Laplacian Smoothing Stochastic ADMMs With Differential Privacy Guarantees

Yuanyuan Liu, *Member, IEEE*, Jiacheng Geng, Fanhua Shang, *Senior Member, IEEE*, Weixin An, Hongying Liu, *Senior Member, IEEE*, Qi Zhu, and Wei Feng, *Member, IEEE*

***Abstract*— Many machine learning tasks such as structured sparse coding and multi-task learning can be converted into an equality constrained optimization problem. The stochastic alternating direction method of multipliers (SADMM) is a popular algorithm to solve such large-scale problems, and has been successfully used in many real-world applications. However, existing SADMMs fail to take into consideration an important issue in their designs, i.e., protecting sensitive information. To address this challenging issue, this paper proposes a novel differential privacy stochastic ADMM framework for solving equality constrained machine learning problems. In particular, to further lift the utility in privacy-preserving equality constrained optimization, a Laplacian smoothing operation is also introduced into our differential privacy ADMM framework, and it can smooth out the Gaussian noise used in the Gaussian mechanism. Then we propose an efficient differentially private variance reduced stochastic ADMM (DP-VRADMM) algorithm with Laplacian smoothing for both strongly convex and general convex objectives. As a by-product, we also present a new differentially private stochastic ADMM algorithm with DP guarantees. In theory, we provide both private guarantees and utility guarantees for the proposed algorithms, which show that Laplacian smoothing can improve the utility bounds of our algorithms. Experimental results on real-world datasets verify our theoretical results and the effectiveness of our algorithms.**

***Index Terms*— Alternating direction method of multipliers (ADMM), differential privacy, variance reduction, Laplacian smoothing, utility guarantees.**

Manuscript received June 30, 2021; revised October 30, 2021 and February 21, 2022; accepted March 31, 2022. Date of publication April 25, 2022; date of current version May 17, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61876221, Grant 61876220, Grant 61976164, Grant 62072334, and Grant 61836009; and in part by the Natural Science Basic Research Program of Shaanxi under Grant 2022GY-061. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dali Kaafar. *(Corresponding authors: Fanhua Shang; Hongying Liu.)*

Yuanyuan Liu, Jiacheng Geng, Weixin An, and Qi Zhu are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, Xi'an 710071, China (e-mail: yyliu@xidian.edu.cn).

Fanhua Shang is with the School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: fhshang@formail.com).

Hongying Liu is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, School of Artificial Intelligence, Xidian University, and with also the Peng Cheng Laboratory, Shenzhen 518055, China (e-mail: hyliu@xidian.edu.cn).

Wei Feng is with the School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin 300350, China, and also with the Key Research Center for Surface Monitoring and Analysis of Cultural Relics, State Administration of Cultural Heritage (SACH), Tianjin 300350, China (e-mail: wfeng@tju.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIFS.2022.3170271, provided by the authors.

Digital Object Identifier 10.1109/TIFS.2022.3170271

I. Introduction

MACHINE learning models are widely applied in many real-world problems. However, the privacy of individuals whose information is included in datasets should be protected when the models are actually applied, especially when utilizing sensitive data such as financial accounts and health-care data. Therefore, it is important to design machine learning algorithms that protect users' privacy. As a rigorous and standard concept of privacy, differential privacy (DP) [1] can guarantee that the algorithm learns statistical information of the population, but nothing about individual users. Empirical risk minimization (ERM) has been widely studied for achieving simultaneously privacy preserving and learning in deterministic and stochastic settings. The DP methods can be roughly classified into three categories. The first type of approaches such as [2], [3] are to perturb the output of non-private algorithms. The second type of methods such as [4]–[6] are to perturb the objective function. The third type of approaches such as [7] are to perturb gradients in first order optimization algorithms. Most DP-ERM algorithms are based on gradient perturbation, e.g., [8]–[12]. The theoretical and numerical results in [11] and [12] have verified that Laplacian smoothing can improve the utility of the DP methods for ERM problems both numerically and theoretically.

Most differentially private methods mentioned above focus on the ERM problem: $F(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x) + r(x)$, where x is the model parameter, n is the number of training samples, $f_i(\cdot)$ is the loss on the i-th sample, and $r(\cdot)$ is a regularizer, e.g., the ℓ_1-norm regularizer $\lambda\|x\|_1$ with a regularization parameter λ. However, many machine learning problems such as graph-guided fused Lasso [13] and generalized Lasso [14] are formulated as the following more complex optimization problem with an equality constraint,

$$\min_{x\in\mathbb{R}^{d_1}, y\in\mathbb{R}^{d_2}} \Big\{ f(x) + r(y),\ \text{s.t.},\ Ax + By = c \Big\} \tag{1}$$

where $A \in \mathbb{R}^{d_3\times d_1}$, $B \in \mathbb{R}^{d_3\times d_2}$ are two given matrices, $c \in \mathbb{R}^{d_3}$ is a constant vector, $f(x) := \frac{1}{n}\sum_{i=1}^{n} f_i(x)$, each $f_i(x)$ is a convex function, and $r(\cdot)$ is convex but possibly non-smooth. The problem (1) covers a variety of machine learning tasks such as structured sparsity problems (e.g., $\|Ax\|_1$ for graph-guided fused Lasso).

The alternating direction method of multipliers (ADMM) is an efficient optimization method for solving Problem (1), and many non-private ADMM algorithms including deterministic ADMMs such as [16]–[18] and stochastic ADMMs

TABLE I

COMPARISON OF THE STOCHASTIC ADMM ALGORITHMS WITH (ϵ, δ)-DP. NOTE THAT DPADMM [15] WAS PROPOSED TO SOLVE GENERAL CONVEX (GC OR NON-STRONGLY CONVEX) NON-SMOOTH ERM PROBLEMS, WHILE OUR DP-SADMM AND DP-VRADMM ARE FOR BOTH THE μ-STRONGLY CONVEX (SC) AND GC PROBLEM (1). WE SUPPOSE THAT EACH COMPONENT FUNCTION f_i IS l-LIPSCHITZ AND g-SMOOTH. NOTE THAT $\tau < 1$ IS A CONSTANT DEFINED IN THEOREM 3 BELOW. THE BOUNDS IGNORE MULTIPLICATIVE DEPENDENCE ON $\log(1/\delta)$

Algorithm	Constraint	Assumption	Gradient complexity	Utility bound
DPADMM [15]	$x = y$	Non-strongly convex	$\mathcal{O}(n^2)$	$\mathcal{O}\left(\frac{l\sqrt{d_1}}{\epsilon n}\right)$
DP-SADMM (ours)	$Ax + By = c$	Non-strongly convex	$\mathcal{O}\left(\frac{\epsilon^2 n^2}{\tau d_1 l^2 \log(\frac{1}{\delta})}\right)$	$\mathcal{O}\left(\frac{l\sqrt{\tau d_1}}{\epsilon n}\right)$
DP-SADMM (ours)	$Ax + By = c$	μ-strongly convex	$\mathcal{O}\left(\frac{\mu\epsilon^2 n^2}{\gamma\|Q_\nu\|_2 \tau d_1 l^2 \log(\frac{1}{\delta})}\right)$	$\mathcal{O}\left(\frac{\tau d_1 l^2 \log(n)}{\mu\epsilon^2 n^2}\right)$
DP-VRADMM (ours)	$Ax + By = c$	Non-strongly convex	$\mathcal{O}\left(\frac{gn\epsilon}{l\sqrt{\tau d_1 \log(1/\delta)}}\right)$	$\mathcal{O}\left(\frac{l\sqrt{\tau d_1}}{\epsilon n}\right)$
DP-VRADMM (ours)	$Ax + By = c$	μ-strongly convex	$\mathcal{O}\left((n + \frac{g}{\mu}) \log \frac{n\epsilon\mu}{d_1}\right)$	$\mathcal{O}\left(\frac{\tau d_1 l^2 \log(n)}{\mu\epsilon^2 n^2}\right)$

such as [19]–[22] have been widely studied in both deterministic and stochastic settings. There are many recently proposed stochastic ADMMs including variance reduction methods such as SAG-ADMM [19], SDCA-ADMM [20] and SVRG-ADMM [21] and momentum accelerated methods such as ASVRG-ADMM [22], [23], which have much faster convergence speed than deterministic ADMMs, especially for large-scale optimization problems. Chen and Lee [15], Wang and Zhang [24] proposed privacy preserving stochastic ADMM algorithms with gradient and objective perturbations, respectively. However, the differentially private ADMM algorithms [15], [24] can be only used to solve non-smooth ERM problems with a simple constraint $x = y$, which can be viewed as a special case of Problem (1). In addition, many differentially private distributed ADMM algorithms such as [25], [26] have been proposed for distributed machine learning problems. To the best of our knowledge, there exists no differential privacy stochastic ADMM for the more complex problem (1).

A. Motivations and Our Contributions

Many non-private stochastic ADMMs mentioned above have been proposed to solve the equality constrained minimization problem (1). However, they do not take into consideration an important issue (i.e., the protection of sensitive information in data) in their designs. Although several DP-ERM methods such as [11], [12] leverage the Laplacian smoothing (LS) [27] as post-processing to smooth the injected Gaussian noise and improve the utility in privacy-preserving ERM, there is still a lack of research in studying how to introduce the LS operator into DP-ADMM for solving Problem (1). To address these issues, in this paper, we focus on how to apply the privacy protection mechanism with a Laplacian smoothing operator to solve the more general equality constrained optimization problem (1). We first propose an efficient differentially private ADMM framework for solving Problem (1). Compared with DP-ERM methods such as [11], the proposed DP-ADMM algorithms are non-trivial in the extensions of both algorithm design and proving convergence because of their essential differences of optimal conditions and convergence criterion. In particular, different from the convergence analysis of the non-DP stochastic ADMMs in [28] and [21], the existence of the LS operator Q_ν^{-1} in the proposed algorithms brings a challenge for our theoretical analysis.

We summarize our main contributions below.

- We propose a general differentially private stochastic ADMM framework with gradient perturbation for solving Problem (1). To further enhance the utility, we also introduce Laplacian smoothing into the proposed framework. To the best of our knowledge, this is the first work to design privacy-preserving stochastic ADMMs with Laplacian smoothing for the equality constrained problem (1).
- By using the proposed framework, we design a novel efficient differentially private stochastic ADMM algorithm (called DP-VRADMM) with variance reduction for both strongly convex (SC) and general convex (GC) objectives. As a by-product, we also present a new differentially private stochastic ADMM algorithm (called DP-SADMM) for both SC and GC problems. Moreover, we prove that the proposed algorithms including DP-VRADMM and DP-SADMM satisfy (ϵ,δ)-DP.
- Moreover, we also provide the theoretical guarantees for the proposed algorithms, which are non-trivial extensions to those of non-private stochastic ADMMs due to the introduction of Laplacian smoothing. Our theoretical results are listed in Table I, which match the near-optimal utility bounds of the DP-ERM algorithm [1] for both SC and GC cases. In particular, the theoretical results show that under the same privacy budget, Laplacian smoothing can further improve the utility bounds of our algorithms for both SC and GC cases.
- Finally, various experimental results on many real-world datasets further verify our theoretical results, and show that the proposed algorithms are efficient for solving Problem (1) in the stochastic setting. Moreover, Laplacian smoothing can improve the performance of many real-world applications.

II. PRELIMINARIES AND RELATED WORK

In this section, we first introduce some definitions of differential privacy, and present some related work about ERM algorithms with Laplacian smoothing and stochastic ADMMs.

A. Notation

Given a vector $x \in \mathbb{R}^{d_1}$, $\|x\|$ is the Euclidean norm, and $\|x\|_1 = \sum_i |x_i|$ is the ℓ_1-norm. For a matrix A, $\|A\|_2$ is its

spectral norm (i.e., the largest singular value of A), $A^\dagger$ denotes its pseudoinverse, and $\|x\|_G^2 = x^T G x$, where $G \in \mathbb{R}^{d_1 \times d_1}$ is a semi-positive definite matrix. Let (x^*, y^*) be the optimal solution of Problem (1), and σ^2 denote the variance of injected Gaussian noise.

Definition 1: A function f is l-Lipschitz if for $\forall x_1, x_2$, it satisfies $|f(x_1) - f(x_2)| \leq l\|x_1 - x_2\|$.

Definition 2: f is g-smooth if for $\forall x_1, x_2$, there exists $g \geq 0$ such that $f(x_2) \leq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{g}{2}\|x_2 - x_1\|^2$.

Definition 3: A function f is μ-strongly convex if for $\forall x_1, x_2$ and for any subgradient $\partial f(x_1)$ at x_1, it satisfies $f(x_2) \geq f(x_1) + \langle \partial f(x_1), x_2 - x_1 \rangle + \frac{\mu}{2}\|x_2 - x_1\|^2$.

B. Differential Privacy

Definition 4 ((ϵ, δ)-DP [29]): A randomized mechanism $\mathcal{M} : \mathcal{D}^n \to \mathcal{R}$ with input domain $\mathcal{D}^n$ and output range $\mathcal{R}$ is (ϵ, δ)-differentially private ((ϵ, δ)-DP) if for any two adjacent data sets $\mathcal{D}, \mathcal{D}' \in \mathcal{D}^n$ differing in one entry and for any subset of outputs $O \subseteq \mathcal{R}$, it holds that

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

Definition 5 (ℓ_2-sensitivity [30]): For two adjacent data sets $\mathcal{D}, \mathcal{D}' \in \mathcal{D}^n$ differing in one entry, the ℓ_2-sensitivity $\triangle_2(q)$ of a function $q : \mathcal{D}^n \to \mathcal{R}$ is defined as:

$$\triangle_2(q) = \sup_{\mathcal{D}, \mathcal{D}'} \|q(\mathcal{D}) - q(\mathcal{D}')\|.$$

Definition 6 ((α, ϵ)-RDP [31]): A randomized mechanism $\mathcal{M} : \mathcal{D}^n \to \mathcal{R}$ is ϵ-Rényi differentially private of order $\alpha > 1$, i.e., (α, ϵ)-RDP, if for any two adjacent data sets $\mathcal{D}, \mathcal{D}' \in \mathcal{D}^n$ differing in one entry, it holds that

$$D_\alpha(\mathcal{M}(\mathcal{D}) \| \mathcal{M}(\mathcal{D}')) \triangleq \frac{1}{\alpha - 1} \log \mathbb{E}\left(\frac{\mathcal{M}(\mathcal{D})}{\mathcal{M}(\mathcal{D}')}\right)^\alpha \leq \epsilon$$

where the expectation is taken over the randomness of $\mathcal{M}(\mathcal{D}')$.

Lemma 1 (From RDP to (ϵ, δ)-DP [31]): If a randomized mechanism $\mathcal{M} : \mathcal{D}^n \to \mathcal{R}$ is (α, ϵ)-RDP, it also satisfies $(\epsilon + \log(1/\delta)/(\alpha - 1), \delta)$-DP for $\forall \delta \in (0, 1)$.

Lemma 2 ([32]): Given a function $q : \mathcal{D}^n \to \mathcal{R}$, the Gaussian mechanism $\mathcal{M} = q(\mathcal{D}) + u$ satisfies $(\alpha, \alpha \triangle_2^2(q)/(2\sigma^2))$-RDP, where $u \sim \mathcal{N}(0, \sigma^2 I)$. And the Gaussian mechanism $\hat{\mathcal{M}} = q(\hat{\mathcal{D}}) + u$, where $u \sim \mathcal{N}(0, \sigma^2 I)$ and $\hat{\mathcal{D}}$ is a subset of $\mathcal{D}$ using uniform sampling without replacement, satisfies $(\alpha, 5\varsigma^2 \alpha \triangle_2^2(q)/\sigma^2)$-RDP given $\sigma^2/\triangle_2^2(q) \geq 1.5$ and $\alpha \leq \log\left(1/(\varsigma(1 + \sigma^2/\triangle_2^2(q)))\right)$, where ς is the subsampling rate.

Lemma 3 ([31]): For t randomized mechanisms $\mathcal{M}_1, \ldots, \mathcal{M}_t$, if $\mathcal{M}_1 : \mathcal{D} \to \mathcal{R}_1$ is (α, ϵ_1)-RDP, $\mathcal{M}_2 : \mathcal{R}_1 \times \mathcal{D} \to \mathcal{R}_2$ is (α, ϵ_2)-RDP, ..., and $\mathcal{M}_t : \mathcal{R}_{t-1} \times \cdots \times \mathcal{R}_1 \times \mathcal{D} \to \mathcal{R}_t$ is (α, ϵ_t)-RDP, then the composite mechanism $(\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_t)$ satisfies $(\alpha, \epsilon_1 + \cdots + \epsilon_t)$-RDP.

C. Stochastic ADMM Algorithms

The ADMM is an effective optimization method [33] and has shown attractive performance in a wide range of real-world problems, such as big data classification [34]. Especially, some machine learning problems such as graph-guided fused Lasso [13] and generalized Lasso [14] can be formulated as the problem (1) with an equality constraint $Ax = y$. The augmented Lagrangian function of Problem (1) is:

$$L_\beta(x, y, \lambda) = f(x) + r(y) + \langle \lambda,\ Ax + By - c \rangle + \frac{\beta}{2}\|Ax + By - c\|^2.$$

Here $\beta > 0$ denotes a penalty parameter, and λ is a scaled dual variable. Deterministic ADMM performs the following update rules in an alternating fashion:

$$\begin{aligned} y_{t+1} &= \arg\min_y \left\{ r(y) + \frac{\beta}{2}\|Ax_t + By - c + \lambda_t\|^2 \right\}, \\ x_{t+1} &= \arg\min_x \left\{ f(x) + \frac{\beta}{2}\|Ax + By_{t+1} - c + \lambda_t\|^2 \right\}, \\ \lambda_{t+1} &= \lambda_t + Ax_{t+1} + By_{t+1} - c. \end{aligned}$$

When updating the variable x, its update step usually has a high computational cost, especially when the number of the samples (i.e., n) is very large. To tackle the issue of high per-iteration complexity of deterministic ADMMs, Wang and Banerjee [35], Suzuki [36] and Ouyang *et al.* [28] proposed some online or stochastic ADMM algorithms. While their update rules for y_{t+1} and λ_{t+1} remain unchanged, their update rule for x_{t+1} becomes as follows:

$$x_{t+1} = \arg\min_x \left\{ \langle x, \nabla f_{i_t}(x_t) \rangle + \frac{1}{2\eta_t}\|x - x_t\|_G^2 + \frac{\beta}{2}\|Ax + By_{t+1} - c + \lambda_t\|^2 \right\}$$

where we pick i_t uniformly at random from $\{1, \ldots, n\}$, $\eta_t \propto 1/\sqrt{t}$ is a step-size, and $\|x\|_G^2 = x^T G x$ with a given positive semi-definite matrix G, e.g., $G \succeq I_{d_1}$ as in [21].

Analogous to stochastic gradient descent (SGD), stochastic ADMMs use an unbiased stochastic gradient at each iteration. However, all the algorithms have much slower convergence rates than their deterministic counterparts [37]. This barrier is mainly due to the variance introduced by the stochasticity of the gradients. Besides, to guarantee convergence, they employ a decaying step-size, which also impacts the convergence rates. More recently, a number of variance reduced stochastic ADMM algorithms such as SVRG-ADMM [21] have been proposed and made exciting progress such as linear convergence rates for SC problems. Especially, SVRG-ADMM is attractive due to its low storage requirement compared with [19], [20], and the variance of its gradients can be gradually reduced and is much smaller than that of SGD and its ADMM variants. Moreover, it uses a constant step-size and consequently has faster convergence than stochastic ADMMs such as [28].

D. Laplacian Smoothing for ERM Problems

In recent years, many effective differential privacy methods such as [4], [9], [11] have been proposed for various ERM problems. Moreover, Osher *et al.* [27] proposed a class of smoothing variants of gradient descent (GD) and SGD, and the proposed surrogates can dramatically reduce the variance, have a larger step-size, and improve generalization accuracy.

More recently, Wang *et al.* [11] proposed a Laplacian smoothing (LS) variant to DP-SGD, which uses the LS operator to effectively reduce the variance of DP-SGD. All the studies show that Laplacian smoothing can smooth out the Gaussian noise used in Gaussian mechanism, and can make the training of the machine learning models more stable and enables the trained models to generalize better. Inspired by the success of all the methods, we propose the first DP stochastic ADMM algorithm with Laplacian smoothing for solving the equality constrained minimization problem (1).

III. Differentially Private Stochastic ADMMs With Laplacian Smoothing

In this section, we propose efficient differentially private stochastic ADMM algorithms with Laplacian smoothing for solving Problem (1). We first introduce the Laplacian smoothing operator and present a new Laplacian smoothing differentially private framework for ADMMs. Then we propose a novel Laplacian smoothing differentially private variance reduction stochastic ADMM (DP-VRADMM) algorithm for both SC and GC problems. Finally, we present a new Laplacian smoothing differentially private stochastic ADMM (DP-SADMM) algorithm as a by-product.

A. Laplacian Smoothing

The Laplacian smoothing operation was recently shown to be a good choice to reduce the variance of the injected Gaussian noise used in the Gaussian mechanism [11], [12]. Inspired by the successful applications of Laplacian smoothing for differentially private SGD [11] and federated learning [12], we propose efficient Laplacian smoothing ADMM algorithms for various equality constrained machine learning problems such as graph-guided fused Lasso [13], generalized Lasso [14] and graph-guided support vector machine (SVM) [28]. Below we first introduce the Laplacian smoothing operator Q_ν^{-1}. Let $Q_\nu = I_{d_1} - \nu L$, where $\nu \geq 0$ is a constant, $I_{d_1} \in \mathbb{R}^{d_1 \times d_1}$ is an identity matrix, and $L \in \mathbb{R}^{d_1 \times d_1}$ is a discrete one-dimensional Laplacian matrix with periodic boundary condition.

$$Q_\nu = \begin{bmatrix} 1+2\nu & -\nu & 0 & \cdots & 0 & -\nu \\ -\nu & 1+2\nu & -\nu & \cdots & 0 & 0 \\ 0 & -\nu & 1+2\nu & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\nu & 0 & 0 & \cdots & -\nu & 1+2\nu \end{bmatrix}.$$

It is clear that when $\nu = 0$, Q_ν becomes an identity matrix. That is, the proposed algorithms degenerate to the algorithms without Laplacian smoothing when $\nu = 0$. Given a vector $a \in \mathbb{R}^{d_1}$, the smoothed vector q is obtained by computing $q = Q_\nu^{-1} a$, which is equivalent to $a = Q_\nu q = q - \nu\phi * q$, where $\phi = [-2, 1, 0, \ldots, 0, 1]^T$, Q_ν can be viewed as a convolution matrix, and $*$ denotes the convolution operator. Note that we use the fast Fourier transform (FFT) to efficiently compute $Q_\nu^{-1} a$. That is,

$$q = Q_\nu^{-1} a = \text{ifft}\Big(\frac{\text{fft}(a)}{1 - \nu \cdot \text{fft}(\phi)}\Big) \tag{2}$$

where fft and ifft denote the FFT and inverse FFT, respectively, and we use the component-wise division.

B. Key Properties

In this subsection, we introduce some key properties of the Laplacian smoothing operator Q_ν^{-1}, which are very useful for our convergence analysis. Below we give the properties.

Property 1: $Q_\nu^{-1} \preceq I_{d_1}$.

Proof: Let the eigenvalue decomposition of Q_ν^{-1} be $Q_\nu^{-1} = U\Lambda U^T$, where Λ is a diagonal matrix with $\Lambda_{ii} = 1/[1 + 2\nu - 2\nu\cos(2\pi i/d_1)]$ (see Lemma 4 in [11]). We have

$$\begin{aligned} Q_\nu^{-1} - I_{d_1} &= U\Lambda U^T - UU^T \\ &= U(\Lambda - I_{d_1})U^T. \end{aligned}$$

Apparently, $\Lambda_{ii} \leq 1$, and thus we obtain $Q_\nu^{-1} - I_{d_1} \preceq 0$. □

Property 2: $Q_\nu^{-1} \preceq Q_\nu^{-1}\widehat{G}Q_\nu^{-1}$, *where* $\widehat{G} = Q_\nu\widetilde{G}$, $\widetilde{G} = \gamma I_{d_1} - \eta\beta Q_\nu^{-1}A^TA$ *and* $\gamma = 1 + \eta\beta\|A^TA\|_2$.

Proof:

$$\begin{aligned} Q_\nu^{-1} - Q_\nu^{-1}\widehat{G}Q_\nu^{-1} &= Q_\nu^{-1} - \widetilde{G}Q_\nu^{-1} \\ &= (1-\gamma)Q_\nu^{-1} + \eta\beta Q_\nu^{-1}A^TAQ_\nu^{-1}. \end{aligned}$$

Since $\gamma = 1 + \eta\beta\|A^TA\|_2$, we have

$$\begin{aligned} &Q_\nu^{-1} - Q_\nu^{-1}\widehat{G}Q_\nu^{-1} \\ &\quad= \eta\beta Q_\nu^{-1}A^TAQ_\nu^{-1} - \eta\beta\|A^TA\|_2 Q_\nu^{-1} \\ &\quad= \eta\beta Q_\nu^{-1}(A^TA - \|A^TA\|_2 I_{d_1})Q_\nu^{-1} \\ &\qquad+ \eta\beta(Q_\nu^{-1} - I_{d_1})\|A^TA\|_2 Q_\nu^{-1}. \end{aligned}$$

According to the definition of the spectral norm $\|\cdot\|_2$, we have that $Q_\nu^{-1}(A^TA - \|A^TA\|_2 I_{d_1})Q_\nu^{-1} \preceq 0$. Let the eigenvalue decomposition of Q_ν^{-1} be $Q_\nu^{-1} = U\Lambda U^T$, where Λ is a diagonal matrix with $\Lambda_{ii} = \frac{1}{1+2\nu-2\nu\cos(2\pi i/d_1)}$, then we have

$$\begin{aligned} (Q_\nu^{-1} - I_{d_1})\|A^TA\|_2 Q_\nu^{-1} &= \|A^TA\|_2(U\Lambda U^T - UU^T)Q_\nu^{-1} \\ &= \|A^TA\|_2 U(\Lambda - I_{d_1})U^TU\Lambda U^T \\ &= \|A^TA\|_2 U(\Lambda - I_{d_1})\Lambda U^T. \end{aligned}$$

It is clear that $\Lambda_{ii} \leq 1$, and then we obtain $(Q_\nu^{-1} - I_{d_1})\|A^TA\|_2 Q_\nu^{-1} \preceq 0$. Therefore, we can conclude that $Q_\nu^{-1} - Q_\nu^{-1}\widehat{G}Q_\nu^{-1} \preceq 0$. □

Property 3: $(Q_\nu^{-1})^2 \preceq Q_\nu^{-1}$.

Proof: Since $(Q_\nu^{-1})^2 = U\Lambda U^TU\Lambda U^T = U\Lambda^2 U^T$, we have

$$(Q_\nu^{-1})^2 - Q_\nu^{-1} = U(\Lambda^2 - \Lambda)U^T.$$

It is clear that $\Lambda_{ii} \leq 1$, and then we have $\Lambda_{ii}^2 - \Lambda_{ii} \leq 0$, thus we obtain that $(Q_\nu^{-1})^2 - Q_\nu^{-1} \preceq 0$. □

C. Differentially Private ADMMs With Laplacian Smoothing

Inspired by the stochastic ADMMs [28], [35], we design the following update rules for our differentially private stochastic ADMMs to solve Problem (1),

$$\begin{aligned} y_{t+1} &= \arg\min_y \Big\{ r(y) + \frac{\beta}{2}\|Ax_t + By - c + \lambda_t\|^2 \Big\}, \\ x_{t+1} &= \arg\min_x \Big\{ \langle x,\, \mathcal{G}_t\rangle + \frac{1}{2\eta}\|x - x_t\|_G^2 \\ &\qquad + \frac{\beta}{2}\|Ax + By_{t+1} - c + \lambda_t\|^2 \Big\}, \\ \lambda_{t+1} &= \lambda_t + Ax_{t+1} + By_{t+1} - c \end{aligned} \tag{3}$$

Algorithm 1 DP-VRADMM for SC and GC Objectives

Input: Privacy parameters (ϵ, δ), the penalty parameter β, the step-size η, and the number of outer-iterations S.

Initialize: m, γ, $\widetilde{x}^0 = \hat{x}^0$, $\widetilde{y}^0$, $\lambda_0^1 = -\frac{1}{\beta}(A^T)^\dagger \nabla f(\widetilde{x}^0)$ for the SC case or λ_0^1 for the GC case, $\sigma^2 = c_1 l^2 S m \ln(1/\delta)/(n^2\epsilon^2)$.

1: **for** $s = 1, 2, \ldots, S$ **do**
2: $\widetilde{p} = \nabla f(\widetilde{x}^{s-1})$;
3: $x_0^s = \widetilde{x}^{s-1}$ for the SC case or $x_0^s = \hat{x}^{s-1}$ for the GC case;
4: **for** $t = 0, 1, \ldots, m-1$ **do**
5: Choose $\mathcal{B}_t \subseteq [n]$ of size b, uniformly at random;
6: $\widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) = \frac{1}{|\mathcal{B}_t|}\sum_{i_t \in \mathcal{B}_t}(\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1})) + \widetilde{p}$;
7: $y_{t+1}^s = \arg\min_y \left\{ r(y) + \frac{\beta}{2}\|Ax_t^s + By - c + \lambda_t^s\|^2 \right\}$;
8: $x_{t+1}^s = x_t - \frac{\eta}{\gamma} Q_\nu^{-1}(\widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) + u_t^s + \beta A^T(Ax_t^s + By_{t+1}^s - c + \lambda_t^s))$, where $u_t^s \sim \mathcal{N}(0, \sigma^2 I_{d_1})$;
9: $\lambda_{t+1}^s = \lambda_t^s + Ax_{t+1}^s + By_{t+1}^s - c$;
10: **end for**
11: $\widetilde{x}^s = \frac{1}{m}\sum_{t=1}^m x_t^s$, $\widetilde{y}^s = \frac{1}{m}\sum_{t=1}^m y_t^s$;
12: $\lambda_0^{s+1} = -\frac{1}{\beta}(A^T)^\dagger \nabla f(\widetilde{x}^s)$ (SC case)
13: or $\hat{x}^s = x_m^s$ and $\lambda_0^{s+1} = \lambda_m^s$ (GC case);
14: **end for**

output $\overline{x} = \widetilde{x}^S$, $\overline{y} = \widetilde{y}^S$ (SC case) or $\overline{x} = \frac{1}{S}\sum_{s=1}^S \widetilde{x}^s$, $\overline{y} = \frac{1}{S}\sum_{s=1}^S \widetilde{y}^s$ (GC case).

where λ is a Lagrangian multiplier (also called dual variable), η is a step-size or learning rate, $\beta > 0$ is a penalty parameter, $\|x\|_G^2 = x^T G x$ with a given positive semi-definite matrix $G = \gamma I_{d_1} - \eta\beta A^T A$ and $\gamma = 1 + \eta\beta\|A^T A\|_2$ as in [28], [38], and $\mathcal{G}_t$ is a stochastic gradient operator by injecting Gaussian noise. For instance, $\mathcal{G}_t = \frac{1}{n}\sum_{i=1}^n \nabla f_i(x_t) + u_t$ is used for deterministic DP-ADMMs, while stochastic gradients below are used for stochastic DP-ADMMs, where u_t is the Gaussian (or Laplace) noise added for privacy. In this paper, we mainly focus on the Gaussian mechanism for differentially private ADMMs.

In Problem (1), $r(y)$ is a specific regularizer (e.g., the graph-guided ℓ_1-norm regularizer) and is generally data independence, and thus involves no privacy. Compared with non-private ADMM algorithms including deterministic ADMMs [33] and stochastic ADMMs [21], [28], [39], [40], the update rule of y_t, as well as that of λ_t, remains unchanged.

Since Problem (1) indicates that $f(x)$ is data-dependent, we only perturb the gradient operator with respect to (w.r.t.) x by injecting Gaussian noise, as shown in the update rule of x_t in (3). However, differentially private ADMMs have a much lower utility than their non-private counterparts, as pointed out by [27]. To mitigate this degradation, we propose two efficient Laplacian smoothing differentially private stochastic ADMM algorithms. More specifically, we design the following update rule for differentially private ADMMs with Laplacian smoothing to replace with that of x_t in (3):

$$x_{t+1} = \arg\min_x \Big\{ \mathcal{G}_t^T x + \frac{1}{2\eta}\|x - x_t\|_{Q_\nu \widetilde{G}}^2 + \frac{\beta}{2}\|Ax + By_{t+1} - c + \lambda_t\|^2 \Big\} \quad (4)$$

where $\mathcal{G}_t$ is a differentially private gradient estimator (e.g., $\mathcal{G}_t = \frac{1}{b}\sum_{i_t \in \mathcal{B}_t} \nabla f_{i_t}(x_t) + u_t$ for stochastic ADMMs, and $\mathcal{B}_t$ is a mini-batch of size b) with Gaussian noise, and $\widetilde{G} = \gamma I_{d_1} - \eta\beta Q_\nu^{-1} A^T A$. It is notable that there is the main difference between the proposed update rules of x_t for differentially private stochastic ADMMs with and without Laplacian smoothing. We introduce the Laplacian matrix Q_ν into the proposed update rule in (4). That is, $\|x - x_t\|_{Q_\nu \widetilde{G}}^2$ in (4) is used to replace $\|x - x_t\|_{\widetilde{G}}^2$ in (3). Moreover, we design the semi-definite matrix $\widetilde{G} = \gamma I_{d_1} - \eta\beta Q_\nu^{-1} A^T A$ in (4) instead of G in (3). Therefore, the update rule of x_t in (4) can be reformulated as follows:

$$x_{t+1} = x_t - \frac{\eta}{\gamma} Q_\nu^{-1}\big[\mathcal{G}_t + \beta A^T(Ax_t + By_{t+1} - c + \lambda_t)\big]. \quad (5)$$

Here, we introduce the Laplacian smoothing operator Q_ν^{-1} in (5) as post-processing to smooth the injected Gaussian noise to improve the utility of differentially private stochastic ADMM algorithms. Laplacian smoothing (LS) can be viewed as a denoising technique that performs post-processing on the Gaussian noise injected stochastic gradients. And the update rule (5) can be calculated efficiently using FFT. Unlike DP-LSSGD [11] for ERM problems, the proposed algorithms also apply the Laplacian smoothing operator for the term $\beta A^T(Ax_t + By_{t+1} - c + \lambda_t)$ in (5) for equality constrained minimization problems. Therefore, the theoretical analysis of privacy and utility guarantees for our algorithms is very challenging. Note that when $\nu = 0$ (i.e., $Q_\nu^{-1} = I$), the proposed algorithms degenerate the differentially private ADMM algorithms without Laplacian smoothing. When the variance $\sigma = 0$ and $Q_\nu^{-1} = I$, the proposed ADMM algorithms become their non-private ADMM counterparts. In this sense, the non-private stochastic ADMMs can be viewed as special cases of our algorithms. As discussed above, the proposed differentially private ADMM algorithms with Laplacian smoothing are simple to implement, and only involve multiplying the gradient by the inverse of a positive definite matrix as in (5), which can be computed efficiently by FFT as in [27]. Below, we propose two efficient differentially private stochastic ADMM algorithms with Laplacian smoothing.

D. Laplacian Smoothing Differentially Private Stochastic ADMM Algorithm With Variance Reduction

In this subsection, we propose an efficient differentially private stochastic ADMM (DP-VRADMM) algorithm with variance reduction and Laplacian smoothing for solving both strongly convex and general convex problems (1). Like SVRG [41] and DP-SVRG [9], which are much faster than SGD algorithms and their DP variants, our DP-VRADMM also has an outer-inner loop structure. That is, it is divided into S outer-iterations, each consisting of m inner-iterations.

Similar to SVRG-ADMM [21], the differentially private stochastic variance reduced gradient is defined as follows:

$$\mathcal{G}_t = \frac{1}{|\mathcal{B}_t|}\sum_{i_t \in \mathcal{B}_t}(\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1})) + \widetilde{p} + u_t^s \quad (6)$$

where $u_t^s \sim \mathcal{N}(0, \sigma^2 I_{d_1})$ is the injected Gaussian noise, σ is defined in Theorem 1 below, $\widetilde{p} = \nabla f(\widetilde{x}^{s-1})$ is the full gradient

of $f(\cdot)$ at the snapshot point $\widetilde{x}^{s-1}$. In fact, the Gaussian noise u_t^s is from the first three terms in (6), as pointed out by [42]. Note that $\widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) = \frac{1}{|\mathcal{B}_t|}\sum_{i_t\in\mathcal{B}_t}(\nabla f_{i_t}(x_t^s) - \nabla f_{i_t}(\widetilde{x}^{s-1})) + \widetilde{p}$ is the mini-batch version of the stochastic variance reduced gradient in [41], and is also an unbiased approximation to the true gradient $\nabla f(x_t^s)$. As analyzed in our previous work [22], [23], the variance reduction gradients can explicitly reduce the variance of SGD and its variants including our differentially private stochastic ADMM proposed below, and their variance approaches zero as the number of iterations increases.

The detailed update rules of our DP-VRADMM algorithm for solving SC and GC objectives are outlined in Algorithm 1. The main differences of our DP-VRADMM for SC and GC problems are listed as follows:

- For the SC problem, the dual variable λ_0^s for each epoch of DP-VRADMM (i.e., Algorithm 1) is initialized as: $\lambda_0^s = -\frac{1}{\beta}(A^T)^\dagger \nabla f(\widetilde{x}^{s-1})$, while the variable is initialized as: $\lambda_0^s = \lambda_m^{s-1}$ for the GC problem as in [21] and [23]. Moreover, the starting point of each epoch is set to the snapshot point (i.e., $x_0^s = \widetilde{x}^{s-1}$) for SC problems, while $x_0^s = x_m^{s-1}$ for GC problems.
- The outputs of DP-VRADMM (i.e., Algorithm 1) for SC problems are $\widetilde{x}^S$ and $\widetilde{y}^S$, while $\overline{x} = \frac{1}{S}\sum_{s=1}^S \widetilde{x}^s$ and $\overline{y} = \frac{1}{S}\sum_{s=1}^S \widetilde{y}^s$ are for the GC case.

E. Laplacian Smoothing Differentially Private Stochastic ADMM Algorithm

In this subsection, we also present a new differentially private stochastic ADMM (DP-SADMM) algorithm with Laplacian smoothing as a by-product, as shown in Algorithm 2. Note that the update rules for y_t and λ_t are identical to those of our DP-VRADMM. In contrast, the update rule of x_t becomes

$$x_{t+1} = x_t - \frac{\eta_{t+1}}{\gamma} Q_v^{-1}\big[\mathcal{G}_t + \beta A^T(Ax_t + By_{t+1} - c + \lambda_t)\big] \tag{7}$$

where η_{t+1} is a decaying step-size, and the differentially private stochastic gradient $\mathcal{G}_t$ is defined as: $\mathcal{G}_t = \frac{1}{b}\sum_{i_t\in\mathcal{B}_t}\nabla f_{i_t}(x_t) + u_t$, where $u_t \sim \mathcal{N}(0, \sigma^2 I_{d_1})$ is the injected Gaussian noise as in Theorem 2 below. It is clear that the gradient used in DP-SADMM is the differentially private stochastic gradient, while that of DP-VRADMM is the differentially private stochastic variance reduced gradient in (6). Thus, DP-VRADMM has better performance than DP-SADMM.

Different from DP-VRADMM (i.e., Algorithm 1), which has a constant step-size, the step-size of DP-SADMM should be decaying as the number of iterations increases. More specifically, for the GC case, the step-size of DP-SADMM is set to $\eta_{t+1} = 1/\sqrt{t+1}$, and its outputs are $\overline{x} = (1/T)\sum_{t=1}^T x_t$ and $\overline{y} = (1/T)\sum_{t=1}^T y_t$. For the SC case, its step-size is set as follows: $\eta_{t+1} = 2\gamma\,\|Q_v\|_2/(\mu(t+1))$, and the outputs are set to the non-uniform averaging to improve convergence, i.e., $\overline{x} = \sum_{t=1}^T t x_t/(\sum_{t=1}^T t)$ and $\overline{y} = \sum_{t=1}^T t y_t/(\sum_{t=1}^T t)$.

F. Privacy Analysis

We first prove that the proposed DP-VRADMM algorithm with Laplacian smoothing satisfies (ϵ, δ)-DP.

Algorithm 2 DP-SADMM for SC and GC Objectives

Input: Privacy parameters (ϵ, δ), the penalty parameter β, and the number of iterations T.

Initialize: x_0, λ_0, and an initial step-size η_1.

$\sigma^2 = 20\alpha T l^2/(\theta\epsilon n^2)$;
for $t = 0, 1, \ldots, T-1$ **do**
  Choose $\mathcal{B}_t \subseteq [n]$ of size b, uniformly at random;
  $\nabla f_{\mathcal{B}_t}(x_t) = \frac{1}{b}\sum_{i_t\in\mathcal{B}_t}\nabla f_{i_t}(x_t)$;
  $y_{t+1} = \arg\min_y \big\{r(y) + \frac{\beta}{2}\|Ax_t + By - c + \lambda_t\|^2\big\}$;
  $x_{t+1} = x_t - \frac{\eta_{t+1}}{\gamma} Q_v^{-1}(\nabla f_{\mathcal{B}_t}(x_t) + u_t + \beta A^T(Ax_t + By_{t+1} - c + \lambda_t))$, where $u_t \sim \mathcal{N}(0, \sigma^2 I_{d_1})$;
  $\lambda_{t+1} = \lambda_t + Ax_{t+1} + By_{t+1} - c$;
  $\eta_{t+1} = \frac{2\gamma\,\|Q_v\|_2}{\mu(t+1)}$ (SC case) or $\eta_{t+1} = \frac{1}{\sqrt{t+1}}$ (GC case);
end for

output $\overline{x} = \frac{1}{\sum_{t=1}^T t}\sum_{t=1}^T t x_t$, $\overline{y} = \frac{1}{\sum_{t=1}^T t}\sum_{t=1}^T t y_t$ (SC case) or $\overline{x} = \frac{1}{T}\sum_{t=1}^T x_t$, $\overline{y} = \frac{1}{T}\sum_{t=1}^T y_t$ (GC case).

Theorem 1 (Privacy guarantees for DP-VRADMM): Suppose that each component function f_i is l-Lipschitz. For the privacy budget $\epsilon \le c_1 Sm/n^2$ with some constant c_1 and $\delta > 0$, DP-VRADMM satisfies (ϵ, δ)-DP with $\sigma^2 = c_2 l^2 S m \ln(1/\delta)/(n^2\epsilon^2)$, where c_2 is a constant.

The detailed proofs of Theorem 1, other theorems and lemmas (except Lemma 4) below are provided in the Supplementary Material. Moreover, we also present the privacy guarantee for DP-SADMM.

Theorem 2 (Privacy Guarantees for DP-SADMM): Suppose that each component function f_i is l-Lipschitz. For any $\delta > 0$ and privacy budget ϵ, DP-SADMM is (ϵ, δ)-DP with $\sigma^2 = 20\alpha T l^2/(\theta\epsilon n^2)$, where $\alpha = \log(1/\delta)/((1-\theta)\epsilon) + 1$, if there exists $\theta \in (0, 1)$ such that $\alpha \le \log(\theta\epsilon n^3/(5\alpha b^3 T + \theta\epsilon b n^2))$ and $5\alpha b^2 T/(\theta\epsilon n^2) \ge 1.5$.

IV. Theoretical Guarantees

In this section, we theoretically analyze the utility guarantees and oracle gradient complexities for the proposed DP-VRADMM and DP-SADMM algorithms, respectively. Different from general ERM problems, the solution of Problem (1) needs to satisfy the constraint $Ax + By = c$. Following [21], we introduce the following function $R(x, y)$ as a convergence criterion for the constrained optimization problem (1), which is the same as the variational inequality used in [43].

$$R(x, y) := f(x) - f(x^*) - \nabla f(x^*)^T(x - x^*) + r(y) - r(y^*) - r'(y^*)^T(y - y^*)$$

where $r'(y)$ denotes the (sub)gradient of $r(\cdot)$ at y. Moreover, $R(x, y) \ge 0$ for all $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$. We first give the utility guarantees for our DP-VRADMM.

Proof Sketch: Due to the introduction of the LS operator Q_v^{-1}, we cannot follow the theoretical framework of existing stochastic ADMMs such as STOC-ADMM [28] and SVRG-ADMM [21]. Thus, we introduce the idea of variable substitution, and combine it with some key properties in Section 3.2 to resolve the issue in Lemma 4 below.

• To estimate the one-iteration upper bound of the variable x in Lemma 4, we split the proof into three steps. Firstly, we construct a variable substitution $z_t^s = Q_\nu x_t^s$ and combine with the update rules of Algorithm 1. Secondly, by using both the variance upper bound and noise upper bound, we estimate the perturbation upper bound in the perturbation step. Finally, we combine the analysis with the key properties of the LS operator in Section 3.2 to obtain the one-iteration upper bound of x.

• To estimate the one-iteration upper bound of the variable y, the upper bound and the detailed derivation are given in Lemma 12 in the Supplementary Material.

• We combine the one-epoch upper bounds of the variables x and y with the privacy bound from Theorem 1 and derive our main result in Theorem 3. The detailed proofs can be found in the Supplementary Material.

A. Utility Bounds of DP-VRADMM

In this subsection, we analyze the utility bounds and gradient complexities of our DP-VRADMM for both strongly convex and general convex problems. Before presenting our main results, we give the following core lemma, which is applicable in both the SC and GC cases.

Lemma 4 (One-Epoch Analysis for x): Suppose the stepsize $\eta \le \frac{1}{2g}$ in Algorithm 1, we have the following result

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{m}\sum_{t=0}^{m-1}((1-4g\eta\alpha(b))(f(x_{t+1}^s)-f(x^*)\\
&\quad-\langle\nabla f(x^*),x_{t+1}^s-x^*\rangle)-\langle A^T\varphi_t^s,x^*-x_{t+1}^s\rangle)\Big]\\
&\le 4g\eta\alpha(b)(f(\widetilde{x}^{s-1})-f(x^*)-\langle\nabla f(x^*),\widetilde{x}^{s-1}-x^*\rangle)\\
&\quad+\frac{4g\eta\alpha(b)}{m}(f(x_0^s)-f(x^*)-\langle\nabla f(x^*),x_0^s-x^*\rangle)\\
&\quad-\frac{4g\eta\alpha(b)}{m}\mathbb{E}[f(x_m^s)-f(x^*)-\langle\nabla f(x^*),x_m^s-x^*\rangle]\\
&\quad+\frac{1}{2m\eta}\mathbb{E}[\|x^*-x_0^s\|_{\widehat{G}}^2-\|x^*-x_m^s\|_{\widehat{G}}^2]+\tau\eta d_1\sigma^2
\end{aligned}$$

where $\widehat{G}=Q_\nu\widetilde{G}=\gamma\,Q_\nu-\eta\beta A^TA$ and $\varphi_t^s=\beta(\lambda_{t+1}^s-\lambda^*)$.

Proof: • **Update steps**: The optimal condition of the x-subproblem (4) is

$$\begin{aligned}
&\mathcal{G}_t+\frac{1}{\eta}Q_\nu\widetilde{G}(x_{t+1}^s-x_t^s)\\
&\qquad+\beta A^T(Ax_{t+1}^s+By_{t+1}^s-c+\lambda_t^s)=0 \quad (8)
\end{aligned}$$

where $\mathcal{G}_t = \widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)+u_t^s = \frac{1}{|\mathcal{B}_t|}\sum_{i_t\in\mathcal{B}_t}(\nabla f_{i_t}(x_t^s)-\nabla f_{i_t}(\widetilde{x}^{s-1}))+\widetilde{p}+u_t^s$, η is a constant for SVRG-ADMM, and $\widetilde{G}=\gamma\,I_{d_1}-\eta\beta Q_\nu^{-1}A^TA$. Let $\widehat{G}=Q_\nu\widetilde{G}$, then we have

$$\begin{aligned}
\widehat{G}&=Q_\nu\widetilde{G}=\gamma\,Q_\nu-\eta\beta A^TA\\
&=\gamma\,Q_\nu-\eta\beta\|A^TA\|_2I_{d_1}+\eta\beta\|A^TA\|_2I_{d_1}-\eta\beta A^TA\\
&\succeq I_{d_1} \quad (9)
\end{aligned}$$

where $\widehat{G}\succeq I_{d_1}$ holds due to the fact that $\gamma\,Q_\nu-\eta\beta\|A^TA\|_2I_{d_1}\succeq I_{d_1}$ (because of $\gamma=1+\eta\beta\|A^TA\|_2$ and the definition of Q_ν) and $\eta\beta\|A^TA\|_2I_{d_1}-\eta\beta A^TA\succeq 0$ (because of the definition of the spectral norm $\|\cdot\|_2$). Since $\lambda_{t+1}^s=\lambda_t^s+Ax_{t+1}^s+By_{t+1}^s-c$ as in Step 9 in Algorithm 1, the optimal condition (8) is rewritten as follows:

$$\mathcal{G}_t+\frac{1}{\eta}\widehat{G}(x_{t+1}^s-x_t^s)+\beta A^T\lambda_{t+1}^s=0. \quad (10)$$

Let $z_t^s=Q_\nu x_t^s$, then $x_t^s=Q_\nu^{-1}z_t^s$. Under the assumption that f is g-smooth, we have

$$\begin{aligned}
&f(x_{t+1}^s)\\
&\le f(x_t^s)+\langle\nabla f(x_t^s),Q_\nu^{-1}(z_{t+1}^s-z_t^s)\rangle+\frac{g}{2}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-2}}^2\\
&=f(x_t^s)+\langle\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)+u_t^s,Q_\nu^{-1}(z_{t+1}^s-z_t^s)\rangle\\
&\quad+\frac{g}{2}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-2}}^2\\
&\quad+\langle\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)-u_t^s,Q_\nu^{-1}(z_{t+1}^s-z_t^s)\rangle\\
&\overset{①}{\le} f(x_t^s)+\langle\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)+u_t^s,Q_\nu^{-1}(z_{t+1}^s-z_t^s)\rangle\\
&\quad+\frac{g}{2}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-2}}^2+\frac{1-\eta g}{2\eta}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-1}}^2\\
&\quad+\frac{\eta}{2(1-\eta g)}\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)-u_t^s\|_{Q_\nu^{-1}}^2\\
&=f(x_t^s)+\langle\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)+u_t^s,x^*-Q_\nu^{-1}z_t^s\rangle\\
&\quad+\frac{g}{2}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-2}}^2-\langle\mathcal{G}_t,x^*-Q_\nu^{-1}z_{t+1}^s\rangle\\
&\quad+\frac{\eta}{2(1-\eta g)}\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)-u_t^s\|_{Q_\nu^{-1}}^2\\
&\quad+\frac{1-\eta g}{2\eta}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-1}}^2\\
&\overset{②}{\le} f(x_t^s)+\langle\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)+u_t^s,x^*-x_t^s\rangle\\
&\quad+\frac{1}{\eta}\langle\widehat{G}Q_\nu^{-1}(z_{t+1}^s-z_t^s),x^*-Q_\nu^{-1}z_{t+1}^s\rangle\\
&\quad+\langle\beta A^T\lambda_{t+1}^s,x^*-Q_\nu^{-1}z_{t+1}^s\rangle+\frac{1}{2\eta}\|z_{t+1}^s-z_t^s\|_{Q_\nu^{-1}}^2\\
&\quad+\frac{\eta}{2(1-\eta g)}\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)-u_t^s\|_{Q_\nu^{-1}}^2. \quad (11)
\end{aligned}$$

Above, ① holds due to the Young's inequality, and ② holds due to the optimal condition (10) and $\|z_{t+1}-z_t\|_{Q_\nu^{-2}}^2\le\|z_{t+1}-z_t\|_{Q_\nu^{-1}}^2$, which can be obtained by Property 3.

• **Perturbation Step**: Given x_t^s, taking expectation w.r.t. $\mathcal{B}_t$ and u_t^s, we have

$$\begin{aligned}
&\mathbb{E}\Big[\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)-u_t^s\|_{Q_\nu^{-1}}^2\Big]\\
&=\mathbb{E}\Big[\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)\|_{Q_\nu^{-1}}^2+\|u_t^s\|_{Q_\nu^{-1}}^2\Big]\\
&\le\mathbb{E}\Big[\|\nabla f(x_t^s)-\widetilde{\nabla}f_{\mathcal{B}_t}(x_t^s)\|^2+\|u_t^s\|_{Q_\nu^{-1}}^2\Big]\\
&\le 4g\alpha(b)[f(x_t^s)-f(x^*)-\langle\nabla f(x^*),x_t^s+\widetilde{x}^{s-1}-2x^*\rangle\\
&\quad+f(\widetilde{x}^{s-1})-f(x^*)]+\tau d_1\sigma^2 \quad (12)
\end{aligned}$$

where $\tau=\frac{1}{d_1}\sum_{i=1}^{d_1}\frac{1}{1+2\nu-2\nu\cos(2\pi i/d_1)}$, the first inequality holds due to Property 1, and the second inequality holds due to the variance upper bound in Lemma 10 and the estimate of $\mathbb{E}[\|u_t^s\|_{Q_\nu^{-1}}^2]$ in Lemma 7, which are provided in the Supplementary Material.

• **One-Iteration Upper Bound**: Using the above analysis with $\eta g\le 1/2$, taking expectation from both sides of the

inequality (11) w.r.t. $\mathcal{B}_t$ and u_t^s, we have

$$\begin{aligned}
&\mathbb{E}[f(x_{t+1}^s)] \\
&\le \mathbb{E}\left[f(x_t^s) + \langle \widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) + u_t^s, x^* - x_t^s\rangle\right] \\
&\quad + \mathbb{E}\left[\frac{1}{\eta}\langle \widehat{G} Q_v^{-1}(z_{t+1}^s - z_t^s), x^* - Q_v^{-1} z_{t+1}^s\rangle\right] \\
&\quad + \mathbb{E}\Big[\frac{\eta}{2(1-\eta g)}\|\nabla f(x_t^s) - \widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) - u_t^s\|_{Q_v^{-1}}^2 \\
&\quad + \frac{1}{2\eta}\|z_{t+1}^s - z_t^s\|_{Q_v^{-1}}^2 + \langle \beta A^T \lambda_{t+1}^s, x^* - Q_v^{-1} z_{t+1}^s\rangle\Big] \\
&\overset{①}{\le} \mathbb{E}\left[f(x^*) + \frac{1}{2\eta}(\|x^* - Q_v^{-1} z_t^s)\|_{\widehat{G}}^2 - \|x^* - Q_v^{-1} z_{t+1}^s\|_{\widehat{G}}^2)\right] \\
&\quad + \mathbb{E}\Big[\frac{\eta}{2(1-\eta g)}\|\nabla f(x_t^s) - \widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) - u_t^s\|_{Q_v^{-1}}^2\Big] \\
&\quad + \mathbb{E}\left[\langle \beta A^T \lambda_{t+1}^s, x^* - x_{t+1}^s\rangle\right] \\
&\overset{②}{\le} \mathbb{E}\left[f(x^*) + \frac{1}{2\eta}(\|x^* - Q_v^{-1} z_t^s)\|_{\widehat{G}}^2 - \|x^* - Q_v^{-1} z_{t+1}^s\|_{\widehat{G}}^2)\right] \\
&\quad + \mathbb{E}\left[\langle \beta A^T \lambda_{t+1}^s, x^* - x_{t+1}^s\rangle\right] \\
&\quad + 4g\eta\alpha(b)[f(x_t^s) - f(x^*) - \langle \nabla f(x^*), x_t^s - x^*\rangle \\
&\quad + f(\widetilde{x}^{s-1}) - f(x^*) - \langle \nabla f(x^*), \widetilde{x}^{s-1} - x^*\rangle] + \eta\tau d_1\sigma^2
\end{aligned}$$

where ① holds due to the convexity of f and $\mathbb{E}[\widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) + u_t^s] = \nabla f(x_t^s)$, i.e., $\mathbb{E}[\langle \widetilde{\nabla} f_{\mathcal{B}_t}(x_t^s) + u_t^s, x^* - x_t^s\rangle] \le f(x^*) - f(x_t^s)$ and Property 4 in the Supplementary Material with $w_1 = Q_v^{-1} z_{t+1}^s$, $w_2 = Q_v^{-1} z_t^s$, $w_3 = x^*$ and $\|z_{t+1}^s - z_t^s\|_{Q_v^{-1}}^2 \le \|Q_v^{-1}(z_{t+1}^s - z_t^s)\|_{\widehat{G}}^2$ (which is obtained by using Property 2); ② uses the inequality (12) and the assumption $\eta \le \frac{1}{2g}$.

Using the optimality condition (i.e., $\nabla f(x^*) + \beta A^T \lambda^* = 0$) of Problem (1) and $\varphi_t^s = \beta(\lambda_{t+1}^s - \lambda^*)$, we have

$$\begin{aligned}
&\left\langle \beta A^T \lambda_{t+1}^s,\ x^* - x_{t+1}^s \right\rangle \\
&\quad = \left\langle \nabla f(x^*),\ x_{t+1}^s - x^* \right\rangle + \left\langle \beta A^T \lambda^*,\ x_{t+1}^s - x^* \right\rangle \\
&\qquad + \left\langle \beta A^T \lambda_{t+1}^s,\ x^* - x_{t+1}^s \right\rangle \\
&\quad = \left\langle \nabla f(x^*),\ x_{t+1}^s - x^* \right\rangle + \left\langle A^T \varphi_t^s,\ x^* - x_{t+1}^s \right\rangle.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathbb{E}[f(x_{t+1}^s) - f(x^*) - \langle \nabla f(x^*), x_{t+1}^s - x^*\rangle \\
&\qquad - \langle A^T \varphi_t^s, x^* - x_{t+1}^s\rangle] \\
&\quad \le 4g\eta\alpha(b)[f(x_t^s) - f(x^*) - \langle \nabla f(x^*), x_t^s - x^*\rangle \\
&\qquad + f(\widetilde{x}^{s-1}) - f(x^*) - \langle \nabla f(x^*), \widetilde{x}^{s-1} - x^*\rangle] \\
&\qquad + \frac{1}{2\eta}\mathbb{E}\left[\|x^* - x_t^s\|_{\widehat{G}}^2 - \|x^* - x_{t+1}^s\|_{\widehat{G}}^2\right] + \tau\eta d_1\sigma^2.
\end{aligned}$$

Summing over $t = 0, \ldots, m-1$ and dividing by m, we have

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{m}\sum_{t=0}^{m-1}(f(x_{t+1}^s) - f(x^*) - \langle \nabla f(x^*), x_{t+1}^s - x^*\rangle \\
&\qquad - \langle A^T \varphi_t^s, x^* - x_{t+1}^s\rangle)\Big] \\
&\quad \le \frac{4g\eta\alpha(b)}{m}\sum_{t=0}^{m-1}\big[f(x_t^s) - f(x^*) - \langle \nabla f(x^*), x_t^s - x^*\rangle \\
&\qquad + f(\widetilde{x}^{s-1}) - f(x^*) - \langle \nabla f(x^*), \widetilde{x}^{s-1} - x^*\rangle\big] \\
&\qquad + \frac{1}{2m\eta}\mathbb{E}\left[\|x^* - x_0^s\|_{\widehat{G}}^2 - \|x^* - x_m^s\|_{\widehat{G}}^2\right] + \tau\eta d_1\sigma^2. \quad (13)
\end{aligned}$$

Subtracting $\mathbb{E}[\frac{4g\eta\alpha(b)}{m}\sum_{t=0}^{m-1}(f(x_{t+1}^s) - f(x^*) - \langle \nabla f(x^*), x_{t+1}^s - x^*\rangle)]$ from both sides of the inequality (12), we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{m}\sum_{t=0}^{m-1}((1 - 4g\eta\alpha(b))(f(x_{t+1}^s) - f(x^*) \\
&\qquad - \langle \nabla f(x^*), x_{t+1}^s - x^*\rangle) - \langle A^T \varphi_t^s, x^* - x_{t+1}^s\rangle)\Big] \\
&\quad \le 4g\eta\alpha(b)\left(f(\widetilde{x}^{s-1}) - f(x^*) - \langle \nabla f(x^*), \widetilde{x}^{s-1} - x^*\rangle\right) \\
&\qquad + \frac{4g\eta\alpha(b)}{m}\left(f(x_0^s) - f(x^*) - \langle \nabla f(x^*), x_0^s - x^*\rangle\right) \\
&\qquad - \frac{4g\eta\alpha(b)}{m}\mathbb{E}\left[f(x_m^s) - f(x^*) - \langle \nabla f(x^*), x_m^s - x^*\rangle\right] \\
&\qquad + \frac{1}{2m\eta}\mathbb{E}\left[\|x^* - x_0^s\|_{\widehat{G}}^2 - \|x^* - x_m^s\|_{\widehat{G}}^2\right] + \tau\eta d_1\sigma^2.
\end{aligned}$$

This completes the proof. □

Note that Lemma 4 is our key intermediate result for the utility guarantees of DP-VRADMM. As shown in the proof of Lemma 4, by using the idea of variable substitution and some properties of Q_v, we manage to get rid of some extra terms introduced by Q_v and obtain the desired utility bounds.

Furthermore, we can obtain the one-epoch upper bound by using Lemmas 4 and 12 in the Supplementary Material, and combine it with the privacy bound from Theorem 1 to derive our main results in Theorems 3 and 4, respectively. The detailed proofs are provided in the Supplementary Material. The utility guarantee of DP-VRADMM for general convex objectives is given in the following theorem.

Theorem 3 (DP-VRADMM for GC Problems): Suppose that $f(\cdot), r(\cdot)$ are convex, and each component function f_i is l-Lipschitz and g-smooth. Given $\epsilon, \delta > 0$, σ is defined in Theorem 1. If we choose $\eta = \Theta(\frac{1}{g}) \le \frac{1}{2g}$ and $m = \Theta(g)$, then the following result holds for $S = \mathcal{O}(\frac{n\epsilon}{l\sqrt{\tau d_1 \log(1/\delta)}})$,

$$\mathbb{E}[R(\overline{x}, \overline{y}) + \zeta\|A\overline{x} + B\overline{y} - c\|] \le \mathcal{O}\Big(\frac{l\sqrt{\tau d_1 \log(1/\delta)}}{\epsilon n}\Big)$$

where $\tau = \frac{1}{d_1}\sum_{i=1}^{d_1}\frac{1}{1+2\nu-2\nu\cos(2\pi i/d_1)}$, and $\zeta > 0$ is a given constant. The overall gradient complexity of DP-VRADMM for GC problems is $\mathcal{O}(\frac{gn\epsilon}{l\sqrt{\tau d_1 \log(1/\delta)}})$.

When $\nu = 0$, DP-VRADMM becomes the algorithm without Laplacian smoothing, and its utility bound is given below.

Corollary 1 (Without Laplacian smoothing): Using the same notation and setting as in Theorem 3 with $\nu = 0$, we have

$$\mathbb{E}[R(\bar{x}, \bar{y}) + \zeta\|A\bar{x} + B\bar{y} - c\|] \le \mathcal{O}\Big(\frac{l\sqrt{d_1 \log(1/\delta)}}{\epsilon n}\Big).$$

Remark 1: Using the property in [27], we have $\tau = \frac{1+p^{d_1}}{(1-p^{d_1})\sqrt{4\nu+1}} \to \frac{1}{\sqrt{4\nu+1}}$ as $d_1 \to \infty$, where $0 < p = \frac{2\nu+1-\sqrt{4\nu+1}}{2\nu} < 1$. This means that τ is rapidly decreasing with the increasing of d_1 and ν, and it is much smaller than 1. That is, the utility bound in Theorem 3 is tighter than that in Corollary 1, which means that Laplacian smoothing can improve the utility bound in theory. In practice, our experimental results

also verify the efficiency of the smoothing operator. As we will show in Remark 2, compared with DP-SADMM under the same private budget (ϵ, δ)*, DP-VRADMM can reduce the gradient complexity from* $\mathcal{O}(\frac{n^2\epsilon^2}{\tau d_1 l^2 \log(1/\delta)})$ *to* $\mathcal{O}(\frac{gn\epsilon}{l\sqrt{\tau d_1 \log(1/\delta)}})$ *for GC problems. The utility bound of DP-VRADMM is* $\mathcal{O}(\frac{l\sqrt{\tau d_1 \log(1/\delta)}}{\epsilon n})$*, which matches the optimal utility bound of DP-ERM [1], [9] for the GC case.*

Below we also analyze the utility bound of DP-VRADMM for strongly convex objectives.

Theorem 4 (DP-VRADMM for SC Problems): Suppose that f *is* μ*-strongly convex,* r *is convex, and each component function* f_i *is* l*-Lipschitz and* g*-smooth, and* A *has full row rank. Given* $\epsilon, \delta > 0$*,* σ *is defined in Theorem 1. If we choose* $\eta = \Theta(\frac{1}{g}) < \frac{1}{16g}$*,* $S = \mathcal{O}(\log(\frac{n^2\epsilon^2\mu}{d_1 l^2 \log(1/\delta)}))$*, and sufficiently large* m *so that they satisfy the inequality*

$$\frac{\gamma \|Q_v\|_2}{\mu \varrho m \eta} + \frac{4g\eta(m+1)\alpha(b)}{\varrho m} + \frac{g}{\beta \varrho \rho_{\min}(AA^T)m} < \frac{1}{2}$$

where $\varrho = 1 - 4g\eta\alpha(b)$*, then the following result holds*

$$\mathbb{E}[R(\bar{x}, \bar{y})] \leq \mathcal{O}\Big(\frac{\tau d_1 l^2 \log(n)\log(1/\delta)}{\mu\epsilon^2 n^2}\Big).$$

Furthermore, the total gradient complexity of DP-VRADMM is $\mathcal{O}((n + g/\mu)\log\frac{n\epsilon\mu}{d_1})$.

Theorem 4 shows that DP-VRADMM has a significantly faster convergence rate than DP-SADMM (i.e., a linear convergence rate), while DP-SADMM only attains a sub-linear rate (see Theorem 6 below). As discussed in Remark 1, the utility bound of Theorem 4 also matches the optimal utility bound of DP-ERM [1], [9] for the SC case. And DP-VRADMM with Laplacian smoothing has a much better upper bound for SC problems, which also verifies the importance of the Laplacian smoothing operator.

B. Utility Bounds for DP-SADMM

We also analyze the utility guarantees of DP-SADMM for both SC and GC objectives. Before presenting our main results for DP-SADMM, we give the following core lemma, which is applicable in both SC and GC cases.

Lemma 5: For Algorithm 2, we have the following result,

$$\begin{aligned}&\mathbb{E}\big[f(x_{t+1}) - f(x^*) + \langle \nabla f(x^*),\ x^* - x_{t+1}\rangle \\ &\quad - \langle A^T\varphi_t,\ x^* - x_{t+1}\rangle\big] \\ &\leq \frac{1}{2\eta_{t+1}}\mathbb{E}\big[\|x^* - x_t\|^2_{\widehat{G}} - \|x^* - x_{t+1}\|^2_{\widehat{G}}\big] \\ &\quad + \frac{\eta_{t+1}}{2(1-\eta_{t+1}g)}\big(l^2/b + d_1\tau\sigma^2\big)\end{aligned}$$

where $\widehat{G} = Q_v\widetilde{G} = \gamma\, Q_v - \eta_{t+1}\beta A^T A$ *and* $\varphi_t = \beta(\lambda_{t+1} - \lambda^*)$.

Note that the proof of Lemma 5 is similar to that of Lemma 4. Due to page limit, we provide the detailed proof of Lemma 5 in the Supplementary Material.

Below we first give the utility bound of DP-SADMM for general convex problems.

Theorem 5 (DP-SADMM for GC Problems): Suppose that f, r *are convex, and each component function* f_i *is* l*-Lipschitz and* g*-smooth. Given* $\epsilon, \delta > 0$*, under the same conditions in Theorem 2 on* σ *and* α*, if we choose* $\eta_{t+1} = 1/\sqrt{t+1} = \mathcal{O}(1/\sqrt{t}) \leq 1/(2g)$ *and* $T = \mathcal{O}(\epsilon^2 n^2/(\tau d_1 l^2 \log(1/\delta)))$*, the output* $\overline{x} = \sum_{t=1}^T x_t/T$*,* $\overline{y} = \sum_{t=1}^T y_t/T$ *satisfies the following utility bound,*

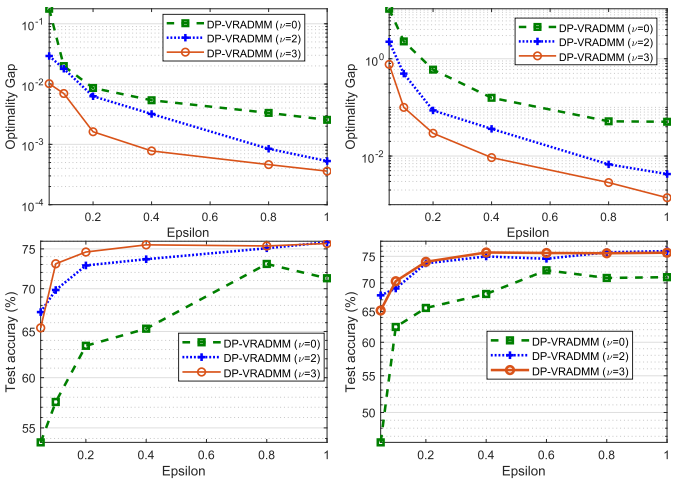
$$\mathbb{E}\Big[R(\overline{x}, \overline{y}) + \zeta\|A\bar{x} + B\bar{y} - c\|\Big] \leq \mathcal{O}\Big(\frac{l\sqrt{\tau d_1 \log(1/\delta)}}{\epsilon n}\Big).$$

Furthermore, we also analyze the utility bound of DP-SADMM for strongly convex objectives as follows.

Theorem 6 (DP-SADMM for SC Problems): Suppose that f *is* μ*-strongly convex,* r *is convex, and each component function* f_i *is* l*-Lipschitz and* g*-smooth. Given* $\epsilon, \delta > 0$*, under the same conditions in Theorem 2 on* σ *and* α*, if we choose* $\eta_{t+1} = 2\gamma\|Q_v\|_2/(\mu(t+1)) = \mathcal{O}(1/(\mu t))$ *and* $T = \mathcal{O}(\mu\epsilon^2 n^2/(\gamma\|Q_v\|_2 \tau d_1 l^2 \log(1/\delta)))$*, the output* $\overline{x} = 1/(\sum_{t=1}^T t)\sum_{t=1}^T t x_t$*,* $\overline{y} = 1/(\sum_{t=1}^T t)\sum_{t=1}^T t y_t$ *satisfies the following utility bound,*

$$\mathbb{E}[R(\overline{x}, \overline{y})] \leq \mathcal{O}\Big(\frac{\tau d_1 l^2 \log(n)\log(1/\delta)}{\mu\epsilon^2 n^2}\Big).$$

Remark 2: As discussed in Remark 1, by using Laplacian smoothing in our DP-SADMM algorithm, the upper bounds in Theorems 5 and 6 are also improved by a factor τ*, which is much smaller than 1. Moreover, the utility bounds of Theorems 5 and 6 also match the optimal utility bound of DP-ERM [1], [9] for both GC and SC cases.*

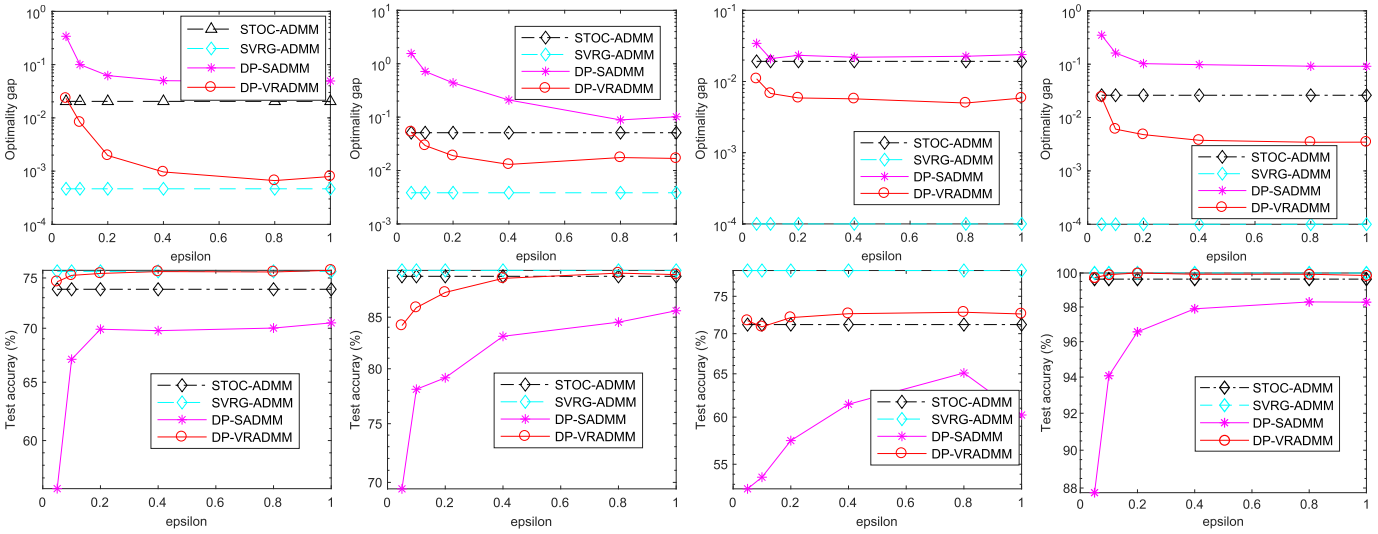


(a) covtype (b) w8a

Fig. 1. Comparison of DP-VRADMM with different Laplacian smoothing coefficients on the two datasets. Top: Objective gap vs. budget; Bottom: Test accuracy vs. budget.

TABLE II
SUMMARY OF DATASETS AND REGULARIZATION PARAMETERS USED IN OUR EXPERIMENTS

Datasets	♯ training	♯ test	♯ mini-batch	λ_1	λ_2
covtype	290,506	290,506	50	1e-4	1e-4
w8a	24,875	24,874	20	1e-4	1e-5
bio-train	72,876	72,875	20	1e-5	1e-4
mushrooms	4,062	4,062	10	1e-5	1e-5



(a) covtype (b) w8a (c) bio-train (d) mushrooms

Fig. 2. Comparison of all the algorithms for solving GC graph-guided fused Lasso problems on the four datasets. Top: Objective gap vs. DP budget, ϵ; Bottom: Test accuracy vs. DP budget, ϵ.

V. Experimental Results

In this section, we evaluate the performance of our DP-VRADMM and DP-SADMM algorithms for solving the general convex (GC) graph-guided fused Lasso and strongly convex (SC) graph-guided logistic regression problems. We report the experimental results for both GC and SC problems on some real-world datasets, as shown in Table II. All the datasets can be downloaded from the LIBSVM Data website.[1] All the experiments were performed on a PC with an Intel Core i7-7700 3.6GHz and 32GB RAM.

A. Effectiveness of Laplacian Smoothing

In this subsection, we first test the effectiveness of the Laplacian smoothing operator used in our DP-VRADMM algorithm for solving structured sparse coding problems. We report the experimental results of our DP-VRADMM with different Laplacian smoothing coefficients (e.g., $\nu = 0, 2, 3$) for solving the following graph-guided fused Lasso problem,

$$\min_x \left\{ \frac{1}{n} \sum_{i=1}^{n} f_i(x) + \lambda_1 \|y\|_1, \quad \text{s.t., } Ax = y \right\}$$

where f_i is the logistic loss function on the feature-label pair (a_i, b_i), i.e., $\log(1 + \exp(-b_i a_i^T x))$, and $\lambda_1 \geq 0$ is a regularization parameter, which is given in Table II. The matrix A is set to $A = [\phi; I]$ as in [28] and [21], where ϕ is the sparsity pattern of the graph obtained by sparse inverse covariance selection [44]. Note that when $\nu = 0$, DP-VRADMM becomes its variant without Laplacian smoothing, i.e., a common differentially private variance reduced stochastic ADMM algorithm.

Fig. 1 plots the objective gap (i.e., the objective value minus the minimum value) and the test accuracy of our DP-VRADMM with different LS coefficients on *covtype* and *w8a*. All the results show that DP-VRADMM with $\nu = 2$ and $\nu = 3$ performs significantly better than DP-VRADMM with $\nu = 0$ (i.e., DP-VRADMM without Laplacian smoothing) in terms of convergence quality and test accuracy, which shows the importance of Laplacian smoothing for differentially private stochastic ADMMs. In fact, there are similar experimental phenomena for the proposed DP-SADMM algorithm and other structured sparse coding problems. This verifies that Laplacian smoothing can smooth the injected Gaussian noise in the differentially private stochastic ADMMs, greatly reduce the impact of the noise to improve their convergence and enable the learned model to generalize better.

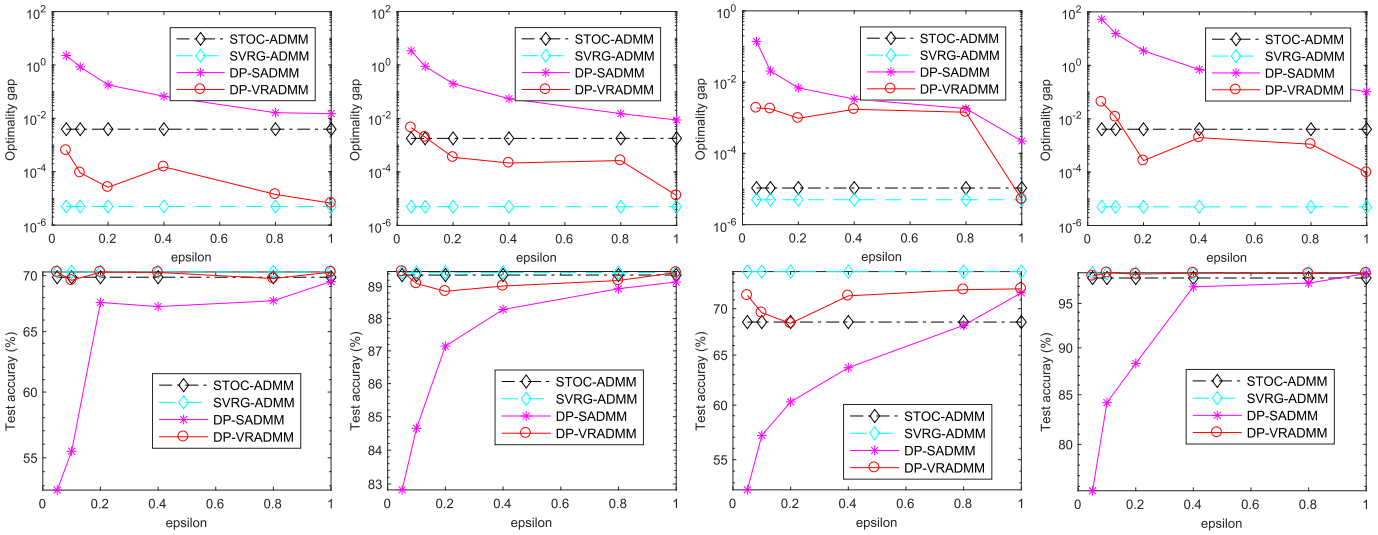
B. Graph-Guided Fused Lasso

We also evaluate the performance of our DP-VRADMM and DP-SADMM for solving the GC graph-guided fused Lasso problem. Note that SVRG-ADMM [21] and STOC-ADMM [28] are used as non-private baselines, and our DP-VRADMM and DP-SADMM are their differentially private variants, respectively. The regularization parameter λ_1 is given in Table II. Moreover, we set $m = 2n/b$ and $\gamma = 1$ as in [21]. Fig. 2 shows the experimental results (including the objective gap and test accuracy) of all the methods with different private budgets $\epsilon \in \{0.05, 0.1, 0.2, 0.4, 0.8, 1\}$ and $\delta = 5 \times 10^{-4}$ on the four datasets. It is clear that DP-VRADMM performs much better than DP-SADMM in terms of both convergence quality and test accuracy, which verified our theoretical results that the former has a faster convergence rate than the latter.

C. Graph-Guided Logistic Regression

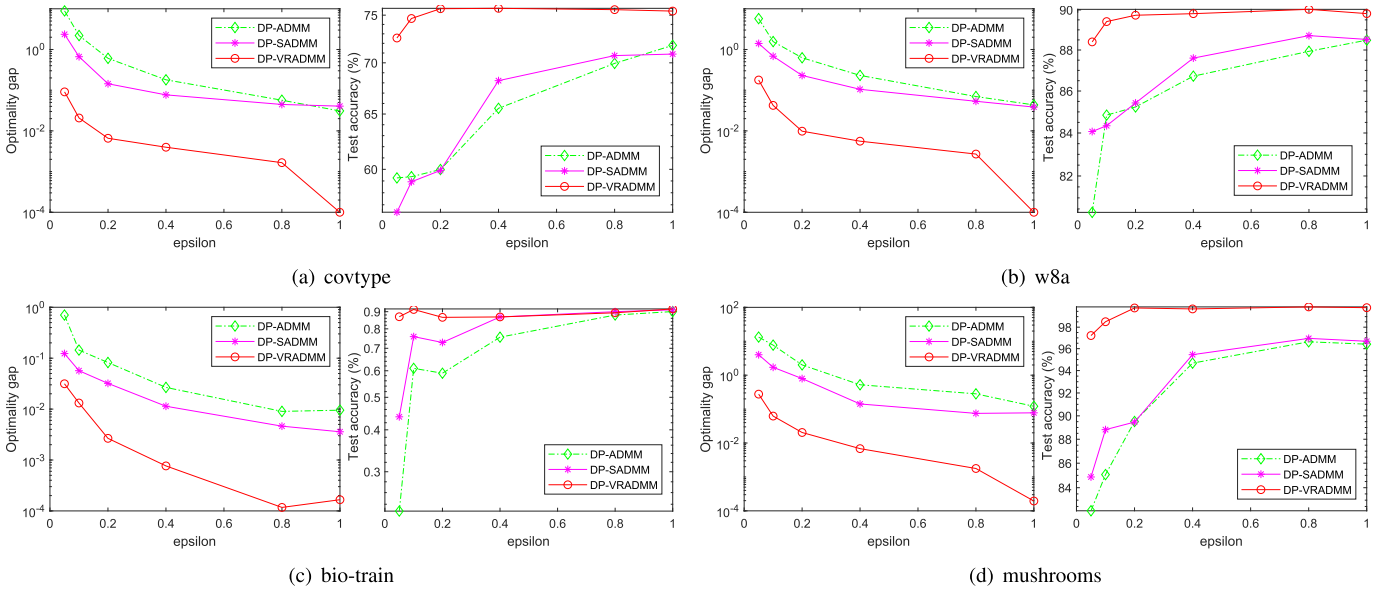
Moreover, we evaluate the performance of our DP-VRADMM and DP-ADMM algorithms for solving

[1] https://www.csie.ntu.edu.tw/ cjlin/libsvm/



(a) covtype (b) w8a (c) bio-train (d) mushrooms

Fig. 3. Comparison of all the algorithms for solving SC graph-guided logistic regression problems on the four datasets. Top: Objective gap vs. DP budget, ϵ; Bottom: Test accuracy vs. DP budget, ϵ.



(a) covtype (b) w8a

(c) bio-train (d) mushrooms

Fig. 4. Comparison of all the algorithms for solving Lasso problems on the four datasets. Left: Objective gap vs. DP budget, ϵ; Right: Test accuracy vs. DP budget, ϵ.

the following SC graph-guided logistic regression problem,

$$\min_x \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(f_i(x) + \frac{\lambda_2}{2} \|x\|^2 \right) + \lambda_1 \|y\|_1, \text{ s.t. } Ax = y \right\}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are two regularization parameters, which are given in Table II. From Fig. 3, we can see that DP-VRADMM also performs significantly better than DP-SADMM in terms of both convergence quality and test accuracy for the SC problem.

D. Lasso

Finally, we conduct some experiments to compare our DP-VRADMM and DP-SADMM methods with the recently proposed method, DP-ADMM [15], for solving the Lasso problem (the problem (1) with the constraint $x = y$, $\min_x \{\frac{1}{n}\sum_{i=1}^{n} f_i(x) + \lambda_1 \|y\|_1, \text{s.t.}, x = y\}$). When the regularization parameter is set to 10^{-4} (i.e., $\lambda_1 = 10^{-4}$), the experimental results are shown in Fig. 4. All the results show that our DP-VRADMM and DP-SADMM methods significantly outperform DP-ADMM in terms of both objective gap and test accuracy, which further verified the importance of Laplacian smoothing for differentially private ADMMs. Moreover, our DP-VRADMM method performs much better than other methods including DP-ADMM and DP-SADMM.

VI. Conclusion and Future Work

In this paper, we proposed an efficient differentially private stochastic ADMM framework for solving equality constrained minimization problems. Then we introduced

Laplacian smoothing into our private stochastic ADMM algorithms to smooth out the injected noise. Moreover, we proposed a new differentially private variance reduction stochastic ADMM (DP-VRADMM) algorithm for both SC and GC objectives. As a by-product, we also presented a new differentially private stochastic ADMM (DP-SADMM) algorithm. Moreover, we provided the theoretical guarantees for both our algorithms. Theoretical and experimental results showed that Laplacian smoothing can improve the utility bounds of our algorithms. In the future, we will extend our algorithms to non-convex objectives or other general equality constrained problems for various machine learning applications. Furthermore, we will also introduce momentum techniques to further accelerate the proposed algorithms and use parallel mechanisms as in [45] for large-scale machine learning problems.

ACKNOWLEDGMENT

The authors would like to thank all the reviewers for their valuable comments.

REFERENCES

[1] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *Proc. IEEE 55th Annu. Symp. Found. Comput. Sci.*, Oct. 2014, pp. 464–473.

[2] X. Wu, F. Li, A. Kumar, K. Chaudhuri, S. Jha, and J. Naughton, "Bolt-on differential privacy for scalable stochastic gradient descent-based analytics," in *Proc. ACM Int. Conf. Manage. Data*, May 2017, pp. 1307–1322.

[3] J. Zhang, K. Zheng, W. Mou, and L. Wang, "Efficient private ERM for smooth objectives," 2017, *arXiv:1703.09947*.

[4] K. Chaudhuri, C. Monteleoni, and A. D. Sarwate, "Differentially private empirical risk minimization," *J. Mach. Learn. Res.*, vol. 12, pp. 1069–1109, Mar. 2011.

[5] D. Kifer, A. Smith, and A. Thakurta, "Private convex empirical risk minimization and high-dimensional regression," in *Proc. Int. Conf. Learn. Theory*, 2012, pp. 1–25.

[6] R. Iyengar, J. P. Near, D. Song, O. Thakkar, A. Thakurta, and L. Wang, "Towards practical differentially private convex optimization," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2019, pp. 299–316.

[7] D. Yu, H. Zhang, W. Chen, T.-Y. Liu, and J. Yin, "Gradient perturbation is underrated for differentially private convex optimization," 2019, *arXiv:1911.11363*.

[8] S. Song, K. Chaudhuri, and A. D. Sarwate, "Stochastic gradient descent with differentially private updates," in *Proc. Global Conf. Signal Inf. Process. (GlobalSIP)*, 2013, pp. 245–248.

[9] D. Wang, M. Ye, and J. Xu, "Differentially private empirical risk minimization revisited: Faster and more general," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 2719–2728.

[10] J. Lee and D. Kifer, "Concentrated differentially private gradient descent with adaptive per-iteration privacy budget," in *24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (SIGKDD)*, 2018, pp. 1656–1665.

[11] B. Wang, Q. Gu, M. Boedihardjo, L. Wang, F. Barekat, and S. J. Osher, "DP-LSSGD: A stochastic optimization method to lift the utility in privacy-preserving ERM," in *Proc. 1st Conf. Math. Sci. Mach. Learn.*, 2020, pp. 328–351.

[12] Z. Liang, B. Wang, Q. Gu, S. Osher, and Y. Yao, "Differentially private federated learning with Laplacian smoothing," 2020, *arXiv:2005.00218*.

[13] S. Kim, K.-A. Sohn, and E. P. Xing, "A multivariate regression approach to association analysis of a quantitative trait network," *Bioinformatics*, vol. 25, no. 12, pp. i204–i212, Jun. 2009.

[14] R. J. Tibshirani and J. Taylor, "The solution path of the generalized lasso," *Ann. Statist.*, vol. 39, no. 3, pp. 1335–1371, 2011.

[15] C. Chen and J. Lee, "Rényi differentially private ADMM for non-smooth regularized optimization," in *Proc. 10th ACM Conf. Data Appl. Secur. Privacy*, 2020, pp. 319–328.

[16] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *SIAM J. Imag. Sci.*, vol. 7, no. 3, pp. 1588–1623, 2014.

[17] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Math. Program.*, vol. 162, pp. 165–199, Mar. 2017.

[18] W. Tian and X. Yuan, "An alternating direction method of multipliers with a worst-case 0 $(1/n^2)$ convergence rate," *Math. Comput.*, vol. 88, no. 318, pp. 1685–1713, 2019.

[19] L. W. Zhong and J. T. Kwok, "Fast stochastic alternating direction method of multipliers," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 46–54.

[20] T. Suzuki, "Stochastic dual coordinate ascent with alternating direction method of multipliers," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 736–744.

[21] S. Zheng and J. T. Kwok, "Fast-and-light stochastic ADMM," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2407–2413.

[22] Y. Liu, F. Shang, H. Liu, L. Kong, L. Jiao, and Z. Lin, "Accelerated variance reduction stochastic ADMM for large-scale machine learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4242–4255, Dec. 2021.

[23] Y. Liu, F. Shang, and J. Cheng, "Accelerated variance reduced stochastic ADMM," in *Proc. AAAI Conf. Artif. Intell.*, 2017, pp. 2287–2293.

[24] P. Wang and H. Zhang, "Differential privacy for sparse classification learning," *Neurocomputing*, vol. 375, pp. 91–101, Jan. 2020.

[25] J. Ding, X. Zhang, M. Chen, K. Xue, C. Zhang, and M. Pan, "Differentially private robust ADMM for distributed machine learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 1302–1311.

[26] X. Zhang, M. M. Khalili, and M. Liu, "Recycled ADMM: Improving the privacy and accuracy of distributed algorithms," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1723–1734, 2020.

[27] S. Osher *et al.*, "Laplacian smoothing gradient descent," 2018, *arXiv:1806.06317*.

[28] H. Ouyang, N. He, L. Q. Tran, and A. Gray, "Stochastic alternating direction method of multipliers," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 80–88.

[29] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *J. Privacy Confidentiality*, vol. 7, no. 3, pp. 17–51, 2016.

[30] C. Dwork and A. Roth, *The Algorithmic Foundations of Differential Privacy*. New York, NY, USA: Now, 2014.

[31] I. Mironov, "Rényi differential privacy," in *Proc. IEEE 30th Comput. Secur. Found. Symp. (CSF)*, Aug. 2017, pp. 263–275.

[32] L. Wang, B. Jayaraman, D. Evans, and Q. Gu, "Efficient privacy-preserving nonconvex optimization," 2019, *arXiv:1910.13659*.

[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jul. 2010.

[34] F. Nie, Y. Huang, X. Wang, and H. Huang, "Linear time solver for primal SVM," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 505–513.

[35] H. Wang and A. Banerjee, "Online alternating direction method," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 1119–1126.

[36] T. Suzuki, "Dual averaging and proximal gradient descent for online alternating direction multiplier method," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 392–400.

[37] F. Shang *et al.*, "VR-SGD: A simple stochastic variance reduction method for machine learning," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 188–202, Jan. 2020.

[38] F. Shang, T. Xu, Y. Liu, H. Liu, L. Shen, and M. Gong, "Differentially private ADMM algorithms for machine learning," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 4733–4745, 2021.

[39] H. Li and Z. Lin, "Faster and non-ergodic O$(1/K)$ stochastic alternating direction method of multipliers," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 4479–4488.

[40] Y. Xu, M. Liu, Q. Lin, and T. Yang, "ADMM without a fixed penalty parameter: Faster convergence with new adaptive penalization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1267–1277.

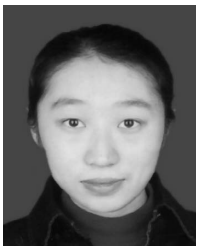
[41] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 315–323.

[42] J. Lee, "Differentially private variance reduced stochastic gradient descent," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 161–166.

[43] B. He and X. Yuan, "On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method," *SIAM J. Numer. Anal.*, vol. 50, no. 2, pp. 700–709, 2012.

[44] O. Banerjee, L. El Ghaoui, and A. d'Aspremont, "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data," *J. Mach. Learn. Res.*, vol. 9, pp. 485–516, Mar. 2008.

[45] F. Shang, H. Huang, J. Fan, Y. Liu, H. Liu, and J. Liu, "Asynchronous parallel, sparse approximated SVRG for high-dimensional machine learning," *IEEE Trans. Knowl. Data Eng.*, early access, Apr. 2, 2021, doi: 10.1109/TKDE.2021.3070539.



Yuanyuan Liu (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2013.

She is currently a Professor with the School of Artificial Intelligence, Xidian University. Prior to joining Xidian University, she was a Post-Doctoral Research Fellow with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, where she was a Post-Doctoral Research Fellow with the Department of Systems Engineering and Engineering Management from 2013 to 2014. Her current research interests include machine learning, pattern recognition, and image processing.

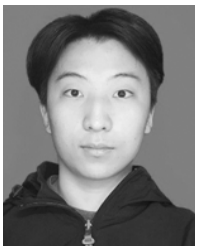
Jiacheng Geng received the master's degree majoring in electronic science and technology from the School of Artificial Intelligence, Xidian University, China, in 2021. He is currently working with Zhejiang Dahua Technology Company Ltd. His current research interests include large-scale machine learning and stochastic optimization.

Fanhua Shang (Senior Member, IEEE) received the Ph.D. degree in circuits and systems from Xidian University, Xi'an, China, in 2012.

He is currently a Professor with the School of Computer Science and Technology, College of Computing and Intelligence, Tianjin University, China. Prior to this, he was a Professor with the School of Artificial Intelligence, Xidian University, from 2018 to 2022. From 2016 to 2018, he was a Research Associate with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, where he was a Post-Doctoral Research Fellow with the Department of Computer Science and Engineering from 2013 to 2015. From 2012 to 2013, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. His current research interests include machine learning, deep learning, data mining, computer vision, and stochastic optimization.

Weixin An received the B.S. degree majoring in information and computation science from the Xi'an University of Technology in 2020. He is currently pursuing the Ph.D. degree with the School of Artificial intelligence, Xidian University, China. His current research interests include large-scale machine learning and stochastic optimization.

Hongying Liu (Senior Member, IEEE) received the B.E. and M.S. degrees in computer science and technology from the Xi'an University of Technology, China, in 2006 and 2009, respectively, and the Ph.D. degree in engineering from Waseda University, Japan, in 2012. She is currently a Faculty Member with the School of Artificial Intelligence and the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, China. Her major research interests include image processing, intelligent signal processing, and machine learning.



Qi Zhu is currently pursuing the B.S. degree in telecommunication engineering with the School of Telecommunication Engineering, Xidian University, Xi'an, China. His current research interests include stochastic optimization for machine learning, sparse signal recovery, and image processing.

Wei Feng (Member, IEEE) received the Ph.D. degree in computer science from the City University of Hong Kong in 2008.

From 2008 to 2010, he was a Research Fellow with The Chinese University of Hong Kong and the City University of Hong Kong. He is currently a Professor with the School of Computer Science and Technology, College of Computing and Intelligence, Tianjin University, China. His major research interests are active robotic vision and visual intelligence, specifically including active camera relocalization and lighting recurrence, general Markov random fields modeling, energy minimization, active 3D scene perception, SLAM, and generic pattern recognition. Recently, he focuses on solving preventive conservation problems of cultural heritages via computer vision and machine learning. He is an Associate Editor of *Neurocomputing* and *Journal of Ambient Intelligence and Humanized Computing*.