated manipulation of the covariance matrix and is more suitable for large-scale problems. Moreover, the PCA algorithms are very sensitive in this sense because the top eigenvector can change by 90° by changing one data point in the dataset. Thus, we apply a Gaussian perturbation to gradients in first order optimization algorithms rather than to X, and the magnitude of the perturbation in each entry can be smaller than that required under the naive input perturbation.

The notion of differential privacy mainly has two types: $(\epsilon, 0)$-DP (also called pure DP) and (ϵ, δ)-DP (also called approximation DP). (ϵ, δ)-DP is a weaker version of $(\epsilon, 0)$-DP because the former allows to break the privacy guarantee with a very small probability (more precisely, δ). Some pioneering differentially private PCA methods such as [Dwork et al., 2014, Hardt and Roth, 2013, Hardt and Roth, 2011, Hardt and Price, 2014, Blum et al., 2005, Balcan et al., 2016, Wang and Xu, 2020] used the concept of (ϵ, δ)-DP. Some methods such as [Chaudhuri et al., 2012, Kapralov and Talwar, 2012, Jiang et al., 2015] are based on $(\epsilon, 0)$-DP. This paper will propose efficient stochastic algorithms with (ϵ, δ)-differential privacy.

Many differentially privat methods focus on the k-subspace PCA problem in the low-dimensional setting (i.e., $n \gg d$), e.g., [Hardt and Roth, 2013, Hardt and Roth, 2011, Jiang et al., 2015, Dwork et al., 2014]. And some methods were proposed for another case, which is assumed to be high-dimensional ($n \ll d$), e.g., [Wang and Xu, 2020]. In the first case, it is generally assumed that the rows of the data matrix are normalized to have a norm of at most 1, as done in this paper. In high-dimensional scenarios, the natural method is to impose some additional structural constraints on the leading eigenvectors, and a commonly used constraint is to assume that the leading eigenvectors are row sparse. In this paper, we mainly consider efficient differentially private PCA algorithms in the first case.

1.2 OUR MAIN CONTRIBUTIONS

The main contributions of this paper are highlighted as follows:

• We propose two efficient stochastic algorithms for differentially private PCA (called DP-SPCA and DP-VRPCA, respectively), which use gradient perturbation at each iteration. We also provide their privacy and utility guarantees, and the theoretical results show that our differential privacy stochastic algorithms have tighter utility upper bounds with less perturbed noise, as shown in Table 1. Especially, DP-VRPCA achieves better utility and is more suitable for large-scale datasets.

• We perform many experiments on a9a, MNIST and CIFAR-10 with different principal components to verify the utility improvement of DP-SPCA and DP-VRPCA. Numerical results consistently confirm that under the same experimental performance, DP-SPCA and DP-VRPCA can achieve better utility guarantees.

• We use a linear support vector machine (SVM) as a discriminative classifier and evaluate our approaches on three standard classification datasets, a9a, MNIST and Real-sim. Our experiments indicate that compared with other algorithms, the classification accuracies of DP-SPCA and DP-VRPCA are closer to that of the algorithm without noise perturbation (i.e., their non-private counterpart). Especially, the performance of DP-VRPCA is superior and more stable.

The remainder of the paper is organized as follows. Section 2 gives some related work about differentially private PCA. Section 3 presents some preliminaries and definitions of differential privacy. Section 4 proposes two efficient stochastic differentially private PCA algorithms and provides theoretical analysis on privacy and utility guarantees of our mechanism together with the comparison to several highly-related work. Then the numerical experimental results are presented in Section 5. Finally, we conclude this work in Section 6.

2 RELATED WORK

In this section, we present some related work about differentially private PCA and differentially private empirical risk minimization (ERM) problems.

2.1 DIFFERENTIALLY PRIVATE PCA

This paper deals with the problem of differentially private PCA. Starting from the SULQ framework [Blum et al., 2005], which uses an early input perturbation framework, and the parameters of noise are refined by [Dwork et al., 2006]. Dwork et al. [2014] proved the state-of-the-art utility bounds for (ϵ, δ)-DP. Hardt and Roth [2011] provided a better bound under the coherence assumption. In [Hardt and Roth, 2013, Hardt and Price, 2014], the authors used a noisy power method to generate the principal eigenvectors iteratively with removing the previously generated ones. Hardt and Price [2014] provided a special case for $(\epsilon, 0)$-DP as well. Chaudhuri et al. [2012] proposed the first differentially private PCA algorithm for $(\epsilon, 0)$-DP based on an exponential mechanism [McSherry and Talwar, 2007]. Kapralov and Talwar [2012] argued that the algorithm in [Chaudhuri et al., 2012] lacks convergence of the chain, which may affect the privacy guarantee, and they also devised a mixed algorithm for low-rank matrix approximation. Jiang et al. [2015] proposed a new input perturbation mechanism for publishing a covariance matrix to achieve $(\epsilon, 0)$-DP, which uses a Wishart distribution to keep the positive semi-definiteness of the published covariance matrix. More recently, Wang and Xu [2020] introduced the PCA problem under the distributed non-interactive local differential privacy model in both low and high dimensional cases.

2.2 DIFFERENTIALLY PRIVATE ERM

Differentially private PCA can be viewed as a differentially private empirical risk minimization (ERM) problem. In the last decade, there are many methods such as [Wu et al., 2015, Wang et al., 2016, Abadi et al., 2016, Zhang et al., 2017, Wang et al., 2019] for differentially private ERM, which attack the problem from different perspectives. Objective perturbation, output perturbation, and gradient perturbation are the three main methods of performing ERM under DP guarantees. Among the three categories, the gradient perturbation method can obtain the optimal solution to ensure both differential privacy and utility guarantees, and is the preferred method in practice.

The output and objective perturbation for privacy protection are considered in [Chaudhuri and Monteleoni, 2008, Chaudhuri et al., 2011], and the privacy and utility of Logistic regression and SVM are theoretically guaranteed. The impacts of learning rate and batch size on DP-ERM were studied in [Song et al., 2013]. Wang et al. [2016] studied the stability, learnability and other properties of DP-ERM. An adaptive privacy budget based on concentrated DP was proposed in [Lee and Kifer, 2018].

Most of existing methods are based on Gradient Descent (GD) and Stochastic Gradient Descent (SGD) [Abadi et al., 2016, Wang et al., 2019]. Abadi et al. [2016] developed new algorithmic techniques for learning and a refined analysis of privacy costs within the framework of differential privacy. Wang et al. [2019] proposed a DP Laplacian smoothing SGD (DP-LSSGD) to train machine learning models with differential privacy guarantees. However, they are slow in general for large datasets [Liu et al., 2017, Shang et al., 2018, 2020]. Wang et al. [2017] presented algorithms with tighter utility upper bound and less running time based SVRG [Johnson and Zhang, 2013].

This paper will propose two new efficient stochastic PCA algorithms with (ϵ, δ)-DP by using gradient perturbation, which are different from existing input and object perturbation power methods, and can achieve privacy protection at a

much lower noise magnitude.

3 PRELIMINARIES AND DEFINITIONS

In this section, we provide some background information on differentially private PCA.

3.1 NOTATIONS

We first give some notations used in this paper. We generally use lower case letters to denote vectors, and upper case letters to denote matrices. The data given to our algorithms is a set of n vectors $X = [x_1, x_2, \cdots, x_n]$, where each x_i corresponds to the private value of one individual, $x_i \in \mathcal{R}^d$, and $\|x_i\| \leq 1$ for all i. For a given matrix $X \in \mathcal{R}^{n\times d}$ ($d \ll n$) whose rows are the data vectors $\{x_i\}$, let $A = \frac{1}{n}X^\top X$ denote the $d \times d$ second moment matrix of the data, which is called the covariance matrix. The matrix A is positive semidefinite, and has Frobenius norm $\|A\|_F$. For a vector $w \in \mathcal{R}^d$, $\|w\|$ denotes the ℓ_2 norm. For a matrix $X \in \mathcal{R}^{n\times d}$, the spectral norm is defined as $\|X\| = \max_{w\in\mathcal{R}^d, \|w\|_2=1} \|Xw\|_2$; the Frobenius norm is defined as $\|X\|_F = \sqrt{\sum_{ij} x_{ij}^2}$, where x_{ij} are the entries of the matrix X.

For a matrix X, the singular value decomposition of X is defined as: $X = U\Sigma V^\top$, where $U \in \mathcal{R}^{n\times n}$ and $V \in \mathcal{R}^{d\times d}$ are called the left and right singular vectors, respectively, and they are unitary matrices. The matrix $\Sigma \in \mathcal{R}^{n\times d}$ is a diagonal matrix with non-negative entries $\sigma_1, \cdots, \sigma_{\min(d,n)}$ ($\sigma_1 \geq \sigma_2 \geq \cdots$) along the diagonal, which are called the singular values. We define $V_k = [v_1, \cdots, v_k]$ and call it the principal k right singular subspace. It is well known that $\|X\|_2 = \sigma_1$, $\|X\|_F^2 = \sum_i \sigma_i^2$, and $\|XV_k\|_F^2 = \sum_{i=1}^k \sigma_i^2 = \max_{P\in\mathcal{R}^{d\times k}} \|XP\|_F^2$.

3.2 PROBLEM SETUP

Considering that the PCA problem can be viewed as a finite sum problem in the following form

$$\min_{V\in\mathcal{R}^{d\times k}: V^\top V = I} \quad f(V) = \frac{1}{n}\sum_{i=1}^{n} f_i(V), \tag{1}$$

where both $f_i(V)$ and $f(V)$ are smooth, and k denotes the number of the top eigenvalues.

We consider the problem of recovering the top k right singular subspace of a $n \times d$ matrix X, where $k \ll d$. This is equivalent to recover the top k eigenvectors of $X^\top X$, or equivalently, to solve the optimization problem

$$\min_{V\in\mathcal{R}^{d\times k}: V^\top V = I} \quad -V^\top \big(\frac{1}{n}\sum_{i=1}^{n} x_i^\top x_i\big)V. \tag{2}$$

3.3 SOME ASSUMPTIONS

Assumption 1 (L-smooth) *Each component function $f_i(\cdot)$ is L-smooth with respect to the norm $\|\cdot\|$, if for $\forall v_1, v_2 \in \mathcal{R}^d$, there is a constant L such that*

$$\|\nabla f_i(v_1) - \nabla f_i(v_2)\| \leq L\|v_1 - v_2\|,\ i \in [n]. \tag{3}$$

The above definition has another equivalent form, i.e., $\forall v_1, v_2 \in \mathcal{R}^d$,

$$-\frac{L}{2}\|v_1 - v_2\|^2 \leq f(v_1) - f(v_2) - \langle \nabla f(v_2), v_1 - v_2\rangle \leq \frac{L}{2}\|v_1 - v_2\|^2. \tag{4}$$

Assumption 2 (Lipschitz Function) *A function f is G-Lipschitz, if for $\forall v_1, v_2 \in \mathcal{R}^d$, there exists $G > 0$ such that $|f(v_1) - f(v_2)| \leq G\|v_1 - v_2\|$.*

3.4 DIFFERENTIAL PRIVACY

Given a matrix X, our goal is to output a subspace that preserves privacy and captures the variance of X as much as possible. To introduce the formal definition of differential privacy in [Dwork et al., 2006, Dwork et al., 2014, Mironov, 2017], we call two datasets $\mathcal{D}, \mathcal{D}' \in \mathcal{R}^n$ neighbors if they differ in only one row, as each row in X corresponds to an individual user. We will ensure (ϵ, δ)-differential privacy with Gaussian perturbation.

Definition 1 (Differential Privacy [Dwork et al., 2006]) *A randomized mechanism $\mathcal{A} : \mathcal{D}^n \to \mathcal{R}$ is (ϵ, δ)-differentially private ((ϵ, δ)-DP) if for every two neighboring datasets $\mathcal{D}, \mathcal{D}' \in \mathcal{D}^n$ differing in one entry and for any subset of outputs $\mathcal{O} \subseteq \mathcal{R}$, it holds that*

$$\Pr[\mathcal{A}(\mathcal{D}) \in \mathcal{O}] \leq e^\epsilon \Pr[\mathcal{A}(\mathcal{D}') \in \mathcal{O}] + \delta. \tag{5}$$

When $\delta = 0$, $\mathcal{A}$ is ϵ-differentially private (ϵ-DP).

Definition 2 (ℓ_2-sensitivity [Dwork et al., 2014]) *For two adjacent datasets $\mathcal{D}, \mathcal{D}' \in \mathcal{D}^n$ differing in one entry, the ℓ_2-sensitivity $\Delta_2(q)$ of a function $q : \mathcal{D}^n \to \mathcal{R}$ is defined as*

$$\Delta_2(q) = \sup_{\mathcal{D},\mathcal{D}'} \|q(\mathcal{D}) - q(\mathcal{D}')\|_2. \tag{6}$$

Definition 3 (Gaussian Mechanism [Dwork et al., 2014]) *Given any function $q : \mathcal{D} \to \mathcal{R}^d$, the Gaussian mechanism is defined as:*

$$\mathcal{M}(\mathcal{D}, q, \epsilon) = q(\mathcal{D}) + z, \tag{7}$$

where z is drawn from Gaussian distribution $\mathcal{N}(0, \lambda^2 I_d)$ with $\lambda \geq \frac{\sqrt{2\ln(1.25/\delta)}\Delta_2(q)}{\epsilon}$. Here $\Delta_2(q)$ is the ℓ_2-sensitivity of the function q. Gaussian mechanism preserves (ϵ, δ)-differential privacy.

Definition 4 ((α, ρ)-RDP [Mironov, 2017]) *A randomized mechanism $\mathcal{A} : \mathcal{D}^n \to \mathcal{R}$ is ρ-Rényi differentially private of order $\alpha > 1$, i.e., (α, ρ)-RDP, if for any adjacent datasets $\mathcal{D}, \mathcal{D}^{'} \in \mathcal{D}^n$ differing in one entry, it holds that*

$$D_\alpha(\mathcal{A}(\mathcal{D})\|\mathcal{A}(\mathcal{D}^{'})) \triangleq \frac{1}{\alpha-1}\log \mathbb{E}\left(\frac{\mathcal{A}(\mathcal{D})}{\mathcal{A}(\mathcal{D}^{'})}\right)^{\alpha} \leq \rho. \quad (8)$$

4 DIFFERENTIALLY PRIVATE STOCHASTIC ALGORITHMS FOR PCA

In this section, we propose two new efficient differential privacy stochastic algorithms for PCA. Then we discuss the main components of our algorithms toward differentially private PCA. The first is a differentially private stochastic gradient descent PCA (DP-SPCA) algorithm, in which Gaussian perturbation is added to the computation of gradients at each iteration step. Our next stochastic algorithm is differentially private stochastic variance reduction PCA (DP-VRPCA), which is based on a variance reduced stochastic gradient technique [Johnson and Zhang, 2013] and has a tighter utility upper bound and a faster convergence speed. We derive that both algorithms are differentially private approximation to the top-k subspace, and DP-SPCA and DP-VRPCA are guaranteed to be (ϵ, δ)-differentially private.

Before we describe our algorithms, we first present a general framework of differentially private PCA. A privacy-preserving PCA takes the row of $X \in \mathcal{R}^{n\times d}$ as input data and then calculates the sample covariance matrix $A = \frac{1}{n}X^\top X$. Finally, it computes the top-k subspace of A as the output. However, this framework needs to repeat multiplying operations on the covariance matrix A, and the processing can be prohibitive for large-scale datasets. Moreover, the frameworks of differentially private PCA usually have to assume that the eigenvalues satisfy a condition for the sake of analysis, e.g., $\sigma_k^2 - \sigma_{k+1}^2 = \omega(\sqrt{n})$ as in [Dwork et al., 2014]. In this paper, we use a stochastic framework with privacy-preserving alternatively, which only chooses one sample at each iteration. Therefore, this avoids performing multiplication operations on the matrix A and does not need to assume such a condition of the eigenvalues.

In the following formulations of the algorithms, we first describe them in the simplest form of the problem in Eq. (2). Where $k = 1$, the problem can be formulated as follows:

$$\min_{v\in\mathcal{R}^d : v^\top v=1} \quad f(v) = -v^\top(\frac{1}{n}\sum_{i=1}^{n} x_i^\top x_i)v, \quad (9)$$

and our goal is to find the top eigenvector v_k.

4.1 DIFFERENTIALLY PRIVATE STOCHASTIC SINGULAR SUBSPACE COMPUTATION VIA GAUSSIAN MECHANISM

DP-SPCA with $k=1$ is summarized in Algorithm 1, which outlines our basic update rules for training a differentially private PCA model with parameters v by minimizing the empirical loss function $f(v)$.

In our DP-SPCA algorithm, we estimate the gradient of $f(\cdot)$ by calculating the gradient of the loss function on the randomly selected examples, which provides an unbiased estimator similar to ordinary SGD. The way of our DP-SPCA algorithm to learn a DP model is injecting Gaussian noise into the stochastic gradient, and we give the following update rule,

$$v_t' = v_{t-1} - \eta_t(\nabla f_{i_t}(v_{t-1}) + z), \quad (10)$$

where η_t is the step-size, and z is the Gaussian noise injected for DP guarantees.

More specifically, a random example is selected in each iteration and we compute the gradient $\nabla_v f_{i_t}$, which is the stochastic gradient of f evaluated from the input x_{i_t} and is an unbiased estimate of the gradient $\nabla_v f$, and adds a Gaussian noise z in order to protect privacy. Then v is updated toward the local minimum along the opposite direction of this noise gradient $(-(\nabla_v f(x_{i_t}) + z))$. Finally, we perform an orthogonal transformation, which aims to ensure that v is a unit vector. We assume that each component function f_i is *G-Lipschitz*.

The block form ($k > 1$) of Algorithm 1 is similar to how to generalize power iterations to orthogonal iterations, where the d-dimensional vectors v_t are replaced by $d \times k$ matrices V_t. And orthogonalization is used instead of the normalization step, which aims to obtain an orthonormal column basis of a matrix, that is, $v_t = \frac{1}{\|v_t'\|}v_t'$ is replaced by $V_t = V_t'(V_t^{'\top}V_t')^{-1/2}$.

Next we provide the theoretical analysis of Algorithm 1 under the framework of differentially private PCA. Privacy guarantee is the basic requirement of a privacy algorithm, and the utility guarantee indicates that how effective is the algorithm against the non-private version. Note that we provide the detailed proofs of some theoretical results in the Supplementary Materials.

Lemma 1 *In DP-SPCA (i.e., Algorithm 1), for $\epsilon \leq \frac{T}{n^2}$ and $\delta > 0$, it is (ϵ, δ)-differential private if*

$$\lambda_1^2 = \frac{20T\alpha G^2}{n^2\epsilon\mu}, \quad (11)$$

where $\alpha = \log(1/\delta)/((1-\mu)\epsilon) + 1$, and $\mu \in (0, 1)$.

Theorem 1 (Privacy Guarantee for DP-SPCA) *Suppose that each component function f_i is G-Lipschitz. Given*

Algorithm 1 DP-SPCA(X, η)

Input: Data matrix $X \in \mathcal{R}^{n\times d}$, the number of iterations T.
Initialize: An initial vector v_0, and $A = \frac{1}{n}X^\top X$.
1: **for** $t = 1, 2, \cdots, T$ **do**
2: Randomly sample i_t from $\{1, 2, ..., n\}$ uniformly;
3: $v'_t = v_{t-1} - \eta_t(\nabla f_{i_t}(v_{t-1}) + z)$, where $z \sim \mathcal{N}(0, \lambda_1^2 I_d)$ and $\lambda_1^2 = \frac{20T\alpha G^2}{n^2\epsilon\mu}$;
4: $v_t = \frac{1}{\|v'_t\|}v'_t$;
5: **end for**
Output: The top eigenvector $v_k = v_T$.

$\delta > 0$ and privacy budget ϵ, with injected Gaussian noise $\mathcal{N}(0, \lambda_1^2)$ for each coordinate, DP-SPCA satisfies (ϵ, δ)-DP with $\lambda_1^2 = 20T\alpha G^2/(n^2\epsilon\mu)$, where $\alpha = \log(1/\delta)/((1-\mu)\epsilon) + 1$, if there exits $\mu \in (0, 1)$ such that $5T\alpha/(n^2\epsilon\mu) \geq 1.5$ and the sampling rate $\tau = 1/n$, we get v_t of each iteration is $(t\mu\epsilon/T + (1-\mu)\epsilon, \delta)$-DP, and finally get the output of DP-SPCA, v_k, is (ϵ, δ)-DP.

Theorem 2 (Utility Guarantee for DP-SPCA) *Suppose each component function f_i is G-Lipschitz. Given $\epsilon, \delta > 0$, with injected Gaussian noise $\mathcal{N}(0, \lambda_1^2)$ for each coordinate, DP-SPCA satisfies (ϵ, δ)-DP with $\lambda_1^2 = \frac{20T\alpha G^2}{n^2\epsilon\mu}$, where $\alpha = \log(1/\delta)/((1-\mu)\epsilon) + 1$. If we choose $\eta_t = 1/T$ and $T = C_1(\|v_0 - \hat{v}_k\|^2 + \hat{\gamma})n^2\epsilon\mu/dG^2\log(1/\delta)$, the output of DP-SPCA (i.e., v_k) satisfies the following utility*

$$\mathbb{E}\|v_k - \hat{v}_k\|^2 \leq \frac{C_2 G\sqrt{\|v_0 - \hat{v}_k\|^2 + \hat{\gamma})d\log(1/\delta)}}{n\sqrt{\epsilon\mu}}. \quad (12)$$

The above utility bound for the case of $k \geq 2$ becomes

$$\mathbb{E}\|V_k - \hat{V}_k\|_F^2 \leq \frac{C_3\sqrt{k}G\sqrt{\|V_0 - \hat{V}_k\|^2 + \hat{\gamma})d\log(1/\delta)}}{n\sqrt{\epsilon\mu}}, \quad (13)$$

where $\hat{v}_k$ and $\hat{V}_k$ are the optimal solution without perturbation, $\hat{\gamma} = \|\nabla f(\hat{V}_k)\|^2 + 1/2$, and C_1, C_2, C_3 are constants.

Remark 1 *This theorem shows that the utility bound of DP-SPCA is $O(\frac{\sqrt{kd}}{n\sqrt{\epsilon}})$, which is much better than that of Local Gaussian [Wang and Xu, 2020] (i.e., $O(\frac{kd}{n\epsilon^2})$).*

4.2 DIFFERENTIALLY PRIVATE STOCHASTIC VARIANCE REDUCTION SINGULAR SUBSPACE COMPUTATION VIA GAUSSIAN MECHANISM

The advantage of DP-SPCA is that each iteration only depends on one derivative $\nabla f_{i_t}(v)$, so the calculation cost is greatly reduced. However, the disadvantage of DP-SPCA is that randomness introduces high variance due to the fact that $\nabla f_{i_t}(v)$ is equal to the gradient $\nabla f(v)$, but each $\nabla f_{i_t}(v)$ is different. This means that SGD-style algorithms have a relatively large variance and greatly slow down convergence.

For large-scale datasets, DP-SPCA is relatively slow and has low accuracy. To improve these performances, our second algorithm, DP-VRPCA, randomly samples a k-dimensional subspace with Gaussian perturbation that ensures differential privacy and is biased towards high utility. The pseudo-code of DP-VRPCA is outlined in Algorithm 2. We call each execution of the inner loop as an iteration, and call each execution of the outer loop as an epoch. Therefore, DP-VRPCA includes multiple epochs, and each epoch contains T iterations. The update rule in Line 6 is a generalized rule of DP-SPCA, which can significantly reduce the variance and use a relatively large learning rate.

The basic idea of DP-VRPCA is to use randomly sampled row x_{i_t} of the matrix X to perform stochastic updates, but think of them as a similar form of exact power iterations, and use them to gradually reduce the variance of stochastic updates. In particular, the algorithm is divided into several epochs $s = 1, 2, \cdots, m$. In each epoch, we perform an exact power iteration for the matrix A, and then perform T stochastic updates, which are rewritten as follows:

$$v'_t = (I + \eta A)v_{t-1} + \eta(x_{i_t}x_{i_t}^\top - A)(v_{t-1} - \tilde{v}) - \eta z. \quad (14)$$

We can see that the first term can actually be regarded as a power iteration, the expectation of the second term is zero-mean, and the variance is determined by $\|v_{t-1} - \tilde{v}\|^2$. With the progressing of the algorithm, both v_{t-1} and $\tilde{v}$ converge to the same optimal point, so $\|v_{t-1} - \tilde{v}\|^2$ shrinks, leading to convergence. And η is a constant step-size and z is the injected Gaussian noise for DP guarantees.

The block version ($k > 1$) of Algorithm 2 is similar to the operations of Algorithm 1. Next we give the theoretical analysis of Algorithm 2.

Theorem 3 (Privacy Guarantee for DP-VRPCA) *Suppose that each component function f_i is G-Lipschitz. For $\epsilon \leq \frac{Tm}{n^2}$ and $\delta > 0$, DP-VRPCA is (ϵ, δ)-differential private with $\lambda_2^2 = G^2Tm\log(1/\delta)/(n^2\epsilon^2)$.*

Theorem 4 (Utility Guarantee for DP-VRPCA) *Suppose each component function f_i is G-Lipschitz. Given $\epsilon, \delta > 0$, with injected Gaussian noise $\mathcal{N}(0, \lambda_2^2)$ for each coordinate, DP-VRPCA satisfies (ϵ, δ)-DP with $\lambda_2^2 = \frac{G^2Tm\log(1/\delta)}{n^2\epsilon^2}$. If we choose $\eta \leq 1/12L$ and $m = C_4\log(\frac{n^2\epsilon^2}{dG^2\log(1/\delta)-\gamma})$, the output of DP-VRPCA (i.e., v^k) satisfies the following utility*

$$\mathbb{E}\|v_k - \hat{v}_k\|^2 \leq \frac{C_5G^2d\log(n\epsilon/dG)\log(1/\delta)}{n^2\epsilon^2}. \quad (15)$$

The upper bound for the case of $k \geq 2$ becomes

$$\mathbb{E}\|V_k - \hat{V}_k\|_F^2 \leq \frac{C_6\sqrt{k}G^2d\log(n\epsilon/dG)\log(1/\delta)}{n^2\epsilon^2}, \quad (16)$$

Algorithm 2 DP-VRPCA(X, η)

Input: Data matrix $X \in \mathcal{R}^{n\times d}$, the step-size η, and the iteration numbers T, m.

Initialize: An initial vector $\tilde{v}^0$, and $A = \frac{1}{n}X^\top X$.

for $s = 1, 2, \cdots, m$ **do**
  $\tilde{\mu}^{s-1} = \nabla f(\tilde{v}^{s-1}) = A\tilde{v}^{s-1}$;
  $v_0^s = \tilde{v}^{s-1}$;
  for $t = 1, 2, \cdots, T$ **do**
    Randomly sample i_t from $\{1, 2, ..., n\}$ uniformly;
    $g_t^s = \nabla f_{i_t}(v_{t-1}^s) - \nabla f_{i_t}(\tilde{v}^{s-1}) + \tilde{\mu}^{s-1} + z$, where $z \sim \mathcal{N}(0, \lambda_2^2 I_d)$ and $\lambda_2^2 = \frac{G^2 Tm\log(1/\delta)}{n^2\epsilon^2}$;
    $v_t^{s'} = v_{t-1}^s - \eta g_t^s$;
    $v_t^s = \frac{1}{\|v_t^{s'}\|} v_t^{s'}$;
  end for
  $\tilde{v}^s = v_T^s$;
end for

Output: The top eigenvector $v_k = v_T^m$.

where C_4, C_5, C_6 are three constants.

Remark 2 *Theorem 4 shows that the utility bound of DP-VRPCA is $O(\frac{\sqrt{k}d\log n}{n^2\epsilon^2})$, which is much better than that of Local Gaussian [Wang and Xu, 2020], i.e., $O(\frac{kd}{n\epsilon^2})$.*

4.3 COMPARISON WITH RELATED WORK

In this subsection, we compare our algorithms with related methods in terms of utility upper bound and the level of noise magnitude required. There are some major measurements to analyze the results of different algorithms. For instance, some existing algorithms such as [Dwork et al., 2014, Jiang et al., 2015, Wang and Xu, 2020] use the subspace distance (i.e., $\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F$) as a quality measurement, and some algorithms such as [Hardt and Roth, 2013, Chaudhuri et al., 2012] apply the variance (i.e., $\|X\hat{V}_k\|_F^2 - \|XV_k\|_F^2$) as a quality measurement. Then we show that our measurement is of the same order of the subspace distance, so that we can compare our utility bound with those of other methods.

Lemma 2 *The criterion (i.e., $\|V_k - \hat{V}_k\|_F^2$) used in this paper is of the same order as the general subspace criterion $\|V_kV_k^\top - \hat{V}_k\hat{V}_k^\top\|_F$, i.e., when $k = 1$,*

$$\|v_1 - \hat{v}_1\|^2 = O(\|v_1v_1^\top - \hat{v}_1\hat{v}_1^\top\|_F), \tag{17}$$

and when $k > 1$

$$\|V_k - \hat{V}_k\|_F^2 = O(\|V_kV_k^\top - \hat{V}_k\hat{V}_k^\top\|_F). \tag{18}$$

We show the compared results of different algorithms in Table 1, which are for the case that the measurement is $\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F$. Note that the compared algorithms such as [Dwork et al., 2014] need to assume that the eigenvalues of A satisfy such condition, e.g., $\sigma_k^2 - \sigma_{k+1}^2 = \omega(d)$, which is not required in Theorems 2 and 4. We can see that our DP-SPCA and DP-VRPCA achieve tighter utility bounds with less noise required compared with existing methods. To be more precise, the noise of DP-VRPCA is of the same order as DP-SPCA, which is inversely proportional to n^2 and is of less magnitude than other methods. And the utility bounds of DP-SPCA and DP-VRPCA are both inversely proportional to n, which is obviously better than the upper bounds of other algorithms. In particular, the utility bound of DP-VRPCA is better than that of DP-SPCA by a factor of $O(\sqrt{d}\log n/n)$ in the low-dimensional case.

Table 2: Summary of the datasets used in our experiments.

Dataset	♯ **Samples,** n	♯ **Features,** d	**Sparsity**
a9a	32,561	123	11.2757%
CIFAR-10	50,000	3,072	99.7617%
MNIST	60,000	784	80.8798%
Real-sim	72,309	20,958	0.2448%

5 NUMERICAL EXPERIMENTS

In this section, we turn to validate our theoretical results on real-world datasets. We compare the proposed algorithms (i.e., DP-SPCA and DP-VRPCA) with other state-of-the-art methods in order to verify our theoretical results.

5.1 DATA AND PREPROCESSING

We report the performance of our algorithms on the widely used real-world datasets (i.e., a9a, CIFAR-10, MNIST and Real-sim)[1]. The detailed information of the four datasets is shown in Table 2. We preprocessed each dataset by normalizing each row so that each entry has maximum value 1, and normalized each column such that the maximum column Euclidean norm is 1.

Settings. In the experiments, the parameters are set as follows: the number of iterations is set as $T = n$ in DP-VRPCA, where n is the number of samples. We pick a fixed step-size, which is set to $\eta_t = \frac{1}{2n}$ in DP-SPCA and $\eta = \frac{1}{n}$ in DP-VRPCA. And the parameters δ in σ_1^2 and σ_2^2 are set to 0.001 in DP-SPCA and DP-VRPCA. Moreover, the initial vectors v_0 and $\tilde{v}^0$ are randomly generated unit vectors.

Performances Metrics. The performance indicator is measured by the subspace distance $\|V_k - \hat{V}_k\|_F^2$, where $\hat{V}_k$ is the

[1]All the datasets can be downloaded from the website at `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`.

Table 3: Classification accuracies of all the differentially private algorithms on a9a ($k = 10$), MNIST ($k = 10$) and Real-sim ($k = 300$) in the k-dimensional subspaces under the privacy budget $\epsilon = 0.1$.

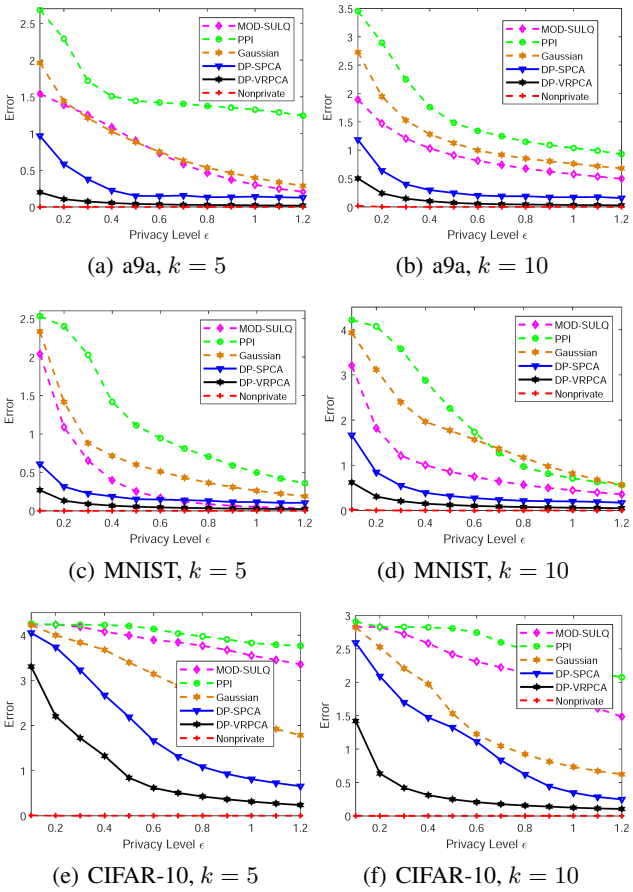
Method	a9a	MNIST	Real-sim
Non-private counterpart	84.9483±0.00	99.2883±0.00	99.4275±0.00
PPI [Hardt and Roth, 2013]	79.8034±0.14	96.1450±0.20	71.7836±0.11
MOD-SULQ [Chaudhuri et al., 2012]	80.1136±0.10	96.5200±0.09	72.7932±0.17
Gaussian [Dwork et al., 2014]	80.3199±0.23	97.0667±0.34	85.5685±0.34
DP-SPCA (ours)	82.5539±0.09	98.3500±0.18	92.1290±0.10
DP-VRPCA (ours)	**82.5539±0.07**	**98.4750±0.14**	**94.1852±0.13**

optimal solution without perturbation and V_k is the result obtained by different algorithms.

5.2 EFFECT OF PRIVACY ON DIFFERENT ALGORITHMS

In this subsection, we first confirm our theoretical results of DP-SPCA and DP-VRPCA on real-world datasets, and we compare our algorithms with the state-the-of-art algorithms including: Nonprivate PCA (without perturbation), MOD-SULQ (Chaudhuri et al. [2012]), Gaussian Mechanism (Dwork et al. [2014]), PPI (Hardt and Roth [2013]), DP-SPCA (i.e., Algorithm 1) and DP-VRPCA (i.e., Algorithm 2) on a9a, MNIST and CIFAR-10 with normalized rows for each dataset.

In the initialization of v_0 and $\tilde{v}^0$, we tried a uniformly generated random projection. We measured the utility by subspace distance, where V_k is the k-dimensional subspace output by the algorithm, thus this reflects how close the output subspace is to the true PCA subspace (i.e., $\hat{V}_k$) in terms of representing the data. The experimental results are shown in Figure 1, which indicates that the error decreases as the privacy budget ϵ increases (i.e., less private), and our DP-SPCA and DP-VRPCA algorithms are significantly better than other compared algorithms. Especially, DP-VRPCA has better performance due to the decreasing variance through iterations. These experiments support the claim that not just the theoretical analysis is superior, but also the performance is affected in a positive way.



(a) a9a, $k = 5$ (b) a9a, $k = 10$

(c) MNIST, $k = 5$ (d) MNIST, $k = 10$

(e) CIFAR-10, $k = 5$ (f) CIFAR-10, $k = 10$

Figure 1: Comparison of all the differentially private PCA algorithms with different privacy levels, where k denotes the number of eigenvalues.

5.3 EFFECT OF PRIVACY ON CLASSIFICATION

A common use of dimensionality reduction algorithms is a preparation for a classification or clustering task. To evaluate the effectiveness of the proposed algorithms, we projected the data onto the subspace output by the algorithms and used the projected data to measure the classification accuracy. We used the linear SVM as a discriminative classifier and evaluate our algorithms and other compared methods on a9a, MNIST and Real-sim. We chose these three datasets because they are publicly available and have long served as benchmarks for machine learning applications.

In the classification experiment, we used half of each dataset as the reserved set for calculating a projection subspace. We projected the classification data onto the subspace computed based on the holdout set. Here, 10% of each dataset was used for training and parameter tuning, and the rest was used for testing. We repeated the classification process 10 times for each algorithm, and then ran the whole procedure over 10 random arrangements of each dataset. The average classification results are reported in Table 3.

The classification results show that the classification performance of our algorithms in top-k PCA subspace is closer to that of non-private PCA, while the performance of the compared algorithms (e.g., MOD-SULQ) is a little worse. The classification accuracies of MOD-SULQ and PPI also appear to have higher variance compared to our algorithms and non-private PCA. The reason is that the projections tend to be far away from the top-k PCA subspace, making the classification error with a larger variance.

6 CONCLUSIONS

In this paper, we studied the problem of differentially private PCA and proposed two new stochastic PCA algorithms with better utility bounds. When gradient perturbation is combined into differentially private PCA methods, a better bound with less noise magnitude can be achieved. The proposed algorithms are simple to implement, and we also showed that they can improve the utility of the private PCA models in both theory and practice.

Several recently proposed accelerated stochastic variance reduction algorithms (e.g., Katyusha [Allen-Zhu, 2018] and MiG [Zhou et al., 2018]) and their asynchronous parallel variants such as [Zhou et al., 2018, Shang et al., 2021] can be used to further accelerate the proposed differentially private PCA algorithms for solving large-scale problems. In the further, we will use the Laplacian smoothing technique in [Wang et al., 2019] to further improve the practice performance and utility bounds of our differentially private stochastic PCA algorithms.

Acknowledgements

We thank all the reviewers for their valuable comments. This work was supported by the National Natural Science Foundation of China (Nos. 61876221, 61876220 and 61976164), the Project supported the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61621005), the Major Research Plan of the National Natural Science Foundation of China (Nos. 91438201 and 91438103), the Program for Cheung Kong Scholars and Innovative Research Team in University (No. IRT_15R53), the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) (No. B07048), and the National Science Basic Research Plan in Shaanxi Province of China (No. 2020JM-194).

References

Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. *arXiv preprint arXiv:1607.00133*, 2016.

Yacine Ait-Sahalia and Dacheng Xiu. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201(2):384–399, 2017.

Zeyuan Allen-Zhu. Katyusha: The first direct acceleration of stochastic gradient methods. *Journal of Machine Learning Research*, 18:1–51, 2018.

Maria-Florina Balcan, Simon Shaolei Du, Yining Wang, and Adams Wei Yu. An improved gap-dependency analysis of the noisy power method. In *29th Annual Conference on Learning Theory*, pages 284–309, 2016.

D.C. Barber, P.J. Howlett, and R.C Smart. Principal component analysis in medical research. *Journal of Applied Statistics*, 2(1):39–43, 1975.

Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the SuLQ framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 128–138, 2005.

Mariusz Bojarski, Anna Choromanska, Krzysztof Choromanski, and Yann LeCun. Differentially- and non-differentially-private random decision trees. *arXiv preprint arXiv:1410.6973*, 2014.

Kamalika Chaudhuri and Claire Monteleoni. Privacy-preserving logistic regression. In *Advances in Neural Information Processing Systems 21*, pages 289–296, 2008.

Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. Differentially private empirical risk minimization. *Journal of Machine Learning Research*, 12(29):1069–1109, 2011.

Kamalika Chaudhuri, Anand D. Sarwate, and Kaushik Sinha. Near-optimal differentially private principal components. In *Advances in Neural Information Processing Systems*, 2012.

Anna B. Costello and Jason Osborne. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(1):1–9, 2005.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third conference on Theory of Cryptography*, pages 265–284, 2006.

Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. Analyze gauss: Optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, pages 11–20, 2014.

Gene H. Golub and Charles F. Van Loan. *Matrix computations*, volume The Fourth Edition. Johns Hopkins University Press, 2013.

Moritz Hardt and Eric Price. The noisy power method: A meta algorithm with applications. In *Advances in Neural Information Processing Systems 27*, pages 2861–2869, 2014.

Moritz Hardt and Aaron Roth. Beating randomized response on incoherent matrices. *arXiv preprint arXiv:1111.0623*, 2011.

Moritz Hardt and Aaron Roth. Beyond worst-case analysis in private singular vector computation. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 331–340, 2013.

Wuxuan Jiang, Cong Xie, and Zhihua Zhang. Wishart mechanism for differentially private principal components analysis. *arXiv preprint arXiv:1511.05680*, 2015.

Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems 26*, volume 26, pages 315–323, 2013.

Michael Kapralov and Kunal Talwar. On differentially private low rank approximation, 2012.

Jaewoo Lee and Daniel Kifer. Concentrated differentially private gradient descent with adaptive per-iteration privacy budget. *arXiv preprint arXiv:1808.09501*, 2018.

Yuanyuan Liu, Fanhua Shang, and James Cheng. Accelerated variance reduced stochastic ADMM. In *Proceedings of AAAI Conference Artificial Intelligence*, pages 2287–2293, 2017.

Dongsheng Lu and Shuhua Xu. Principal component analysis reveals the 1000 genomes project does not sufficiently cover the human genetic diversity in asia. *Frontiers in Genetics*, 4:127–127, 2013.

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

Ilya Mironov. Renyi differential privacy. In *Proceedings of 30th IEEE Computer Security Foundations Symposium*, pages 263–275, 2017.

Fanhua Shang, Yuanyuan Liu, Licheng Jiao, Kaiwen Zhou, James Cheng, Yan Ren, and Yufei Jin. ASVRG: Accelerated proximal SVRG. In *Proceedings of Machine Learning Research*, pages 815–830, 2018.

Fanhua Shang, Kaiwen Zhou, Hongying Liu, James Cheng, Ivor W. Tsang, Lijun Zhang, Dacheng Tao, and Licheng Jiao. VR-SGD: A simple stochastic variance reduction method for machine learning. *IEEE Transactions on Knowledge and Data Engineering*, 32(1):188–202, 2020.

Fanhua Shang, Hua Huang, Jun Fan, Yuanyuan Liu, Hongying Liu, and Jianhui Liu. Asynchronous parallel, sparse approximated svrg for high-dimensional machine learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

Shuang Song, Kamalika Chaudhuri, and Anand D. Sarwate. Stochastic gradient descent with differentially private updates. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 245–248, 2013.

Bao Wang, Quanquan Gu, March Boedihardjo, Farzin Barekat, and Stanley J. Osher. Dp-lssgd: A stochastic optimization method to lift the utility in privacy-preserving erm. *arXiv preprint arXiv:1906.12056*, 2019.

Di Wang and Jinhui Xu. Principal component analysis in the local differential privacy model. *Theoretical Computer Science*, 809:296–312, 2020.

Di Wang, Minwei Ye, and Jinhui Xu. Differentially private empirical risk minimization revisited: Faster and more general. In *Advances in Neural Information Processing Systems 30*, pages 2722–2731, 2017.

Yu-Xiang Wang, Jing Lei, and Stephen Fienberg. Learning with differential privacy: stability, learnability and the sufficiency and necessity of ERM principle. *Journal of Machine Learning Research*, 17(1):6353–6392, 2016.

Xi Wu, Matthew Fredrikson, Wentao Wu, Somesh Jha, and Jeffrey F. Naughton. Revisiting differentially private regression: Lessons from learning theory and their consequences. *arXiv preprint arXiv:1512.06388*, 2015.

Jiaqi Zhang, Kai Zheng, Wenlong Mou, and Liwei Wang. Efficient private ERM for smooth objectives. *arXiv preprint arXiv:1703.09947*, 2017.

Kaiwen Zhou, Fanhua Shang, and James Cheng. A simple stochastic variance reduced algorithm with fast convergence rates. In *Proceedings of International Conference on Machine Learning*, pages 5975–5984, 2018.