Distributed Stochastic Gradient Descent with Subsampling-Enhanced Differential Privacy and Event-Triggered Communication ⋆

Jimin Wang [a,b], Jialong Chen [d,e], Ji-Feng Zhang [c,d,e]

[a] *School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

[b] *Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China*

[c] *School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, Henan Province*

[d] *State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

[e] *School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

Abstract

This paper proposes a new distributed nonconvex stochastic optimization algorithm that can achieve privacy protection and event-triggered communication. Specifically, each node masks its local state to avoid information leakage, and then designs an event-triggered mechanism to determine whether the current masked state is transmitted to its neighbor nodes. Two masked cases are considered, one is the additive Gaussian noise, and another is the unbiased stochastic quantizer. For both cases, differential privacy analysis is given rigorously, where (ϵ, δ)-differential privacy for the additive Gaussian noise while $(0, \delta)$-differential privacy for the unbiased stochastic quantizer. By using a sample-size parameter-controlled subsampling method, both cases enhance the differential privacy level compared with the existing ones. By using a two-time-scale step-sizes method, the convergence rate and the oracle complexity of the proposed algorithm are given when the global cost function satisfies the Polyak-Łojasiewicz condition. We show the tradeoff between the privacy level, event-triggered communication and the convergence rate of the algorithm. In addition, the proposed algorithm achieves both the mean square convergence and enhanced differential privacy as the sample-size goes to infinity. A numerical example of the distributed training on the "MNIST" dataset is given to show the effectiveness and advantage of the algorithm.

Key words: Differential privacy; event-triggered communication; distributed stochastic gradient descent; convergence rate

1 Introduction

Distributed optimization is gaining more and more attraction due to its wide applications in multi-robot system, smart grids, and large-scale machine learning. In these applications, the problem can be formulated as a network of nodes cooperatively solve a common optimization problem through on-node computation and local communication (Nedic & Ozdaglar, 2009; Reisizadeh et al., 2019a,b; Xin et al., 2022; Zhang et al., 2023, 2024). Among others, distributed stochastic optimization focuses on finding optimal solutions for stochastic cost functions in a distributed manner. To solve this stochastic optimization problem, a distributed stochastic gradient descent (SGD) algorithm is one of the popular methods, and has been widely studied, such as distributed SGD with quantized communication (Reisizadeh et al., 2019a,b), distributed SGD with gradient compression (Zhang et al., 2023), and so on.

When nodes exchange information in the distributed SGD algorithm, the involved sensitive data may be leaked. Therefore, a fundamental challenge in such an optimization algorithm is to protect the privacy of the

⋆ The work is supported by National Natural Science Foundation of China under Grants 62433020, 62203045, T2293770. The material in this paper was not presented at any conference. Corresponding author: Ji-Feng Zhang.

Email addresses: `jimwang@ustb.edu.cn` (Jimin Wang), `chenjialong23@mails.ucas.ac.cn` (Jialong Chen), `jif@iss.ac.cn` (Ji-Feng Zhang).

involved sensitive data (Zhang et al., 2021; Miao et al., 2024). To solve this issue, various techniques have been employed, such as homomorphic encryption (Lu & Zhu, 2018), state decomposition (Wang, 2019), differential privacy (Dwork & Roth, 2014; Le Ny & Pappas, 2014; Liu et al., 2020b; Wang et al., 2022; Chen et al., 2023; Wang et al., 2024), and so on. Due to its simplicity and wide applicability in privacy protection, differential privacy has attracted a lot of attention and been used to solve privacy issues in distributed stochastic optimization. Up to now, there are two main kinds of differential privacy, one is the (ϵ, δ)-differential privacy masked by the additive Gaussian (Laplacian) noise, and another is the (0, δ)-differential privacy masked by the unbiased stochastic quantizer. By adding the additive Gaussian (Laplacian) noise, differentially private distributed stochastic optimization algorithms are proposed in (Zhang et al., 2018; Li et al., 2018; Huang et al., 2019; Ding et al., 2021; Gratton et al., 2021; Xu et al., 2022; Yan et al., 2023; Wang & Başar, 2023a; Liu et al., 2024). Among them, distributed convex stochastic optimization algorithms were considered in (Zhang et al., 2018; Li et al., 2018; Huang et al., 2019; Ding et al., 2021; Gratton et al., 2021; Liu et al., 2024), while distributed nonconvex stochastic optimizations were provided in (Xu et al., 2022; Wang & Başar, 2023a; Yan et al., 2023). However, (ϵ, δ)-differential privacy is only given for each iteration in Zhang et al. (2018); Li et al. (2018); Huang et al. (2019); Ding et al. (2021); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Yan et al. (2023); Liu et al. (2024), leading to infinite cumulative differential privacy budgets over infinite iterations. For the unbiased stochastic quantizer case, Wang & Başar (2023b); Liu et al. (2025) treat the dither signals in the quantizers as privacy noises, and prove that using the dithered lattice quantizer (i.e., ternary quantizer in Wang & Başar (2023b) and stochastic quantizer in Liu et al. (2025)) can achieve (0, δ)-differential privacy. Since (0, δ)-differential privacy is only given for each iteration in Wang & Başar (2023b); Liu et al. (2025), (0, 1)-differential privacy is achieved over infinite iterations. In this case, the sensitive information therein cannot be protected over infinite iterations since (0,1)-differential privacy means the algorithm directly outputs the sensitive information. Recently, the differentially private distributed stochastic nonconvex optimization with quantized communication was studied in Chen et al. (2024). Although (ϵ, δ)-differential privacy has been improved, (0, δ)-differential privacy is not considered therein.

Besides privacy preservation, another fundamental challenge in the distributed SGD algorithm is the expensiveness of communications. To solve this issue, an event-triggered communication is a well-known method (Espitia et al., 2021; Wang & Krstic, 2023). There have been some interesting works on communication-efficient distributed optimization by an event-triggered communication, such as distributed optimization for second-order continuous-time multi-agent systems (Yi et al., 2018), distributed non-convex optimization (George & Gurram, 2020; Xu et al., 2024), distributed convex constrained optimization (Liu et al., 2020a), distributed online convex optimization (Cao & Başar, 2021), distributed optimization with asynchronous computation (Dong et al., 2025), large-scale machine learning under compressed communication (Singh et al., 2023), and distributed federated learning with decaying communication rate (He et al., 2023). However, none of the above-mentioned literature takes privacy issues into account. Recently, both the event-triggered communication and the differential privacy are considered for average consensus (Wang et al., 2019; Gao et al., 2019; Liang et al., 2024), and distributed optimization (Mao et al., 2023; Yuan et al., 2024), respectively. As far as we known, the event-triggered communication and privacy preservation are seldom studied together in distributed stochastic optimization though it is of great interest.

In this paper, we are interested in designing a privacy preserving and communication-efficient algorithm for distributed stochastic optimization by using event-triggered communication. The main contributions of this paper are as follows:

- A new differentially private distributed nonconvex stochastic optimization algorithm with event-triggered communication has been proposed. In the proposed algorithm, each node masks its local state to avoid information leakage, and then designs an event-triggered mechanism to determine whether the current masked state is transmitted to its neighbor nodes. Two masked cases are considered, one is the additive Gaussian noise, and another is the unbiased stochastic quantizer.
- For both masked cases, by using a sample-size parameter-controlled subsampling method, the differential privacy level is enhanced. For the additive Gaussian noise case, (ϵ, δ)-differential privacy is enhanced compared with (Zhang et al., 2018; Li et al., 2018; Huang et al., 2019; Ding et al., 2021; Gratton et al., 2021; Xu et al., 2022; Wang & Başar, 2023a; Yan et al., 2023; Liu et al., 2024). For the unbiased stochastic quantization case, $(0, \delta)$-differential privacy is enhanced compared with (Wang & Başar, 2023b; Liu et al., 2025) while saving the transmitted rounds and bits of communication simultaneously.
- Under the Polyak-Łojasiewicz condition, the convergence rate of the algorithm is given for general privacy noises, including increasing, constant and decreasing privacy noises. This is non-trivial even without considering the privacy protection. By using a two-time-scale step-sizes method, the assumption of bounded gradients required in (Zhang et al., 2018; Li et al., 2018; Huang et al., 2019; Ding et al., 2021; Gratton et al., 2021; Wang & Başar, 2023a; Liu et al., 2024) has been removed.

This paper is organized as follows: Section 2 formulates the problem to be investigated. Section 3 presents the main results including the privacy, convergence rate and oracle complexity analysis of the algorithm. Section 4 provides a numerical example of the distributed training of a convolutional neural network on the "MNIST" dataset. Section 5 gives some concluding remarks.

Notation: $\mathbb{R}$ and $\mathbb{R}^r$ denote the set of all real numbers and r-dimensional Euclidean space, respectively. Range(F) denotes the range of a mapping F. For sequences $\{a_k\}_{k=1}^{\infty}$ and $\{b_k\}_{k=1}^{\infty}$, $a_k = O(b_k)$ means there exists $C \geq 0$ such that $\limsup_{k\to\infty} |\frac{a_k}{b_k}| \leq C$. $\mathbf{1}_n$ represents an n-dimensional vector whose elements are all 1. $A^\top$ stands for the transpose of the matrix A. $\|x\| = \sqrt{x^\top x}$ denotes the standard Euclidean norm of $x = [x_1, x_2, \ldots, x_m]^\top$, and $\|A\|$ denotes the 2-norm of the matrix A. $\mathbb{E}(X)$ refers to the expectation of a random variable X. $\otimes$ denotes the Kronecker product of matrices. For a differentiable function $f(x)$, $\nabla f(x)$ denotes its gradient at the point x.

2 Preliminaries and Problem formulation

2.1 Graph theory

Consider a network of n nodes which exchange information on an undirected and connected communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of all nodes, and $\mathcal{E}$ is the set of all edges. An edge $e_{ij} \in \mathcal{E}$ if and only if Node i can receive the information from Node j. Different nodes in $\mathcal{V}$ exchange information based on the weight matrix $\mathcal{A} = (a_{ij})_{1\leq i,j\leq n}$, whose entry a_{ij} is either positive if $e_{ij} \in \mathcal{E}$, or 0, otherwise. The neighbor set of Node i is defined as $\mathcal{N}_i = \{j \in \mathcal{V} : a_{ij} > 0\}$, and the Laplacian matrix of $\mathcal{A}$ is defined as $\mathcal{L} = \mathrm{diag}(\mathcal{A}\mathbf{1}_n) - \mathcal{A}$. The assumption about the weight matrix $\mathcal{A}$ is given as follows:

Assumption 1 *The weight matrix $\mathcal{A}$ is doubly stochastic, i.e., $\mathcal{A}\mathbf{1}_n = \mathbf{1}_n$, $\mathbf{1}_n^\top \mathcal{A} = \mathbf{1}_n^\top$.*

2.2 Distributed stochastic optimization

In this paper, the following distributed nonconvex stochastic optimization problem is considered:

$$\min_{x\in\mathbb{R}^r} F(x) = \min_{x\in\mathbb{R}^r} \frac{1}{n}\sum_{i=1}^{n} f_i(x), \quad f_i(x) = \mathbb{E}_{\xi_i\sim\mathscr{D}_i}[\ell_i(x, \xi_i)], \quad (1)$$

where x is available to all nodes, $\ell_i(x, \xi_i)$ is a local cost function which is private to Node i, and ξ_i is a random variable drawn from an unknown probability distribution $\mathscr{D}_i$. In practice, since the probability distribution $\mathscr{D}_i$ is difficult to obtain, it is replaced by the dataset $\mathcal{D}_i = \{\xi_{i,l}, 1 \leq l \leq D\}$. Then, (1) can be rewritten as the following empirical risk minimization problem:

$$\min_{x\in\mathbb{R}^r} F(x) = \min_{x\in\mathbb{R}^r} \frac{1}{n}\sum_{i=1}^{n} f_i(x), \quad f_i(x) = \frac{1}{D}\sum_{l=1}^{D} \ell_i(x, \xi_{i,l}). \quad (2)$$

To solve the empirical risk minimization problem (2), we need the following standard assumption.

Assumption 2 *(i) For any node $i \in \mathcal{V}$, f_i has Lipschitz continuous gradients, i.e., $\|\nabla f_i(x) - \nabla f_i(y)\| \leq L\|x - y\|$, $\forall x, y \in \mathbb{R}^r$, where L is a positive constant.*

(ii) Each cost function is bounded from below, i.e., $\min_{x\in\mathbb{R}^r} f_i(x) = f_i^* > -\infty$.

(iii) For any node $i \in \mathcal{V}, x \in \mathbb{R}^r$ and ζ_i uniformly sampled from $\mathcal{D}_i$, there exists a stochastic first-order oracle which returns a sampled gradient $\nabla\ell_i(x, \zeta_i)$ of $f_i(x)$. In addition, there exists $\sigma_\ell > 0$ such that each sampled gradient $\nabla\ell_i(x, \zeta_i)$ satisfies $\mathbb{E}[\nabla\ell_i(x, \zeta_i)] = \nabla f_i(x)$, $\mathbb{E}[\|\nabla\ell_i(x, \zeta_i) - \nabla f_i(x)\|^2] \leq \sigma_\ell^2$.

Remark 1 *Assumption 2(i) is commonly used (see e.g. (Zhang et al., 2018; Reisizadeh et al., 2019a; Ding et al., 2021; Xin et al., 2022; Xu et al., 2022; Yan et al., 2023; Wang & Başar, 2023a)). Assumption 2(ii) ensures the existence of the optimal solution. Assumption 2(iii) requires that each sampled gradient $\nabla\ell_i(x, \zeta_i)$ is unbiased with a bounded variance σ_ℓ^2 (see e.g. (Ding et al., 2021; Xu et al., 2022; Wang & Başar, 2023a; Yan et al., 2023)).*

Distributed SGD for solving the problem (2) was first studied and rigorously analyzed by Nedic & Ozdaglar (2009). In this algorithm, each node i iteratively updates its decision variables $x_{i,k+1}$ by combining an average of the states of its neighbors with a gradient step as follows: $x_{i,k+1} = \sum_{j\in\mathcal{N}_i} a_{ij}x_{j,k} - \alpha_k\nabla\ell_i(x_{i,k}, \zeta_{i,k})$, where α_k is the time-varying step size corresponding to the influence of the gradients on the state update rule at each time step. When using this distributed SGD algorithm to solve the distributed stochastic optimization problem, there are two key issues worthy of attention. One is the leakage of the sensitive information concerning the sampled gradient, and the other is the expensive communications. To solve the first issue, the differential privacy method masked by the additive Gaussian (Laplacian) noise and the unbiased stochastic quantizer is used. To solve the second issue, an event-triggered communication mechanism is introduced. Next, we first introduce the differential privacy method.

2.3 Differential privacy

In this paper, we consider the following adversary widely used in the privacy issue for distributed stochastic optimization (Wang & Başar, 2023a,b; Liu et al., 2024):

- A *semi-honest* adversary. This kind of adversary is defined as a node within the network which has access to certain internal states (such as $x_{i,k}$ from Node i), follows the prescribed protocols and accurately computes iterative state correctly. However, it aims to infer the sensitive information of other nodes.
- An external adversary (which also refer to an *eavesdropper*) who has capability to wiretap and monitor all communication channels, allowing them to capture distributed messages from any node. This enables the eavesdropper to infer the sensitive information of internal nodes.

When solving the empirical risk minimization problem (2), the stochastic first-order oracle needs data samples to return sampled gradients. Meanwhile, the adversaries mentioned above can infer the sensitive information of data samples from sampled gradients (Zhu et al., 2019). In order to provide privacy protection for data samples, inspired by Wang et al. (2022); Liu et al. (2024), a symmetric binary relation called *adjacency relation* is defined as follows:

Definition 1 *(Adjacency relation) Let* $\mathcal{D} = \{\xi_{i,l}, i \in \mathcal{V}, 1 \leq l \leq D\}$, $\mathcal{D}' = \{\xi'_{i,l}, i \in \mathcal{V}, 1 \leq l \leq D\}$ *be two sets of data samples. If for any given* $C > 0$ *and any* $x \in \mathbb{R}^r$, *there exists exactly one pair of data samples* $\xi_{i_0,l_0}, \xi'_{i_0,l_0}$ *in* $\mathcal{D}, \mathcal{D}'$ *such that*

$$\begin{cases} \|\nabla \ell_i(x, \xi_{i,l}) - \nabla \ell_i(x, \xi'_{i,l})\| \leq C, & \text{if } i = i_0 \text{ and } l = l_0; \\ \|\nabla \ell_i(x, \xi_{i,l}) - \nabla \ell_i(x, \xi'_{i,l})\| = 0, & \text{if } i \neq i_0 \text{ or } l \neq l_0, \end{cases} \tag{3}$$

then $\mathcal{D}$ *and* $\mathcal{D}'$ *are said to be adjacent, denoted by* $Adj(\mathcal{D}, \mathcal{D}')$.

Remark 2 *The boundary* C *characterizes the "closeness" of a pair of data samples* ξ_{i_0,l_0}, ξ'_{i_0,l_0}. *By (3), the larger the boundary* C *is, the larger the allowed magnitude of sampled gradients between adjacent datasets is.*

To give the privacy protection level of the algorithm, we adopt the definition of (ϵ, δ)-differential privacy as follows:

Definition 2 (Le Ny & Pappas, 2014) *((ϵ, δ)-differential privacy) Given* $\epsilon \geq 0, 0 < \delta \leq 1$, *a randomized algorithm* $\mathcal{M}$ *achieves* (ϵ, δ)*-differential privacy for* $Adj(\mathcal{D}, \mathcal{D}')$ *if* $\mathbb{P}(\mathcal{M}(\mathcal{D}) \in T) \leq e^{\epsilon}\mathbb{P}(\mathcal{M}(\mathcal{D}') \in T) + \delta$ *for any Borel-measurable set* $T \subseteq Range(\mathcal{M})$.

2.4 Event-triggered communication

We assume that each node can continuously monitor its own masked state and decide when to transmit its current masked state over the network based on an event-triggered mechanism. Let Φ be the event-triggered threshold for any $k \geq 0$, $\tau_{i,0} \triangleq 0$ be the first event triggering instant, and $e_{i,0} \triangleq \mathcal{C}(x_{i,0})$ be the masked state error at the zero-th iteration for any node $i \in \mathcal{V}$. Then, for any given $K \geq 1$ and $k = 1, 2, \ldots, K$, we define $e_{i,k} \triangleq \mathcal{C}(x_{i,k}) - \mathcal{C}(x_{i,\tau_{i,k}})$ as the masked state error, and $\tau_{i,k}$ as the latest event triggering instant no more than k as follows:

$$\tau_{i,k} \triangleq \begin{cases} \tau_{i,k-1}, & \text{if } \|e_{i,k-1}\| < \Phi, \\ k, & \text{if } \|e_{i,k-1}\| \geq \Phi. \end{cases} \tag{4}$$

We give the following example to further clarify the event-triggered mechanism. As shown in Fig. 1, Node 1's masked state error at the zero-th and third iteration is larger than the event-triggered threshold, and thus 0, 1, 4 are event triggering instants. In this case, $\tau_{1,k}$ is given for $k = 0, 1, 2, 3, 4$ by (4).

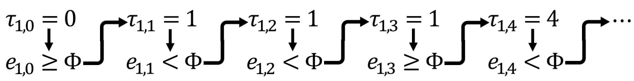


Fig. 1. An example of the event-triggered mechanism

Remark 3 *The event-triggered mechanism (4) is to determine whether the masked state is worth sharing with its neighbors by comparing it with the last sent-out masked state. To ensure the convergence, the event-triggered threshold* Φ *should decay fast enough, as required in Assumption 3. However, to avoid nodes communicating frequently all the time,* Φ *should not decay too fast.*

2.5 Oracle complexity

Since the sample-size parameter-controlled subsampling method is employed in this paper, the total number of data samples to obtain an optimal solution is an issue worthy of attention. To show this, we give the definitions of η-optimal solutions and the oracle complexity as follows:

Definition 3 *(Chen et al., 2024) (η-optimal solution) Given* $\eta > 0$ *and the global minimum* $F^* = \min_{x \in \mathbb{R}^r} F(x)$ *of the problem (2),* $x_K = [x_{1,K}^\top, \ldots, x_{n,K}^\top]^\top$ *is an* η*-optimal solution if* $\mathbb{E}|F(x_{i,K}) - F^*| < \eta$, $\forall i \in \mathcal{V}$, *where* $F(x_{i,K})$ *is the estimate of the global minimum* F^* *at the* K*-th iteration.*

Definition 4 *(Chen et al., 2024) Given* $\eta > 0$, *the oracle complexity* $\sum_{k=0}^{N(\eta)} s_k$ *is the total number of data samples to obtain an* η*-optimal solution, where* $N(\eta) = \min\{K : x_K$ *is an* η*-optimal solution*$\}$, s_k *is the sample-size at the* k*-th iteration.*

3 Main result

3.1 The proposed algorithm

In this subsection, we give a distributed SGD algorithm with differential privacy and event-triggered communication. The detailed implementation steps are given in Algorithm 1.

Algorithm 1 A distributed SGD algorithm with differential privacy and event-triggered communication

---

Initialization: $x_{i,0} \in \mathbb{R}^r$, $\forall i \in \mathcal{V}$, weight matrix $(a_{ij})_{1\leq i,j\leq n}$, iteration maximum K, step-sizes $\alpha = \frac{a_1}{K^{p_1}}$, $\beta = \frac{a_2}{K^{p_2}}$, the sample-size $s = \lfloor a_3 K^{p_3} \rfloor + 1$, and the event-triggered threshold $\Phi = \frac{a_4}{K^{p_4}}$.

for $k = 0, 1, 2, \ldots, K$, **do**

1: **If** $k = 0$, **then**
   Node i sets $\tau_{i,0} = 0$, and masks $x_{i,0}$ as $\mathcal{C}(x_{i,0})$. Then, Node i sends $\mathcal{C}(x_{i,0})$ to its neighbor $j \in \mathcal{N}_i$, and receives $\mathcal{C}(x_{j,0})$ from its neighbor $j \in \mathcal{N}_i$.
   else
      If $\|e_{i,k}\| \geq \Phi$, **then**
      Node i masks $x_{i,k}$ as $\mathcal{C}(x_{i,k})$, and then sends $\mathcal{C}(x_{i,k})$ to its neighbor $j \in \mathcal{N}_i$. Node i sets $\mathcal{C}(x_{i,\tau_{i,k}}) = \mathcal{C}(x_{i,k})$.
      else
      Node i sets $\mathcal{C}(x_{i,\tau_{i,k}}) = \mathcal{C}(x_{i,\tau_{i,k-1}})$.
      end if
   end if

2: Node i takes s different data samples $\zeta_{i,k,1}, \ldots, \zeta_{i,k,s}$ uniformly from $\mathcal{D}_i$ to generate sampled gradients $\nabla \ell_i(x_{i,k}, \zeta_{i,k,1}), \ldots, \nabla \ell_i(x_{i,k}, \zeta_{i,k,s})$. Then, Node i puts these data samples back into $\mathcal{D}_i$.

3: Node i computes the averaged sampled gradient by

$$\nabla \ell_{i,k} = \frac{1}{s} \sum_{l=1}^{s} \nabla \ell_i(x_{i,k}, \zeta_{i,k,l}). \tag{5}$$

4: Node i updates its state by

$$x_{i,k+1} = (1-\beta)x_{i,k} + \beta \sum_{j\in\mathcal{N}_i} a_{ij}\mathcal{C}(x_{j,\tau_{j,k}}) - \alpha \nabla \ell_{i,k}. \tag{6}$$

end for

3.2 Privacy analysis

In this subsection, we will show the differential privacy analysis of Algorithm 1. Inspired by Wang et al. (2022), we first provide the sensitivity of the algorithm, which helps us to analyze the differential privacy of the algorithm.

Definition 5 *(Sensitivity) Given $Adj(\mathcal{D}, \mathcal{D}')$, and a mapping q. For any $k = 0, \cdots, K$, let $\mathcal{D}_k = \{\zeta_{i,k,l}, i \in \mathcal{V}, 1 \leq l \leq s\}$, $\mathcal{D}'_k = \{\zeta'_{i,k,l}, i \in \mathcal{V}, 1 \leq l \leq s\}$ be the data samples taken from $\mathcal{D}, \mathcal{D}'$ at the k-th iteration, respectively. Define the sensitivity of Algorithm 1 at the k-th iteration as follows:*

$$\Delta_k^q \triangleq \sup_{Adj(\mathcal{D},\mathcal{D}')} \|q(\mathcal{D}_k) - q(\mathcal{D}'_k)\|. \tag{7}$$

Remark 4 *Definition 5 captures the magnitude by which one node's data sample can change the mapping q in the worst case. It is the key quantity to achieve the (ϵ, δ)-differential privacy at the k-th iteration. In Algorithm 1, the mapping $q(\mathcal{D}_k) = x_{k+1} = [x_{1,k+1}^\top, \ldots, x_{n,k+1}^\top]^\top$.*

The following lemma gives the sensitivity Δ_k^q of Algorithm 1 for any $k = 0, \ldots, K$.

Lemma 1 *At the k-th iteration, the sensitivity of Algorithm 1 satisfies $\Delta_k^q \leq \frac{\alpha C}{s}\left(\sum_{m=0}^{k} |1-\beta|^m\right)$.*

Proof: Since the differential privacy is well-defined for the worse case where the event triggering instant happens at each iteration $k = 0, \ldots, K$, the whole observations $(\mathcal{C}(x_0), \mathcal{C}(x_1), \ldots, \mathcal{C}(x_K))$ is considered. Then, the left proof is similar to that of Lemma 1 in Chen et al. (2024), and thus, is omitted here. □

Next, we consider the following two masked cases. The first one is the additive Gaussian noise, i.e., $\mathcal{C}(x_{i,k}) = x_{i,k} + d_{i,k}$ with $d_{i,k} \sim N(0, \sigma^2 I_r)$. The second one is the unbiased stochastic quantizer, the unbiased stochastic quantizes a vector $x_{i,k} = [x_{i1,k}, \ldots, x_{ir,k}]^\top \in \mathbb{R}^r$ as $\mathcal{C}(x_{i,k}) = [\mathcal{C}(x_{i1,k}), \ldots, \mathcal{C}(x_{ir,k})]^\top$ to the range by a scale factor $\sigma \in \mathbb{N}^+$. For any $l_c\sigma < x_{i\iota,k} \leq (l_c+1)\sigma$, $\iota = 1, \ldots, r$, $l_c \in \mathbb{Z}$, the quantizer outputs

$$\mathcal{C}(x_{i\iota,k}) = \begin{cases} l_c\sigma, & \text{with probability} 1 + l_c - x_{i\iota,k}/\sigma; \\ (l_c+1)\sigma, & \text{with probability } x_{i\iota,k}/\sigma - l_c. \end{cases} \tag{8}$$

We first give the privacy analysis for the additive Gaussian noise case.

Theorem 1 *For any given $K \geq 1, k = 0, \cdots, K$, let*

$$\alpha = \frac{a_1}{K^{p_1}},\ \beta = \frac{a_2}{K^{p_2}},\ s = \lfloor a_3 K^{p_3} \rfloor + 1,$$
$$\sigma = K^{p_4}, \delta_k = \frac{1}{(k+2)^\nu},\ a_1,\ a_2,\ a_3 > 0.$$

If $0 < a_2 < K^{p_2}$ and $\nu > 0$, then Algorithm 1 achieves

(ϵ, δ)-differential privacy over finite iterations K, where

$$\epsilon = \sum_{k=0}^{K} \epsilon_k \leq \sum_{k=0}^{K} \frac{2Ca_1\sqrt{\ln(1.25(k+1)^\nu)}}{a_2 a_3 K^{p_1 - p_2 + p_3 + p_4}},$$
$$\delta = \sum_{k=0}^{K} \frac{1}{(k+2)^\nu}. \tag{9}$$

Furthermore, if $p_1 - p_2 + p_3 + p_4 > 1$, $\nu \geq 2$, then Algorithm 1 achieves finite cumulative differential privacy budgets ϵ, δ over infinite iterations.

Proof. The proof is the same as that of Theorem 1 in Chen et al. (2024), and thus, is omitted here. □

Now we give the privacy analysis for the unbiased stochastic quantizer case.

Theorem 2 *For any given $K \geq 1, k = 0, \cdots, K$, let*

$$\alpha = \frac{a_1}{K^{p_1}}, \ \beta = \frac{a_2}{K^{p_2}}, \ s = \lfloor a_3 K^{p_3} \rfloor + 1,$$
$$\sigma = K^{p_4}, \ a_1, \ a_2, \ a_3 > 0.$$

If $0 < a_2 < K^{p_2}$, then Algorithm 1 achieves $(0, \delta)$-differential privacy over finite iterations K, where

$$\delta \leq \min\{1, \frac{Ca_1(K+1)}{a_2 a_3 K^{p_1 - p_2 + p_3 + p_4}}\}. \tag{10}$$

Furthermore, if $p_1 - p_2 + p_3 + p_4 > 1$, then the cumulative differential privacy budget δ of Algorithm 1 goes to 0 over infinite iterations.

Proof. Without loss of generality, we proceed with the following two cases.

Case 1: If $x_{i\iota,k} \in ((l_c - 1)\sigma, l_c\sigma]$ and $x'_{i\iota,k} \in (l_c\sigma, (l_c + 1)\sigma]$, then

$$\begin{aligned}\delta_k =& |\mathbb{P}[\mathcal{C}(x_{i\iota,k}) = l_c\sigma | x_{i\iota,k}] - \mathbb{P}[\mathcal{C}(x'_{i\iota,k}) = l_c\sigma | x'_{i\iota,k}]| \\ =& |x_{i\iota,k}/\sigma - (l_c - 1) - (1 + l_c - x'_{i\iota,k}/\sigma)| \\ =& |(x_{i\iota,k} + x'_{i\iota,k})/\sigma - 2l_c| \\ \leq& |x'_{i\iota,k} - l_c\sigma|/\sigma + |l_c\sigma - x_{i\iota,k}|/\sigma \\ =& \frac{|x_{i\iota,k} - x'_{i\iota,k}|}{\sigma}.\end{aligned}$$

Similarly, one can obtain the same relationship when $x_{i\iota,k} \in (l_c\sigma, (l_c + 1)\sigma]$ and $x'_{i\iota,k} \in ((l_c - 1)\sigma, l_c\sigma]$.

Case 2: If $x_{i\iota,k}, x'_{i\iota,k} \in ((l_c - 1)\sigma, l_c\sigma]$, then

$$\begin{aligned}\delta_k =& |\mathbb{P}[\mathcal{C}(x_{i\iota,k}) = l_c\sigma | x_{i\iota,k}] - \mathbb{P}[\mathcal{C}(x'_{i\iota,k}) = l_c\sigma | x'_{i\iota,k}]| \\ =& |1 + l_c - x_{i\iota,k}/\sigma - (1 + l_c - x'_{i\iota,k}/\sigma)| \\ =& \frac{|x_{i\iota,k} - x'_{i\iota,k}|}{\sigma}\end{aligned}$$

Similarly, one can obtain the same relationship when $x_{i\iota,k}, x'_{i\iota,k} \in (l_c\sigma, (l_c + 1)\sigma]$. Hence, we have

$$\delta_k \leq \frac{\frac{\alpha C}{s}\left(\sum_{m=0}^{k} |1 - \beta|^m\right)}{\sigma} \leq \frac{Ca_1}{a_2 a_3 K^{p_1 - p_2 + p_3 + p_4}}.$$

Note that $\delta \leq 1$. Then, the theorem is proved. □

Remark 5 *In the unbiased stochastic quantizer case, the unbiased stochastic quantizer plays two roles, one is saving the bits of communication, another is achieving the privacy preserving. While* Singh et al. (2023) *considers both event-triggered and quantized communication in distributed optimization, it does not consider the privacy protection. Further, by properly designing the scaling factors σ and the sample size s, we solve the issue of privacy protection failure caused by δ_k increasing to 1 in* Wang & Başar (2023b); Liu et al. (2025). *Therefore, compared with* Singh et al. (2023); Wang & Başar (2023b); Liu et al. (2025), *Algorithm 1 achieves a smaller δ over infinite iterations while saving the transmitted rounds and bits of communication simultaneously.*

Remark 6 *Theorems 1-2 also shows how step-size parameters p_1, p_2, the sample-size parameter p_3 and the masked parameter p_4 affect cumulative differential privacy budgets. As shown in (9), the larger the step-size parameter p_1, the sample-size parameter p_3 and the masked parameter p_4 are, the smaller cumulative differential privacy budgets are. In addition, the smaller the step-size parameter p_2 is, the smaller cumulative differential privacy budgets are.*

Remark 7 *The sample-size s is not required to go to infinity to achieve differential privacy over infinite iterations for both masked cases. Specifically, let the sample-size parameter $p_3 = 0$. Then, the sample-size s is constant. For the additive Gaussian noise case, if $p_1 - p_2 + p_4 > 1, \nu \geq 2$, then Algorithm 1 can achieve finite cumulative differential privacy budgets over infinite iterations. This shows advantage over* Zhang et al. (2018); Li et al. (2018); Huang et al. (2019); Ding et al. (2021); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Yan et al. (2023); Liu et al. (2024), *since cumulative differential privacy budgets go to infinity therein. For the unbiased stochastic quantizer case, if $p_1 - p_2 + p_4 > 1$, then the cumulative differential privacy budget δ of Algorithm 1 goes to 0 over infinite iterations. This shows advantage over* Wang & Başar (2023b); Liu et al. (2025), *since (0, 1)-differential privacy is achieved over infinite iterations.*

3.3 Convergence analysis

In this subsection, we will give the convergence rate analysis of Algorithm 1. As shown in Reisizadeh et al. (2019b), by (8), one obtains $\mathbb{E}[\mathcal{C}(x_{i,k}) - x_{i,k} | x_{i,k}] = 0$, and $\mathbb{E}[|\mathcal{C}(x_{i.k})-x_{i,k}|^2|x_{i,k}] \leq \sigma^2$. Thus, for both masked cases, we treat $\mathcal{C}(x_{i,k}) - x_{i,k}$ as a stochastic noise $d_{i,k}$ for convergence analysis. Define σ algebras $\mathcal{F}_0=\sigma(\{\varnothing\})$ and $\mathcal{F}_k=\sigma(\{d_l, w_l : l = 0, \dots, , k-1\})$, $\forall k = 1, \dots, K$. Then, $\mathcal{C}(x_{i,k}) = x_{i,k} + d_{i,k}$ with $\mathbb{E}[d_{i,k}|\mathcal{F}_k] = 0$ and $\mathbb{E}[\|d_{i,k}\|^2|\mathcal{F}_k] \leq \sigma^2$. Let

$$x_k = [x_{1,k}^\top, \dots, x_{n,k}^\top]^\top,\ e_k = [e_{1,k}^\top, \dots, e_{n,k}^\top]^\top,$$
$$d_k = [d_{1,k}^\top, \dots, d_{n,k}^\top]^\top,\ \nabla \ell_k = [\nabla \ell_{1,k}^\top, \dots, \nabla \ell_{n,k}^\top]^\top,$$
$$\nabla f(x_k) \triangleq [\nabla f_1(x_{1,k})^\top, \nabla f_2(x_{2,k})^\top, \dots, \nabla f_n(x_{n,k})^\top]^\top,$$
$$w_k \triangleq \nabla \ell_k - \nabla f(x_k).$$

Then, by $\mathcal{C}(x_{\tau_k}) = x_k + d_k - e_k$, we express (6) in a compact form as follows:

$$\begin{aligned} x_{k+1} =& ((I_n - \beta\mathcal{L}) \otimes I_r)x_k - \alpha\nabla f(x_k) \\ &+ \beta(\mathcal{A} \otimes I_r)(d_k - e_k) - \alpha w_k. \end{aligned} \tag{11}$$

First, we introduce an assumption on step-sizes, the sample-size, the privacy noise, and the event-triggered threshold.

Assumption 3 *For any given $K \geq 1$, step-sizes $\alpha = \frac{a_1}{K^{p_1}}$, $\beta = \frac{a_2}{K^{p_2}}$, the sample-size $s = \lfloor a_3 K^{p_3} \rfloor + 1$, the privacy noise parameter $\sigma = K^{p_4}$, and the event-triggered threshold $\Phi = \frac{a_4}{K^{p_5}}$ satisfy $a_1, a_2, a_3, a_4 > 0$, $p_2+p_5 > p_1$, $\frac{1}{2} < p_2 < p_1 < 1$, $2p_2 - 2p_4 > p_1$.*

Assumption 4 *(Polyak-Lojasiewicz) The global cost function $F(x)$ satisfies the Polyak-Lojasiewicz condition, i.e., there exists $\mu > 0$ such that $2\mu(F(x) - F^*) \leq \|\nabla F(x)\|^2$, $\forall x \in \mathbb{R}^r$.*

Remark 8 *Assumption 4 is commonly used (e.g. (Xin et al., 2022)), and means that the gradient $\nabla F(x)$ grows faster than a quadratic function as the algorithm moves away from the optimal solution. Such functions exist, for example, $F(x) = x^2 + 3\sin^2 x$ is a nonconvex function satisfying Assumption 4 for any $\mu \in (0, 0.3)$. As shown in Theorem 2 of Karimi et al. (2016), Assumption 4 is more general than the convex cost functions assumed in Zhang et al. (2018); Li et al. (2018); Huang et al. (2019); Reisizadeh et al. (2019a); Ding et al. (2021); Gratton et al. (2021); Liu et al. (2024).*

Theorem 3 *If Assumptions 1-4 hold, then for any given $K \geq 1$ and $\psi \in [1, 2]$, we have $\mathbb{E}\|\nabla F(x_{i,K+1})\|^\psi = O(K^{-\frac{\psi}{2}\min\{p_1-p_2, p_2+p_5-p_1, 2p_2-2p_4-p_1\}})$, $\forall i \in \mathcal{V}$. Furthermore, when $\psi = 2$,*

$$\begin{aligned} &\mathbb{E}(F(x_{i,K+1}) - F^*) \\ =& O(K^{-\min\{p_1-p_2, p_2+p_5-p_1, 2p_2-2p_4-p_1\}}), \forall i \in \mathcal{V}, \end{aligned} \tag{12}$$

and the mean square convergence is achieved as K goes to infinity, i.e., $\lim_{K\to\infty} \mathbb{E}\|\nabla F(x_{i,K+1})\|^2 = 0, \forall i \in \mathcal{V}$.

Proof. See Appendix A. □

Remark 9 *When the event-triggered mechanism (4) is employed, the introduced error e_k brings difficulty to the convergence rate analysis of Algorithm 1. To combat this effect, the step-size β is introduced. Moreover, from (12) it follows that the smaller the event-triggered threshold parameter p_5 is, the slower the convergence rate is. Therefore, the introduced event-triggered mechanism does slow down the convergence rate of Algorithm 1.*

Remark 10 *The convergence rate of Algorithm 1 is given for general privacy noises, including increasing, constant and decreasing privacy noises. This is nontrivial even without considering the privacy protection.*

Remark 11 *If the global cost function $F(x)$ is λ-strongly convex, i.e., there exists $\lambda > 0$ such that $F(y) \geq F(x) + \langle \nabla F(x), y - x\rangle + \frac{\lambda}{2}\|y - x\|^2$, $\forall x, y \in \mathbb{R}^r$, then by Lemma 6.9 in Bubeck (2015) we have $2\lambda(F(x) - F^*) \leq \|\nabla F(x)\|^2$, which means the global cost function $F(x)$ satisfies Assumption 4. In this case, Algorithm 1 achieves the same convergence rate as Theorem 3. Thus, Theorem 3 also holds for λ-strongly convex cost function.*

Remark 12 *Note that distributed nonconvex stochastic optimization algorithms may converge to a saddle point instead of the desired global minimum. Then, the discussion of the avoidance of saddle points is necessary. Assumption 4 implies that each stationary point x^* of F satisfying $\nabla F^* = 0$ is a global minimum of F, and thus guarantees the avoidance of saddle points discussed in Wang & Başar (2023a). Furthermore, compared with Zhang et al. (2018); Li et al. (2018); Huang et al. (2019); Ding et al. (2021); Gratton et al. (2021); Wang & Başar (2023a); Liu et al. (2024), Assumption 4 helps us to give the convergence rate of Algorithm 1 without the assumption of bounded gradients.*

Based on Theorems 1-3, the mean square convergence of Algorithm 1 as well as the differential privacy over infinite iterations can be established, which is given in the following corollary:

Corollary 1 *For any given $K \geq 1$, $k = 0, \cdots, K$, let*

$$\begin{aligned} &\alpha = \frac{a_1}{K^{p_1}},\ \beta = \frac{a_2}{K^{p_2}},\ s = \lfloor a_3 K^{p_3} \rfloor + 1, \sigma = K^{p_4}, \\ &\Phi = \frac{a_4}{K^{p_5}}, \delta_k = \frac{1}{(k+2)^\nu},\ a_1,\ a_2,\ a_3,\ a_4 > 0. \end{aligned}$$

Then, under Assumptions 1-2, and 4, we have the following conclusions for the two masked methods.

- *For the additive Gaussian noise case, if $\nu \geq 2$, $p_2 + p_5 > p_1$, $\frac{1}{2} < p_2 < p_1 < 1$, $2p_2 - 2p_4 > p_1$, $p_1 - p_2 + p_3 + p_4 > 1$, then Algorithm 1 achieves the mean square* *convergence and finite cumulative differential privacy budgets ϵ, δ over infinite iterations simultaneously as the sample-size s goes to infinity.*

- *For the unbiased stochastic quantizer case, if $p_2+p_5 > p_1$, $\frac{1}{2} < p_2 < p_1 < 1$, $2p_2-2p_4 > p_1$, $p_1-p_2+p_3+p_4 > 1$, then Algorithm 1 achieves the mean square convergence and the cumulative differential privacy budget δ going to 0 over infinite iterations simultaneously as the sample-size s goes to infinity.*

Proof. By Theorems 1-3, this corollary is proved. □

Remark 13 *The result of Corollary 1 does not contradict the trade-off between privacy and utility. In fact, to achieve differential privacy, Algorithm 1 incurs a compromise on the utility. However, different from Gratton et al. (2021); Yan et al. (2023) which compromise convergence accuracy to enable differential privacy, Algorithm 1 compromises the convergence rate and the sample-size (which are also utility metrics) instead. From Corollary 1, it follows that the larger the privacy noise parameter p_4 is, the slower the mean square convergence rate is. Besides, the sample-size s is required to go to infinity when the mean square convergence of Algorithm 1 and finite cumulative privacy budgets over infinite iterations are considered simultaneously. The ability to retain convergence accuracy makes our approach suitable for accuracy-critical scenarios.*

Based on Theorem 3, Definitions 3 and 4, the oracle complexity of Algorithm 1 for obtaining an η-optimal solution is given as follows:

Theorem 4 *Given $\eta \in (0, \frac{1}{2})$, let $p_1 = 1-\eta$, $p_2 = \frac{2-2\eta}{3}$, $p_3 = \eta$, $p_4 = 0$, $p_5 = 1$. Then, under Assumptions 1-2 and 4, the oracle complexity of Algorithm 1 is $O(\eta^{-\frac{3+3\eta}{1-\eta}})$.*

Proof. For given $\eta > 0$, let the iteration maximum in Algorithm 1 be $N(\eta)$. Then, we have $s = \lfloor a_3 N(\eta)^\eta \rfloor + 1 \leq a_3 N(\eta)^\eta + 1$.

Note that by Theorem 3, there exists a constant $C > 0$ such that

$$\mathbb{E}|F(x_{i,K+1}) - F^*| = \mathbb{E}(F(x_{i,K+1}) - F^*) \leq \frac{C}{K^{\frac{1-\eta}{3}}}. \quad (13)$$

Then, when $K \geq \lfloor (\frac{C}{\eta})^{\frac{3}{1-\eta}} \rfloor + 1 > (\frac{C}{\eta})^{\frac{3}{1-\eta}}$, (13) can be rewritten as

$$\mathbb{E}|F(x_{i,K+1}) - F^*| \leq \frac{C}{K^{\frac{1-\eta}{3}}} < \frac{C}{(\frac{C}{\eta})^{(\frac{1-\eta}{3})\frac{3}{1-\eta}}} = \eta. \quad (14)$$

Thus, by (14) and Definition 3, x_{K+1} is an η-optimal solution. Since $N(\eta)$ is the smallest integer such that $x_{N(\eta)}$ is an η-optimal solution, we have

$$\begin{aligned} N(\eta) &\leq 1 + \min\{K : K \geq \lfloor (\frac{C}{\eta})^{\frac{3}{1-\eta}} \rfloor + 1\} \\ &= \lfloor (\frac{C}{\eta})^{\frac{3}{1-\eta}} \rfloor + 2. \end{aligned} \quad (15)$$

Hence, by Definition 4 and (15), we have

$$\begin{aligned} \sum_{k=0}^{N(\eta)} s &= (N(\eta)+1)s \leq (N(\eta)+1)(a_3 N(\eta)^\eta + 1) \\ &= O\left(N(\eta)^{1+\eta}\right) = O\left(\eta^{-\frac{3+3\eta}{1-\eta}}\right). \end{aligned}$$

Therefore, the theorem is proved. □

Remark 14 *From Theorems 3 and 4, the faster the convergence rate is, the smaller the oracle complexity is. Further, if $\eta = 0.02$, then the total number of data samples to obtain an η-optimal solution is $O(10^5)$, which does not go to infinity. This requirement for the total number of data samples is acceptable since the computational cost of centralized SGD is $O(10^5)$ to achieve the same accuracy as Algorithm 1.*

4 Numerical Example

In this section, we train the convolution neural network (CNN) model in the distributed manner on the benchmark dataset "MNIST" (LeCun et al., 1998). Specifically, five nodes cooperatively train the CNN model over the undirected graph shown in Fig. 2, which satisfies Assumption 1. Then, the "MNIST" dataset is divided into two subsets for training and testing, respectively. The training dataset is uniformly divided into 5 subsets, each of which can only be accessed by one node to update its model parameters. The testing dataset can be accessed by all nodes to evaluate the performance of their models. In the following, we show the effect of the event-triggered threshold parameter p_5 on the convergence rate of Algorithm 1, and the comparison of the convergence rate between Algorithm 1 and methods in Li et al. (2018); Huang et al. (2019); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Liu et al. (2024), respectively.

4.1 Effect of the event-triggered threshold parameter on the convergence rate

Let step-sizes $\alpha = \frac{80}{2000} = 6 \cdot 10^{-2}$, $\beta = \frac{0.7}{2000^{0.65}} = 5 \cdot 10^{-3}$, the sample-size $s = \lfloor 3 \cdot 10^{-4} \cdot 2000^{1.6} \rfloor + 1 = 58$, the noise parameter $p_4 = -1$, and the event-triggered threshold $\Phi = \frac{130}{2000^{p_5}}$ with the event-triggered threshold parameter $p_5 = 0.5$, 1, 2, respectively. Then, the training and testing accuracy on the "MNIST" dataset are presented

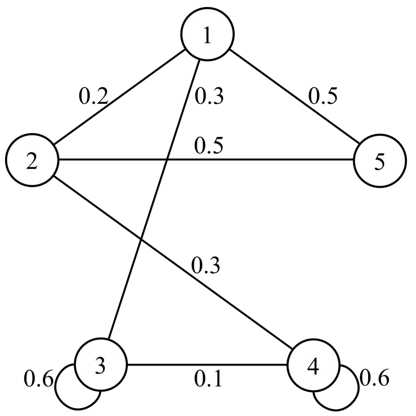
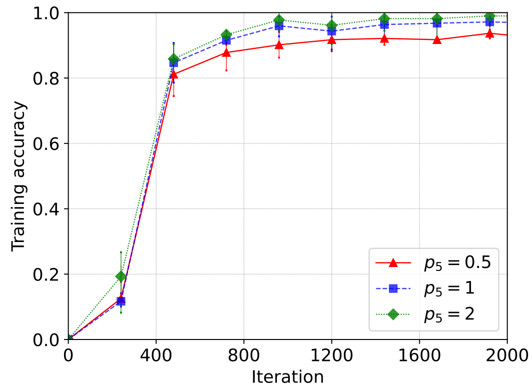
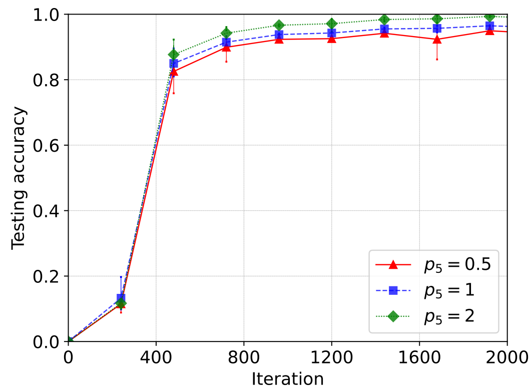


Fig. 2. Topology structure of the undirected graph

in Fig. 3, from which one can see that the larger the event-triggered threshold parameter p_5 is, the faster Algorithm 1 converges.



(a) Training accuracy

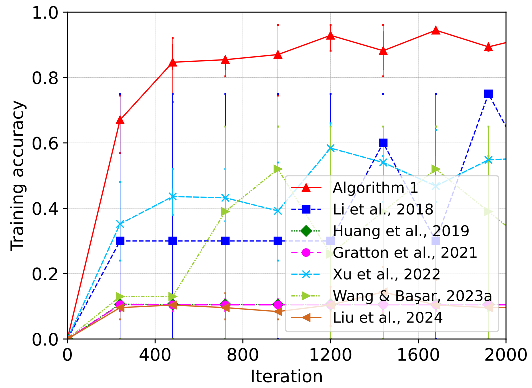


(b) Testing accuracy

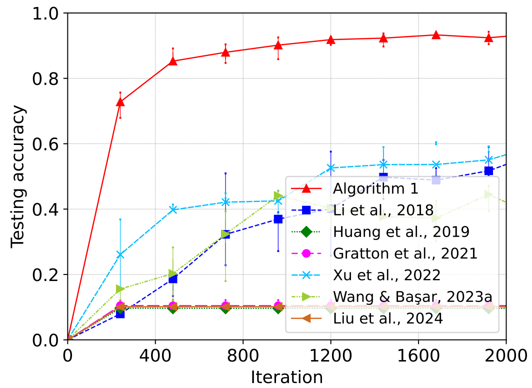
Fig. 3. Accuracy of Algorithm 1 with the event-triggered threshold parameter $p_5 = 0.5, 1, 2$

4.2 Comparison with existing methods

Let the event-triggered threshold parameter $p_5 = 0.5$ in Algorithm 1. Then, the comparison of the convergence rate between Algorithm 1 and methods in Li et al. (2018); Huang et al. (2019); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Liu et al. (2024) is presented in Fig. 4. To ensure a fair comparison, we set the same step-sizes and sample-size in Li et al. (2018); Huang et al. (2019); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Liu et al. (2024) as the ones of this paper. From Figs. 4(a) and 4(b), it can be seen that Algorithm 1 converges faster than those in Li et al. (2018); Huang et al. (2019); Gratton et al. (2021); Xu et al. (2022); Wang & Başar (2023a); Liu et al. (2024).



(a) Training accuracy



(b) Testing accuracy

Fig. 4. Comparison of the convergence rate in Algorithm 1 with the ones in (Li et al., 2018; Huang et al., 2019; Gratton et al., 2021; Xu et al., 2022; Wang & Başar, 2023a; Liu et al., 2024)

5 Conclusion

In this paper, we have proposed a differentially private distributed nonconvex stochastic optimization algorithm with event-triggered communication. Two masked cases are considered to achieve the differential privacy, one is the additive Gaussian noise, and another is the unbiased stochastic quantizer. For both cases, differential privacy analysis is given rigorously. By using the sample-size parameter-controlled subsampling method, the differential privacy level of the algorithm is enhanced compared with the existing ones. Then, by using the two-time-scale step-sizes method, the convergence rate and the oracle complexity of the algorithm are given under the Polyak-Łojasiewicz condition. Further, we show how the event-triggered mechanism affects the convergence rate. Finally, a numerical example of the distributed training of CNN on the "MNIST" dataset is given to verify the effectiveness of the algorithm.

Appendix A. Proof of Theorem 3

For the convenience of the analysis, let

$$\begin{aligned}
&\bar{x}_k \triangleq \frac{1}{n}(\mathbf{1}_n^\top \otimes I_r)x_k, \bar{w}_k \triangleq \frac{1}{n}(\mathbf{1}_n^\top \otimes I_r)w_k,\\
&Y_k \triangleq (W \otimes I_r)x_k, W \triangleq I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top,\\
&\nabla f(\bar{x}_k) \triangleq [\nabla f_1(\bar{x}_k)^\top, \nabla f_2(\bar{x}_k)^\top, \dots, \nabla f_n(\bar{x}_k)^\top]^\top,\\
&\overline{\nabla f(x_k)} \triangleq \frac{1}{n}(\mathbf{1}_n^\top \otimes I_r)\nabla f(x_k) = \frac{1}{n}\sum_{i=1}^n \nabla f_i(x_{i,k}).
\end{aligned}$$

The following three steps are given to prove Theorem 3.

Step 1: We first consider the term $\|Y_k\|^2$. Note that $W(I_n - \beta\mathcal{L}) = (I_n - \beta\mathcal{L})W$. Then, multiplying both sides of (11) by $W \otimes I_r$ gives

$$\begin{aligned}
Y_{k+1} = &((I_n - \beta\mathcal{L}) \otimes I_r)Y_k - \alpha(W \otimes I_r)\nabla f(x_k)\\
&+ \beta(W\mathcal{A} \otimes I_r)(d_k - e_k) - \alpha(W \otimes I_r)w_k. \qquad (A.1)
\end{aligned}$$

Since d_k is independent of $\mathcal{F}_k$, $\mathbb{E}[d_{i,k}|\mathcal{F}_k] = 0$ and $\mathbb{E}[\|d_{i,k}\|^2|\mathcal{F}_k] \le \sigma^2$, we then have

$$\mathbb{E}[d_k|\mathcal{F}_k] = 0, \qquad (A.2)$$
$$\mathbb{E}[\|d_k\|^2|\mathcal{F}_k] = nr\sigma^2. \qquad (A.3)$$

Since $w_k = \nabla \ell_k - \nabla f(x_k)$, by Assumption 2(iii) we have

$$\mathbb{E}(w_k|\mathcal{F}_k) = \mathbb{E}w_k = 0, \qquad (A.4)$$
$$\mathbb{E}(\|w_k\|^2|\mathcal{F}_k) = \mathbb{E}\|w_k\|^2 \le \frac{n\sigma_\ell^2}{s}. \qquad (A.5)$$

By the event-trigger mechanism (4), we have

$$\|e_k\| \le \sqrt{n}\Phi. \qquad (A.6)$$

By (A.4), taking conditional mathematical expectation of $\|Y_{k+1}\|^2$ with respect to $\mathcal{F}_k$, it is obtained that

$$\begin{aligned}
&\mathbb{E}(\|Y_{k+1}\|^2|\mathcal{F}_k)\\
=&\mathbb{E}(\| ((I_n - \beta\mathcal{L}) \otimes I_r) Y_k - \alpha(W \otimes I_r)\nabla f(x_k)\\
&+\beta(W\mathcal{A} \otimes I_r)(d_k - e_k) - \alpha(W \otimes I_r)w_k\|^2|\mathcal{F}_k)\\
=&\mathbb{E}(\| ((I_n - \beta\mathcal{L}) \otimes I_r) Y_k - \alpha(W \otimes I_r)\nabla f(x_k)\\
&+\beta(W\mathcal{A} \otimes I_r)(d_k - e_k)\|^2|\mathcal{F}_k)+\alpha^2\mathbb{E}(\|(W\otimes I_r)w_k\|^2|\mathcal{F}_k)\\
&- 2\mathbb{E}(\langle((I_n-\beta\mathcal{L}) \otimes I_r)Y_k - \alpha(W\otimes I_r)\nabla f(x_k)\\
&+\beta(W\mathcal{A} \otimes I_r)(d_k-e_k), \alpha(W \otimes I_r)w_k\rangle|\mathcal{F}_k)\\
=&\mathbb{E}(\|((I_n-\beta\mathcal{L})\otimes I_r)Y_k - \alpha(W\otimes I_r)\nabla f(x_k)\\
&+\beta(W\mathcal{A} \otimes I_r)(d_k-e_k)\|^2|\mathcal{F}_k)+\alpha^2\mathbb{E}\,\|(W\otimes I_r)w_k\|^2. \qquad (A.7)
\end{aligned}$$

By the law of total expectation, taking mathematical expectation on both sides of (A.7) we have

$$\begin{aligned}
&\mathbb{E}\|Y_{k+1}\|^2\\
=&\mathbb{E}\|((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k)\\
&+ \beta(W\mathcal{A}\otimes I_r)(d_k-e_k)\|^2+\alpha^2\mathbb{E}\|(W\otimes I_r)w_k\|^2\\
=&\mathbb{E}\|((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k) - \beta(W\mathcal{A}\otimes I_r)e_k\|^2\\
&+ \beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)d_k\|^2 + 2\mathbb{E}\langle((I_n-\beta\mathcal{L})\otimes I_r)Y_k\\
&- \alpha(W\otimes I_r)\nabla f(x_k), \beta(W\mathcal{A}\otimes I_r)d_k\rangle - 2\mathbb{E}\langle\beta(W\mathcal{A}\otimes I_r)e_k,\\
&\beta(W\mathcal{A}\otimes I_r)d_k\rangle +\alpha^2\mathbb{E}\|(W\otimes I_r)w_k\|^2. \qquad (A.8)
\end{aligned}$$

Since x_k is $\mathcal{F}_k$-measurable, $((I_n - \beta\mathcal{L})\otimes I_r)Y_k - \alpha(W \otimes I_r)\nabla f(x_k)$ is $\mathcal{F}_k$-measurable. By (A.2), we have

$$\begin{aligned}
&2\mathbb{E}\langle((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k), \beta(W\mathcal{A}\otimes I_r)d_k\rangle\\
=&2\mathbb{E}(\mathbb{E}(\langle((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k),\\
&\beta(W\mathcal{A}\otimes I_r)d_k\rangle|\mathcal{F}_k))\\
=&2\mathbb{E}(\langle((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k),\\
&\mathbb{E}(\beta(W\mathcal{A}\otimes I_r)d_k|\mathcal{F}_k)\rangle)\\
=&0. \qquad (A.9)
\end{aligned}$$

Note that for any $m \ge 1$ and $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^r$, the following inequality holds:

$$\begin{aligned}
&\|\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_m\|^2\\
\le& m(\|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2 + \cdots + \|\mathbf{a}_m\|^2). \qquad (A.10)
\end{aligned}$$

Then, by letting $m = 2$ in (A.10), we have

$$\begin{aligned}
&- 2\mathbb{E}\langle\beta(W\mathcal{A}\otimes I_r)e_k, \beta(W\mathcal{A}\otimes I_r)d_k\rangle\\
\le& 2\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|\|(W\mathcal{A}\otimes I_r)d_k\|\\
\le& \beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2 + \beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)d_k\|^2. \qquad (A.11)
\end{aligned}$$

Substituting (A.9) and (A.11) into (A.8) implies

$$\begin{aligned}
&\mathbb{E}\|Y_{k+1}\|^2\\
\le&\mathbb{E}\|((I_n-\beta\mathcal{L})\otimes I_r)Y_k-\alpha(W\otimes I_r)\nabla f(x_k)\\
&- \beta(W\mathcal{A}\otimes I_r)e_k\|^2 + 2\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)d_k\|^2\\
&+ \beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2 +\alpha^2\mathbb{E}\|(W\otimes I_r)w_k\|^2. \qquad (A.12)
\end{aligned}$$

Note that the following Cauchy-Schwarz inequality (Zorich, 2015) holds: $\|\mathbf{a}+\mathbf{b}\|^2 \le (1+\rho_{\mathcal{L}}\beta)\|\mathbf{a}\|^2 + (1+\frac{1}{\rho_{\mathcal{L}}\beta})\|\mathbf{b}\|^2$, $\forall \mathbf{a},\mathbf{b} \in \mathbb{R}^r$, where $\rho_{\mathcal{L}} > 0$ is the second smallest eigenvalue of $\mathcal{L}$. Then, this together with (A.12) gives

$$\begin{aligned}
&\mathbb{E}\|Y_{k+1}\|^2\\
\le&(1+\rho_{\mathcal{L}}\beta)\,\mathbb{E}\,\|((I_n-\beta\mathcal{L})\otimes I_r)\,Y_k\|^2\\
&+\left(1+\frac{1}{\rho_{\mathcal{L}}\beta}\right)\mathbb{E}\|-\alpha(W\otimes I_r)\nabla f(x_k)-\beta(W\mathcal{A}\otimes I_r)e_k\|^2\\
&+2\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)d_k\|^2+\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2\\
&+\alpha^2\mathbb{E}\,\|(W\otimes I_r)w_k\|^2\,. \qquad (A.13)
\end{aligned}$$

By letting $m=2$ in (A.10), $\mathbb{E}\|-\alpha(W\otimes I_r)\nabla f(x_k)-\beta(W\mathcal{A}\otimes I_r)e_k\|^2$ in (A.13) can be rewritten as

$$\begin{aligned}
&\mathbb{E}\|-\alpha(W\otimes I_r)\nabla f(x_k)-\beta(W\mathcal{A}\otimes I_r)e_k\|^2\\
\le&2\alpha^2\mathbb{E}\|(W\otimes I_r)\nabla f(x_k)\|^2\\
&+2\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2. \qquad (A.14)
\end{aligned}$$

Substituting (A.14) into (A.13) implies

$$\begin{aligned}
&\mathbb{E}\,\|Y_{k+1}\|^2\\
\le&(1+\rho_{\mathcal{L}}\beta)\,\mathbb{E}\,\|((I_n-\beta\mathcal{L})\otimes I_r)\,Y_k\|^2\\
&+\frac{2(1+\rho_{\mathcal{L}}\beta)}{\rho_{\mathcal{L}}\beta}(\alpha^2\mathbb{E}\|(W\otimes I_r)\nabla f(x_k)\|^2\\
&+\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2)\\
&+2\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)d_k\|^2+\beta^2\mathbb{E}\|(W\mathcal{A}\otimes I_r)e_k\|^2\\
&+\alpha^2\mathbb{E}\,\|(W\otimes I_r)w_k\|^2\,. \qquad (A.15)
\end{aligned}$$

Note that $\|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\|$, $\forall A \in \mathbb{R}^{n\times n}$, $\mathbf{x}\in\mathbb{R}^n$. Then, by $\|\mathcal{A}\|=\|W\|=1$, substituting (A.3), (A.5) and (A.6) into (A.15) implies

$$\begin{aligned}
&\mathbb{E}\,\|Y_{k+1}\|^2\\
\le&(1+\rho_{\mathcal{L}}\beta)\,\mathbb{E}\,\|((I_n-\beta\mathcal{L})\otimes I_r)Y_k\|^2\\
&+\frac{2(1+\rho_{\mathcal{L}}\beta)\alpha^2}{\rho_{\mathcal{L}}\beta}\mathbb{E}\|\nabla f(x_k)\|^2\\
&+\frac{n(2+3\rho_{\mathcal{L}}\beta)\beta\Phi^2}{\rho_{\mathcal{L}}}+2nr\beta^2\sigma^2+\frac{n\alpha^2\sigma_\ell^2}{s}. \qquad (A.16)
\end{aligned}$$

Then, by Courant-Fischer's Theorem (Horn & Johnson, 2012) we have

$$\|((I_n-\beta\mathcal{L})\otimes I_r)\,Y_k\|^2 \le (1-\rho_{\mathcal{L}}\beta)^2\|Y_k\|^2. \qquad (A.17)$$

Thus, substituting (A.17) into (A.16), one can get

$$\begin{aligned}
&\mathbb{E}\|Y_{k+1}\|^2\\
\le&(1+\rho_{\mathcal{L}}\beta)(1-\rho_{\mathcal{L}}\beta)^2\mathbb{E}\|Y_k\|^2+\frac{2(1+\rho_{\mathcal{L}}\beta)\alpha^2}{\rho_{\mathcal{L}}\beta}\mathbb{E}\|\nabla f(x_k)\|^2\\
&+\frac{n(2+3\rho_{\mathcal{L}}\beta)\beta\Phi^2}{\rho_{\mathcal{L}}}+2nr\beta^2\sigma^2+\frac{n\alpha^2\sigma_\ell^2}{s}\\
\le&(1-\rho_{\mathcal{L}}\beta)\mathbb{E}\|Y_k\|^2\\
&+\frac{2(1+\rho_{\mathcal{L}}\beta)\alpha^2}{\rho_{\mathcal{L}}\beta}\mathbb{E}\|\nabla f(x_k)-\nabla f(\bar{x}_k)+\nabla f(\bar{x}_k)\|^2\\
&+\frac{n(2+3\rho_{\mathcal{L}}\beta)\beta\Phi^2}{\rho_{\mathcal{L}}}+2nr\beta^2\sigma^2+\frac{n\alpha^2\sigma_\ell^2}{s}. \qquad (A.18)
\end{aligned}$$

Letting $m=2$ in (A.10), $\|\nabla f(x_k)-\nabla f(\bar{x}_k)+\nabla f(\bar{x}_k)\|^2$ in (A.18) can be rewritten as

$$\begin{aligned}
&\|\nabla f(x_k)-\nabla f(\bar{x}_k)+\nabla f(\bar{x}_k)\|^2\\
\le&2\|\nabla f(x_k)-\nabla f(\bar{x}_k)\|^2+2\|\nabla f(\bar{x}_k)\|^2\\
=&2\sum_{i=1}^n\|\nabla f_i(x_{i,k})-\nabla f_i(\bar{x}_k)\|^2+2\sum_{i=1}^n\|\nabla f_i(\bar{x}_k)\|^2. \qquad (A.19)
\end{aligned}$$

By Assumption 2(i) we have $\|\nabla f_i(x_{i,k})-\nabla f_i(\bar{x}_k)\| \le L\|x_{i,k}-\bar{x}_k\|$. Then, $\sum_{i=1}^n\|\nabla f_i(x_{i,k})-\nabla f_i(\bar{x}_k)\|^2$ can be rewritten as

$$\begin{aligned}
&\sum_{i=1}^n\|\nabla f_i(x_{i,k})-\nabla f_i(\bar{x}_k)\|^2\\
\le&L^2\sum_{i=1}^n\|x_{i,k}-\bar{x}_k\|^2=L^2\|Y_k\|^2. \qquad (A.20)
\end{aligned}$$

By Assumption 2(ii) and Lemma A.1(ii) in Chen et al. (2024), $\|\nabla f_i(\bar{x}_k)\|^2 \le 2L(f_i(\bar{x}_k)-f_i^*)$, we have

$$\sum_{i=1}^n\|\nabla f_i(\bar{x}_k)\|^2 \le 2L\sum_{i=1}^n(f_i(\bar{x}_k)-f_i^*). \qquad (A.21)$$

Thus, substituting (A.20) and (A.21) into (A.19) gives

$$\begin{aligned}
&\|\nabla f(x_k)-\nabla f(\bar{x}_k)+\nabla f(\bar{x}_k)\|^2\\
\le&2L^2\|Y_k\|^2+4L\left(\sum_{i=1}^n(f_i(\bar{x}_k)-f_i^*)\right). \qquad (A.22)
\end{aligned}$$

Let $M^* = F^* - \frac{1}{n}\sum_{i=1}^n f_i^*$. Then, (A.22) can be rewritten as

$$\begin{aligned}
&\|\nabla f(x_k)-\nabla f(\bar{x}_k)+\nabla f(\bar{x}_k)\|^2\\
\le&2L^2\|Y_k\|^2+4L\left(\sum_{i=1}^n(f_i(\bar{x}_k)-f_i^*)\right)\\
=&2L^2\|Y_k\|^2+4nL(F(\bar{x}_k)-F^*)+4nLM^*.
\end{aligned}$$

This together with (A.18) implies

$$\begin{aligned}\mathbb{E}\|Y_{k+1}\|^2 \leq & \left(1-\rho_{\mathcal{L}}\beta+\frac{4(1+\rho_{\mathcal{L}}\beta)\alpha^2L^2}{\rho_{\mathcal{L}}\beta}\right)\mathbb{E}\|Y_k\|^2 \\ &+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2L}{\rho_{\mathcal{L}}\beta}\mathbb{E}(F(\bar{x}_k)-F^*) \\ &+\frac{n(2+3\rho_{\mathcal{L}}\beta)\beta\Phi^2}{\rho_{\mathcal{L}}}+2nr\beta^2\sigma^2+\frac{n\alpha^2\sigma_\ell^2}{s} \\ &+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2LM^*}{\rho_{\mathcal{L}}\beta}. \end{aligned} \tag{A.23}$$

Step 2: We next focus on the term $F(\bar{x}_k)-F^*$. Multiplying both sides of (11) by $\frac{1}{n}(\mathbf{1}_n^\top\otimes I_r)$ implies

$$\bar{x}_{k+1}=\bar{x}_k-\alpha\overline{\nabla f(x_k)}-\alpha\bar{w}_k+\frac{\beta}{n}(\mathbf{1}_n^\top\otimes I_r)(d_k-e_k). \tag{A.24}$$

Then by (A.24) and Lemma 3.4 in Bubeck (2015), we can derive that

$$\begin{aligned}&F(\bar{x}_{k+1})-F^* \\ \leq& (F(\bar{x}_k)-F^*)+\frac{L}{2}\|\bar{x}_{k+1}-\bar{x}_k\|^2+\langle\nabla F(\bar{x}_k),\bar{x}_{k+1}-\bar{x}_k\rangle \\ =&(F(\bar{x}_k)-F^*)+\frac{L}{2}\|\alpha\overline{\nabla f(x_k)} \\ &-\frac{\beta}{n}\left(\mathbf{1}_n^\top\otimes I_r\right)(d_k-e_k)+\alpha\bar{w}_k\|^2 \\ &-\langle\nabla F(\bar{x}_k),-\frac{\beta}{n}\left(\mathbf{1}_n^\top\otimes I_r\right)(d_k-e_k) \\ &+\alpha\overline{\nabla f(x_k)}+\alpha\bar{w}_k\rangle. \end{aligned} \tag{A.25}$$

Since w_k is independent of $\mathcal{F}_k$, by (A.2) and (A.4), taking conditional expectation of (A.25) with respect to $\mathcal{F}_k$ gives

$$\begin{aligned}&\mathbb{E}(F(\bar{x}_{k+1})-F^*|\mathcal{F}_k) \\ \leq&\mathbb{E}(F(\bar{x}_k)-F^*|\mathcal{F}_k)-\alpha\mathbb{E}(\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle|\mathcal{F}_k) \\ &+\frac{L}{2}\mathbb{E}(\|\alpha\overline{\nabla f(x_k)}-\frac{\beta}{n}(\mathbf{1}_n^\top\otimes I_r)(d_k-e_k)+\alpha\bar{w}_k\|^2|\mathcal{F}_k) \\ &+\frac{\beta}{n}\mathbb{E}(\langle\nabla F(\bar{x}_k),\left(\mathbf{1}_n^\top\otimes I_r\right)(d_k-e_k)\rangle|\mathcal{F}_k) \\ =&(F(\bar{x}_k)-F^*)-\alpha\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle \\ &+\frac{L}{2}\mathbb{E}(\|\alpha\overline{\nabla f(x_k)}-\frac{\beta}{n}(\mathbf{1}_n^\top\otimes I_r)(d_k-e_k)\|^2|\mathcal{F}_k) \\ &+\frac{\alpha^2L}{2}\mathbb{E}\|\bar{w}_k\|^2-\frac{\beta}{n}\langle\nabla F(\bar{x}_k),(\mathbf{1}_n^\top\otimes I_r)e_k\rangle. \end{aligned} \tag{A.26}$$

Letting $m=3$ in (A.10), $\frac{L}{2}\mathbb{E}(\|\alpha\overline{\nabla f(x_k)}-\frac{\beta}{n}(\mathbf{1}_n^\top\otimes I_r)(d_k-e_k)\|^2|\mathcal{F}_k)$ in (A.26) can be rewritten as

$$\begin{aligned}&\frac{L}{2}\mathbb{E}(\|\alpha\overline{\nabla f(x_k)}-\frac{\beta}{n}\left(\mathbf{1}_n^\top\otimes I_r\right)(d_k-e_k)\|^2|\mathcal{F}_k) \\ \leq&\frac{3\beta^2L}{2n^2}\left(\mathbb{E}(\|(\mathbf{1}_n^\top\otimes I_r)e_k\|^2|\mathcal{F}_k)+\mathbb{E}(\|(\mathbf{1}_n^\top\otimes I_r)d_k\|^2|\mathcal{F}_k)\right) \\ &+\frac{3\alpha^2L}{2}\mathbb{E}(\|\overline{\nabla f(x_k)}\|^2|\mathcal{F}_k). \end{aligned} \tag{A.27}$$

Note that $\|(\mathbf{1}_n^\top\otimes I_r)e_k\|^2=n\|\sum_{i=1}^n e_{i,k}\|^2\leq n^2\|e_k\|^2$ and $\|\bar{w}_k\|^2=\|\frac{1}{n}\sum_{i=1}^n w_{i,k}\|^2\leq\frac{1}{n}\|w_k\|^2$. Then, by (A.3), (A.5), (A.6) and the law of total expectation, Substituting (A.27) into (A.26) and taking mathematical expectation on both sides implies

$$\begin{aligned}&\mathbb{E}\left(F(\bar{x}_{k+1})-F^*\right) \\ \leq&\mathbb{E}(F(\bar{x}_k)-F^*)-\alpha\mathbb{E}\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle \\ &+\frac{3n\beta^2L}{2}\left(\Phi^2+r\sigma^2\right)-\frac{\beta}{n}\mathbb{E}\left\langle\nabla F(\bar{x}_k),\left(\mathbf{1}_n^\top\otimes I_r\right)e_k\right\rangle \\ &+\frac{3\alpha^2L}{2}\mathbb{E}\|\overline{\nabla f(x_k)}\|^2+\frac{\alpha^2\sigma_\ell^2L}{2s} \\ \leq&\mathbb{E}(F(\bar{x}_k)-F^*)-\alpha\mathbb{E}\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle \\ &+\frac{3n\beta^2L}{2}\left(\Phi^2+r\sigma^2\right)+\frac{\sqrt{n}\beta\Phi}{2}(\mathbb{E}\|\nabla F(\bar{x}_k)\|^2+1) \\ &+\frac{3\alpha^2L}{2}\mathbb{E}\|\overline{\nabla f(x_k)}\|^2+\frac{\alpha^2\sigma_\ell^2L}{2s}. \end{aligned} \tag{A.28}$$

Note that $\langle\mathbf{a},\mathbf{b}\rangle=\frac{1}{2}\|\mathbf{a}\|^2+\frac{1}{2}\|\mathbf{b}\|^2-\frac{1}{2}\|\mathbf{a}-\mathbf{b}\|^2$, $\forall\mathbf{a},\mathbf{b}\in\mathbb{R}^r$. Then, $-\alpha\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle$ in (A.28) can be rewritten as

$$\begin{aligned}&-\alpha\langle\nabla F(\bar{x}_k),\overline{\nabla f(x_k)}\rangle \\ =&-\frac{\alpha}{2}\|\nabla F(\bar{x}_k)\|^2-\frac{\alpha}{2}\|\overline{\nabla f(x_k)}\|^2+\frac{\alpha}{2}\|\nabla F(\bar{x}_k)-\overline{\nabla f(x_k)}\|^2 \\ \leq&-\frac{\alpha}{2}\|\nabla F(\bar{x}_k)\|^2+\frac{\alpha}{2}\|\nabla F(\bar{x}_k)-\overline{\nabla f(x_k)}\|^2. \end{aligned} \tag{A.29}$$

Let $m=n$ in (A.10). Then, $\|\nabla F(\bar{x}_k)-\overline{\nabla f(x_k)}\|^2$ in (A.29) can be rewritten as

$$\begin{aligned}\|\nabla F(\bar{x}_k)-\overline{\nabla f(x_k)}\|^2=&\|\frac{1}{n}\sum_{i=1}^n(\nabla f_i(\bar{x}_k)-\nabla f_i(x_{i,k}))\|^2 \\ \leq&\frac{1}{n}\sum_{i=1}^n\|\nabla f_i(\bar{x}_k)-\nabla f_i(x_{i,k})\|^2. \end{aligned} \tag{A.30}$$

Thus, substituting (A.20) into (A.30) leads to

$$\left\|\nabla F(\bar{x}_k)-\overline{\nabla f(x_k)}\right\|^2\leq\frac{L^2}{n}\|Y_k\|^2. \tag{A.31}$$

Substituting (A.29) and (A.31) into (A.28) implies

$$\begin{aligned}&\mathbb{E}(F(\bar{x}_{k+1})-F^*)\\ \leq&\mathbb{E}(F(\bar{x}_k)-F^*)-\frac{\alpha}{2}\mathbb{E}\|\nabla F(\bar{x}_k)\|^2+\frac{\alpha L^2}{2n}\mathbb{E}\|Y_k\|^2\\ &+\frac{\sqrt{n}\beta\Phi}{2}\mathbb{E}\|\nabla F(\bar{x}_k)\|^2+\frac{\alpha^2\sigma_\ell^2 L}{2s}\\ &+\frac{3n\beta^2 L}{2}\left(\Phi^2+r\sigma^2\right)+\frac{\sqrt{n}\beta\Phi}{2}\\ &+\frac{3\alpha^2 L}{2}\mathbb{E}\left\|\overline{\nabla f(x_k)}-\nabla F(\bar{x}_k)+\nabla F(\bar{x}_k)\right\|^2. \quad (A.32)\end{aligned}$$

Furthermore, by letting $m=2$ in (A.10) and using (A.31), $\|\overline{\nabla f(x_k)}-\nabla F(\bar{x}_k)+\nabla F(\bar{x}_k)\|^2$ in (A.32) can be rewritten as

$$\begin{aligned}&\left\|\overline{\nabla f(x_k)}-\nabla F(\bar{x}_k)+\nabla F(\bar{x}_k)\right\|^2\\ \leq&2\left\|\overline{\nabla f(x_k)}-\nabla F(\bar{x}_k)\right\|^2+2\left\|\nabla F(\bar{x}_k)\right\|^2\\ \leq&\frac{2L^2}{n}\|Y_k\|^2+2\left\|\nabla F(\bar{x}_k)\right\|^2. \quad (A.33)\end{aligned}$$

Letting $m=n$ in (A.10) and using Lemma A.1(ii) in Chen et al. (2024), $\|\nabla F(\bar{x}_k)\|^2$ in (A.33) can be rewritten as

$$\|\nabla F(\bar{x}_k)\|^2\leq 2L\left(F(\bar{x}_k)-F^*\right). \quad (A.34)$$

Thus, substituting (A.33) and (A.34) into (A.32) implies

$$\begin{aligned}&\mathbb{E}\left(F(\bar{x}_{k+1})-F^*\right)\\ \leq&\left(1+6\alpha^2L^2+\sqrt{n}L\beta\Phi\right)\mathbb{E}(F(\bar{x}_k)-F^*)\\ &-\frac{\alpha}{2}\mathbb{E}\|\nabla F(\bar{x}_k)\|^2+\frac{\alpha L^2(1+6\alpha L)}{2n}\mathbb{E}\|Y_k\|^2\\ &+\frac{\alpha^2\sigma_\ell^2L}{2s}+\frac{3n\beta^2L}{2}(\Phi^2+r\sigma^2)+\frac{\sqrt{n}\beta\Phi}{2}. \quad (A.35)\end{aligned}$$

If Assumption 4 holds, then (A.35) can be rewritten as

$$\begin{aligned}&\mathbb{E}\left(F(\bar{x}_{k+1})-F^*\right)\\ \leq&\left(1-\mu\alpha+6\alpha^2L^2+\sqrt{n}L\beta\Phi\right)\mathbb{E}(F(\bar{x}_k)-F^*)\\ &+\frac{\alpha L^2(1+6\alpha L)}{2n}\mathbb{E}\|Y_k\|^2+\frac{\alpha^2\sigma_\ell^2L}{2s}\\ &+\frac{3n\beta^2L}{2}(\Phi^2+r\sigma^2)+\frac{\sqrt{n}\beta\Phi}{2}. \quad (A.36)\end{aligned}$$

Step 3: We get the main result of the theorem based on **Steps 1** and **2**.

Let

$$\begin{aligned}\theta_1=&\max\{1-\mu\alpha+6\alpha^2L^2+\sqrt{n}L\beta\Phi+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2L}{\rho_{\mathcal{L}}\beta},\\ &1-\rho_{\mathcal{L}}\beta+\frac{\alpha L^2(1+6\alpha L)}{2n}+\frac{4(1+\rho_{\mathcal{L}}\beta)\alpha^2L^2}{\rho_{\mathcal{L}}\beta}\}, \quad (A.37)\end{aligned}$$

$$\begin{aligned}\theta_2=&\frac{(2n+L)\alpha^2\sigma_\ell^2}{2s}+\frac{3n\beta^2L}{2}(\Phi^2+r\sigma^2)\\ &+\frac{n(2+3\rho_{\mathcal{L}}\beta)\beta\Phi^2}{\rho_{\mathcal{L}}}+2nr\beta^2\sigma^2+\frac{\sqrt{n}\beta\Phi}{2}\\ &+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2LM^*}{\rho_{\mathcal{L}}\beta}. \quad (A.38)\end{aligned}$$

Then, by (A.37) and (A.38), summing (A.23) and (A.36) implies

$$\begin{aligned}&\mathbb{E}(\|Y_{k+1}\|^2+F(\bar{x}_{k+1})-F^*)\\ \leq&\theta_1\mathbb{E}(\|Y_k\|^2+F(\bar{x}_k)-F^*)+\theta_2. \quad (A.39)\end{aligned}$$

By iteratively computing (A.39), for any $k=0,\ldots,K$, the following inequality holds:

$$\begin{aligned}&\mathbb{E}(\|Y_{k+1}\|^2+F(\bar{x}_{k+1})-F^*)\\ \leq&\theta_1^{k+1}\mathbb{E}(\|Y_0\|^2+F(\bar{x}_0)-F^*)+\theta_2\sum_{m=0}^{k}\theta_1^{k-m}. \quad (A.40)\end{aligned}$$

When K is sufficiently large, we have $0<\theta_1<1$. Since $\ln(1-x)\leq -x$, $\forall x<1$, we can obtain that $\theta_1^{K+1}=\exp((K+1)\ln(1-(1-\theta_1)))\leq\exp(-(K+1)(1-\theta_1))$. Substituting (A.37) into the inequality above implies

$$\begin{aligned}\theta_1^{K+1}\leq&\max\{\exp((K+1)(-\mu\alpha+6\alpha^2L^2+\sqrt{n}L\beta\Phi\\ &+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2L}{\rho_{\mathcal{L}}\beta})),\\ &\exp((K+1)(-\rho_{\mathcal{L}}\beta+\frac{\alpha L^2(1+6\alpha L)}{2n}\\ &+\frac{4(1+\rho_{\mathcal{L}}\beta)\alpha^2L^2}{\rho_{\mathcal{L}}\beta}))\}. \quad (A.41)\end{aligned}$$

Note that $p_1>p_2$ and $p_2+p_5>p_1$ in Assumption 3. Then, we have $-\mu\alpha+6\alpha^2L^2+\sqrt{n}L\beta\Phi+\frac{8n(1+\rho_{\mathcal{L}}\beta)\alpha^2L}{\rho_{\mathcal{L}}\beta}=O(-\frac{\mu\alpha}{2})$, $-\rho_{\mathcal{L}}\beta+\frac{\alpha L^2(1+6\alpha L)}{2n}+\frac{4(1+\rho_{\mathcal{L}}\beta)\alpha^2L^2}{\rho_{\mathcal{L}}\beta}=O(-\frac{\rho_{\mathcal{L}}\beta}{2})$. Thus, (A.41) can be rewritten as

$$\begin{aligned}&\theta_1^{K+1}\\ =&O(\max\{\exp(-\frac{(K+1)\mu\alpha}{2}),\exp(-\frac{(K+1)\rho_{\mathcal{L}}\beta}{2})\})\\ =&O(\max\{\exp(-\frac{\mu a_1}{2}K^{1-p_1}),\exp(-\frac{\rho_{\mathcal{L}}a_2}{2}K^{1-p_2})\}). \quad (A.42)\end{aligned}$$

Moreover, by $2p_2-2p_4>p_1$, $\theta_2\sum_{m=0}^{K}\theta_1^{K-m}$ in (A.40) can be rewritten as

$$\theta_2\sum_{m=0}^{K}\theta_1^{K-m}=\frac{\theta_2(1-\theta_1^{K+1})}{1-\theta_1}=O\left(\frac{\theta_2}{1-\theta_1}\right)$$
$$=O\left(\frac{1}{K^{\min\{p_1-p_2,p_2+p_5-p_1,2p_2-2p_4-p_1\}}}\right). \tag{A.43}$$

For any node $i\in\mathcal{V}$, by Lemma 3.4 in Bubeck (2015), we have

$$F(x_{i,K+1})-F(\bar{x}_{K+1})\leq\langle\nabla F(\bar{x}_{K+1}),x_{i,K+1}-\bar{x}_{K+1}\rangle+\frac{L}{2}\|\bar{x}_{K+1}-x_{i,K+1}\|^2. \tag{A.44}$$

Note that $\langle\mathbf{a},\mathbf{b}\rangle\leq\|\mathbf{a}\|\|\mathbf{b}\|\leq\frac{\|\mathbf{a}\|^2+\|\mathbf{b}\|^2}{2}$, $\forall\mathbf{a},\mathbf{b}\in\mathbb{R}^r$. Then, (A.44) can be rewritten as

$$\begin{aligned}&F(x_{i,K+1})-F(\bar{x}_{K+1})\\ \leq&\frac{\|\nabla F(\bar{x}_{K+1})\|^2+\|\bar{x}_{K+1}-x_{i,K+1}\|^2}{2}\\ &+\frac{L}{2}\|\bar{x}_{K+1}-x_{i,K+1}\|^2\\ =&\frac{L+1}{2}\|\bar{x}_{K+1}-x_{i,K+1}\|^2+\frac{\|\nabla F(\bar{x}_{K+1})\|^2}{2}.\end{aligned} \tag{A.45}$$

By Lemma A.1(ii) in Chen et al. (2024) we have $\|\nabla F(\bar{x}_{K+1})\|^2\leq 2L(F(\bar{x}_{K+1})-F^*)$. This together with (A.45) gives $F(x_{i,K+1})-F(\bar{x}_{K+1})\leq\frac{L+1}{2}\|\bar{x}_{K+1}-x_{i,K+1}\|^2+L(F(\bar{x}_{K+1})-F^*)$. Thus, we have

$$\begin{aligned}&F(x_{i,K+1})-F(\bar{x}_{K+1})\\ \leq&\frac{L+1}{2}\sum_{i=1}^{n}\|\bar{x}_{K+1}-x_{i,K+1}\|^2+L(F(\bar{x}_{K+1})-F^*)\\ =&\frac{L+1}{2}\|Y_{K+1}\|^2+L(F(\bar{x}_{K+1})-F^*).\end{aligned} \tag{A.46}$$

Furthermore, for any node $i\in\mathcal{V}$, by (A.46), we have

$$\begin{aligned}&F(x_{i,K+1})-F^*\\ =&(F(x_{i,K+1})-F(\bar{x}_{K+1}))+(F(\bar{x}_{K+1})-F^*)\\ \leq&\frac{L+1}{2}\|Y_{K+1}\|^2+(L+1)(F(\bar{x}_{K+1})-F^*)\\ \leq&(L+1)\left(\|Y_{K+1}\|^2+(F(\bar{x}_{K+1})-F^*)\right).\end{aligned} \tag{A.47}$$

Hence, by substituting (A.42), (A.43) and (A.47) into (A.40), we have

$$\begin{aligned}&\mathbb{E}(F(x_{i,K+1})-F^*)\\ =&O\left(\frac{1}{K^{\min\{p_1-p_2,p_2+p_5-p_1,2p_2-2p_4-p_1\}}}\right).\end{aligned} \tag{A.48}$$

Note that by Lemma A.1(ii) in Chen et al. (2024), we have

$$\|\nabla F(x_{i,K+1})\|^2\leq 2L(F(x_{i,K+1})-F^*). \tag{A.49}$$

Then, taking the mathematical expectation on (A.49) and substituting (A.48) into (A.49) imply

$$\begin{aligned}&\mathbb{E}\|\nabla F(x_{i,K+1})\|^2\\ =&O\left(\frac{1}{K^{\min\{p_1-p_2,p_2+p_5-p_1,2p_2-2p_4-p_1\}}}\right).\end{aligned} \tag{A.50}$$

Note that for any $\psi\in[1,2]$, the function $x^{\frac{\psi}{2}}$ is concave for $x\geq 0$. Then, by Jensen's inequality (Chow & Teicher, 2012) we have $\mathbb{E}\|\nabla F(x_{i,K+1})\|^{\psi}=\mathbb{E}\left(\|\nabla F(x_{i,K+1})\|^2\right)^{\frac{\psi}{2}}\leq(\mathbb{E}\|\nabla F(x_{i,K+1})\|^2)^{\frac{\psi}{2}}$. Thus, substituting (A.50) into this inequality implies

$$\begin{aligned}&\mathbb{E}\|\nabla F(x_{i,K+1})\|^{\psi}\\ =&O\left(\frac{1}{K^{\frac{\psi}{2}\min\{p_1-p_2,p_2+p_5-p_1,2p_2-2p_4-p_1\}}}\right),\end{aligned}$$

the theorem is thereby proved. □

References

Bubeck, S. (2015). Convex optimization: algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4), 231-357.

Cao, X., & Başar, T. (2021). Decentralized online convex optimization with event-triggered communications. *IEEE Transactions on Signal Processing*, 69, 284-299.

Chen, J., Wang, J., & Zhang, J. F. (2024). Differentially private distributed nonconvex stochastic optimization with quantized communications. *arXiv preprint arXiv:2403.18254*.

Chen, X., Wang, C., Yang, Q., Hu, T., & Jiang, C. (2023). Locally differentially private high-dimensional data synthesis. *Science China Information Sciences*, 66(1), 112101.

Chow, Y. S., & Teicher, H. (2012). *Probability Theory: Independence, Interchangeability, Martingales*. New York USA Springer-Verlag.

Ding, J., Liang, G., Bi, J., & Pan, M. (2021). Differentially private and communication efficient collaborative learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7219-7227.

Dong, Z., Jin, Y., Mao, S., Ren, W., Du, W., & Tang, Y. (2025). Distributed optimization with asynchronous computation and event-triggered communication. *IEEE Transactions on Automatic Control*, 70(2), 1084-1099.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4), 211-407.

Espitia, N., Karafyllis, I., & Krstic, M. (2021). Event-triggered boundary control of constant-parameter reaction-diffusion PDEs: A small-gain approach. *Automatica*, 128, 109562.

Gao, L., Deng, S., & Ren, W. (2019). Differentially private consensus with an event-triggered mechanism. *IEEE Transactions on Control of Network Systems*, 6(1), 60-71.

George, J., & Gurram, P. (2020). Distributed stochastic gradient descent with event-triggered communication. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7169-7178.

Gratton, C., Venkategowda, N. K., Arablouei, R., & Werner, S. (2021). Privacy-preserved distributed learning with zeroth-order optimization. *IEEE Transactions on Information Forensics and Security*, 17, 265-279.

He, X., Yi, X., Zhao, Y., Johansson, K. H., & Gupta, V. (2023). Asymptotic analysis of federated learning under event-triggered communication. *IEEE Transactions on Signal Processing*, 71, 2654-2667.

Horn, R. A., & Johnson, C. R. (2013). *Matrix Analysis.* Cambridge University Press.

Huang, Z., Hu, R., Guo, Y., Chan-Tin, E., & Gong, Y. (2019). DP-ADMM: ADMM-based distributed learning with differential privacy. *IEEE Transactions on Information Forensics and Security*, 15, 1002-1012.

Karimi, H., Nutini, J., & Schmidt, M. (2016). Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Machine Learning and Knowledge Discovery in Databases: European Conference*, 795-811.

LeCun, Y., Cortes, C., & Burges, C. J. C. (1998). The MNIST database of handwritten digits. *National Institute of Standards and Technology.*

Le Ny, J., & Pappas, G. J. (2014). Differentially private filtering. *IEEE Transactions on Automatic Control*, 59(2), 341-354.

Li, C., Zhou, P., Xiong, L., Wang, Q., & Wang, T. (2018). Differentially private distributed online learning. *IEEE Transactions on Knowledge and Data Engineering*, 30(8), 1440-1453.

Liang, L. M., Ding, R. Q., Liu, S., & Su, R. (2024). Event-triggered privacy-preserving consensus control with edge-based additive noise. *IEEE Transactions on Automatic Control*, 69(10), 7059-7066.

Liu, C., Johansson, K. H., & Shi, Y. (2024). Distributed empirical risk minimization with differential privacy. *Automatica*, 162, 111514.

Liu, C., Li, H., Shi, Y., & Xu, D. (2020a). Distributed event-triggered gradient method for constrained convex minimization. *IEEE Transactions on Automatic Control*, 65(2), 778-785.

Liu, L., Kawano, Y., & Cao, M. (2025) Design of stochastic quantizers for privacy preservation, *arXiv:2403.03048.*

Liu, X. K., Zhang, J. F., & Wang, J. (2020b). Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems. *Automatica*, 122, 109283.

Lu, Y., & Zhu, M. (2018). Privacy preserving distributed optimization using homomorphic encryption. *Automatica*, 96, 314-325.

Mao, S., Yang, M., Yang, W., Tang, Y., Zheng, W. X., Gu, J., & Werner, H.(2023). Differentially private distributed optimization with an event-triggered mechanism, *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 70(7), 2943-2956.

Miao, Y., Kuang, D., Li, X., Leng, T., Liu, X., & Ma, J. (2024). Efficient privacy-preserving federated learning under dishonest-majority setting. *Science China Information Sciences*, 67(5), 159102.

Nedic, A., & Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1), 48-61.

Reisizadeh, A., Taheri, H., Mokhtari, A., Hassani, H. & Pedarsani, R. (2019a). Robust and communication-efficient collaborative learning. *Advances in Neural Information Processing Systems*, 8388-8399.

Reisizadeh, A., Mokhtari, A., Hassani, H., & Pedarsani, R. (2019b). An exact quantized decentralized gradient descent algorithm. *IEEE Transactions on Signal Processing*, 67(19), 4934-4947.

Singh, N., Data, D., George, J., & Diggavi, S. (2022). SPARQ-SGD: Event-triggered and compressed communication in decentralized optimization. *IEEE Transactions on Automatic Control*, 68(2), 721-736.

Wang, A., Liao, X., & He, H. (2019). Event-triggered differentially private average consensus for multi-agent network. *IEEE/CAA Journal of Automatica Sinica*, 6(1), 75-83.

Wang, J., Ke, J., & Zhang, J. F. (2024). Differentially private bipartite consensus over signed networks with time-varying noises. *IEEE Transactions on Automatic Control*, 69(9), 5788-5803.

Wang, J., & Krstic, M. (2022). Event-triggered adaptive control of a parabolic PDE-ODE cascade with piecewise-constant inputs and identification. *IEEE Transactions on Automatic Control*, 68(9), 5493-5508.

Wang, J., Zhang, J. F., & He, X. (2022). Differentially private distributed algorithms for stochastic aggregative games. *Automatica*, 142, 110440.

Wang, Y. (2019). Privacy-preserving average consensus via state decomposition. *IEEE Transactions on Automatic Control*, 64(11), 4711-4716.

Wang, Y., & Başar, T. (2023a). Decentralized nonconvex optimization with guaranteed privacy and accuracy. *Automatica*, 150, 110858.

Wang, Y., & Başar, T. (2023b). Quantization enabled privacy protection in decentralized stochastic optimization. *IEEE Transactions on Automatic Control*, 68(7), 4038-4052.

Xin, R., Khan, H. A., & Kar, S. (2022). A fast randomized incremental gradient method for decentralized nonconvex optimization. *IEEE Transactions on Automatic Control*, 67(10), 5150-5165.

Xu, J., Zhang, W., & Wang, F. (2022). A (DP) ^{2} SGD: asynchronous decentralized parallel stochastic gradient descent with differential privacy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 8036-8047.

Xu, L., Yi, X., Shi, Y., Johansson, K. H., Chai, T., & Yang, T. (2024). Distributed nonconvex optimization with event-triggered communication. *IEEE Transactions on Automatic Control*, 69(4), 2745-2752.

Yan, G., Li, T., Wu, K., & Song, L. (2024). Killing two birds with one stone: Quantization achieves privacy in distributed learning. *Digital Signal Processing*, 146, 104353.

Yi, X., Yao, L., Yang, T., George, J., & Johansson, K. H. (2018). Distributed optimization for second-order multi-agent systems with dynamic event-triggered communication. In *2018 IEEE Conference on Decision and Control*, 3397-3402.

Yuan, Y., He, W. L., Du, W. L., Tian, Y. C., Han, Q. L.,& Qian, F. (2024). Distributed gradient tracking for differentially private multi-agent optimization with a dynamic event-triggered mechanism, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 54(5), 3044-3055.

Zhang, J. F., Tan, J., & Wang, J. (2021). Privacy security in control systems. *Science China Information Sciences*, 64(7), 176201.

Zhang, W. T., Shi, Y., Zhang, B. Y., & Yuan, D. M. (2024). Improved dynamic regret of distributed online multiple Frank-Wolfe convex optimization. *Science China Information Sciences*, 67(11), 212201.

Zhang, X., Khalili, M. M., & Liu, M. (2018). Improving the privacy and accuracy of ADMM-based distributed algorithms. In *International Conference on Machine Learning*, 5796-5805.

Zhang, Z., Zhang, Y., Guo, D., Zhao, S., & Zhu, X. (2023). Communication-efficient federated continual learning for distributed learning system with Non-IID data. *Science China Information Sciences*, 66(2), 122102.

Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. *Advances in Neural Information Processing systems*, 32. 14774-14784.

Zorich, V. A. (2015). *Mathematical Analysis I*. Berlin German Springer.