Privacy-Preserving Distributed Estimation with Limited Data Rate

Jieming Ke, *Student Member, IEEE*, Jimin Wang, *Member, IEEE*, and Ji-Feng Zhang, *Fellow, IEEE*

***Abstract*— This paper focuses on the privacy-preserving distributed estimation problem with a limited data rate, where the observations are sensitive information. Specifically, a binary-valued quantizer-based privacy-preserving distributed estimation algorithm is developed, which improves the algorithm's privacy-preserving capability and simultaneously reduces the communication costs. The algorithm's privacy-preserving capability, measured by the Fisher information matrix, is dynamically enhanced over time. Notably, the Fisher information matrix of the output signals with respect to the sensitive information converges to zero at a polynomial rate. Regarding the communication costs, each sensor transmits only 1 bit of information to its neighbours at each time. Additionally, the information receiver does not require any *a priori* knowledge on the upper bounds of the estimates' norms to decode the quantized information. While achieving the requirements of privacy preservation and reducing communication costs, the estimates of the algorithm converge almost surely to the true value of the unknown parameter even when the privacy noises in the binary-valued quantizers increase over time. A polynomial almost sure convergence rate is obtained, and then the trade-off between privacy and convergence rate is established. A numerical example demonstrates the main results.**

***Index Terms*— Distributed estimation; privacy preservation; binary-valued quantization; Fisher information.**

I. Introduction

DISTRIBUTED estimation has received close attention in the past decade due to its extensive applications in various fields, such as biological networks, online machine learning, and smart grids [1], [2]. Different from traditional centralized estimation, the observations of distributed estimation are collected by different sensors in the communication network. Therefore, a network communication is required to fuse the observations from each sensor. However, in actual distributed systems, observations may contain sensitive information, and the network communication may lead to sensitive information leakage. For example, medical research usually requires clinical observation data of patients from different hospitals, which involve the patients' personal data [3], [4]. Motivated by this sensitive topic, this paper investigates how to achieve distributed estimation while ensuring that the observations do not leaked.

The current literature offers several privacy-preserving methods for distributed systems. One of the methods is the homomorphic encryption method [5]–[8], which provides high-dimensional security while ensuring control accuracy. Another commonly used method is the stochastic obfuscation method [9]–[12], which has the advantages of low computational complexity and high timeliness. Differential privacy metric or Fisher information is used to quantify the privacy preserving capability of the algorithms based on the stochastic obfuscation method. Other methods include the state decomposition method [13] and the privacy mask method [14].

Among the existing methods, quantizer-based privacy-preserving methods have recently received significant attention [15]–[18], owing to their improved privacy preservation while reducing communication costs. For example, [15] introduces a dithered lattice quantizer-based differential privacy method for federated learning. Additionally, [16] proposes a dynamic quantization-based homomorphic encryption method for the distributed economic dispatch problem. Moreover, [17] uses an unbiased ternary quantizer to preserve sensitive information, while [18] analyzes the privacy-preserving capability of the stochastic quantizer and applies it to the output tracking control problem.

Despite the remarkable progress, the existing quantizer-based privacy-preserving methods suffer from various limitations. Firstly, many existing methods are based on infinite-level quantizers [15], [18], but real digital networks have a communication data rate constraint. Although the quantizers in [16], [17] are finite-level, the information receiver should know the upper bounds of the states' norms to decode the quantized information. Besides, [15] highlights that its infinite-level quantizer can be transformed into a finite-level quantizer, but the transformation also requires similar *a priori* knowledge. Such *a priori* knowledge is not always available in practice, and therefore it is challenging to implement distributed algorithms under a finite data rate without *a priori* knowledge on the upper bounds of the states' norms. This is because in such

This work was supported by National Natural Science Foundation of China under Grants T2293772 and 62203045. Corresponding author: Ji-Feng Zhang.

Jieming Ke is with the Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: kejieming@amss.ac.cn)

Jimin Wang is with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083; and also with the Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China (e-mail: jimwang@ustb.edu.cn)

Ji-Feng Zhang is with the School of Automation and Electrical Engineering, Zhongyuan University of Technology, Zhengzhou 450007, Henan Province; and also with the Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. (e-mail: jif@iss.ac.cn)

a situation, it is hard to guarantee the unbiasedness of the quantizer [19]. Secondly, there are still few results to quantify the improvement in privacy preservation by quantizer-based methods compared to unquantized ones. [17], [18] reveal that their quantizer-based methods can achieve $(0, \delta)$-differential privacy, which is quite different from the ϵ-differential privacy that is commonly achieved in the unquantized case [4], [9], [20]. Therefore, differential privacy is difficult to be used to quantify the improvement in privacy-preserving capability brought by these quantizer-based methods compared to unquantized ones.

This paper overcomes the above limitations under the framework of the privacy-preserving distributed estimation problem. The key solution lies in proposing a novel binary-valued quantizer-based (BQB) privacy-preserving method. The BQB method preserves sensitive information by using binary-valued quantizers with privacy noises, and demonstrates strong privacy, low communication costs and wide applicability. The BQB method has a stronger privacy-preserving capability than the unquantized ones. The improvement is quantified using the Fisher information [10], [21], instead of differential privacy metric [11], [23]–[25]. Regarding the communication costs, each sensor transmits only 1 bit of information to its neighbours at each time. The data rate will not increase with the dimension of the estimates, because the high-dimensional estimates are periodically compressed into scalars before quantization. Considering the applicability, the information receiver does not require any *a priori* knowledge on the upper bounds of the estimates' norms. Therefore, the *a priori* knowledge can be removed because the convergence analysis does not rely on the unbiasedness of the quantizers as in [15], [17]. Instead, the proposed convergence analysis solely relies on the strict monotonicity of the privacy noise distribution function [22], [26].

Furthermore, a BQB privacy-preserving distributed estimation algorithm (BQB-PPDEA) is proposed. Different from the static privacy-preserving capabilities of the existing algorithms [3], [23], [27], BQB-PPDEA achieves a dynamically enhanced privacy. The realization of dynamically enhanced privacy is due to overcoming the difficulties in the following two aspects. The first difficulty is to find a privacy metric that characterizes the dynamically enhanced privacy. Common differential privacy metrics for distributed estimation, learning, and optimization algorithms [3], [15], [25], [28] focus on describing the privacy-preserving capability for the observation sequences. Therefore, it is challenging to characterize the privacy-preserving capability for each observation dynamically. To overcome this difficulty, this paper employs the Fisher information as the privacy metric. For BQB-PPDEA, the Fisher information matrix of the output signals with respect to the sensitive information converges to zero at a polynomial rate, which characterizes the dynamic changes of the privacy-preserving capability of BQB-PPDEA. The second difficulty involves algorithm design. Existing privacy-preserving distributed estimation algorithms [25], [29] adopt decaying privacy noises at the outputs to avoid the algorithms losing convergence while achieving privacy preservation. However, the decaying privacy noises cannot enable the corresponding algorithms to achieve the dynamically enhanced privacy. To overcome this difficulty, in BQB-PPDEA, privacy noises in the innovation step are removed, and the step-sizes in the algorithm are adjusted accordingly. Based on these adjustments, the estimates of the algorithm can still achieve convergence when the privacy noises are constant or even increasing, further enabling the algorithm to achieve the dynamically enhanced privacy.

This paper further studies the privacy-preserving capability of BQB-PPDEA in complex communication network environments, especially link failures. Among the existing literature, [1] models link failures as an independent and identically distributed (i.i.d.) graph sequence, and [30] extends the model to the Markovian switching graph case. But, both of these works does not consider the privacy issue. Link failures in the communication network will reduce the communication frequency, and thereby improve the privacy-preserving capability of the algorithm. Using the chain rule for Fisher information matrices [31] can rigorously characterize the impact of the Markovian switching graphs on the privacy-preserving capability.

This paper proposes a novel BQB-PPDEA. The main contributions of this paper are summarized as follows.

1) BQB-PPDEA's privacy-preserving capability is dynamically enhanced. The Fisher information matrix of the output signals with respect to the sensitive information converges to zero at a polynomial rate. Furthermore, the stationary distribution of Markovian switching graphs is shown to be the key factor affecting the privacy-preserving capability.
2) Under BQB-PPDEA, each sensor transmits only 1 bit of information to its neighbours at each time. This is the lowest data rate among existing quantizer-based privacy-preserving distributed algorithms [15]–[17]. Additionally, the information receiver is not required to know the upper bounds of the estimate's norms [15]–[17] to decode the quantized information.
3) The almost sure convergence of BQB-PPDEA is proved even with increasing privacy noises. A polynomial almost sure convergence rate is also obtained. BQB-PPDEA is the first distributed estimation algorithm that achieves convergence under a finite data rate and increasing noises, even without considering privacy preservation [32], [33].
4) The trade-off between privacy and convergence rate is established. Better privacy implies a slower convergence rate, and vice versa. Furthermore, the sensor operators can determine their own preference for the privacy and convergence rate by properly selecting step-sizes and privacy noise distributions.

The rest of paper is organized as follows. Section II formulates the problem, and introduces the Fisher information-based privacy metric. Section III proposes BQB-PPDEA. Section IV analyzes the privacy-preserving capability of BQB-PPDEA. Section V proves the almost sure convergence of the algorithm, and calculates the almost sure convergence rate. Section VI establishes the trade-off between privacy and convergence rate. Section VII uses a numerical example to demonstrate the main results. Section VIII gives a concluding remark for this paper.

Notation

In the rest of the paper, $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$ are the sets of natural numbers, real numbers, n-dimensional real vectors, and $n\times m$-dimensional real matrices, respectively. $\|x\|$ is the Euclidean norm for vector x, and $\|A\|$ is the induced matrix norm for matrix A. A^+ is the pseudo-inverse of matrix A. I_n is an $n\times n$ identity matrix. $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true; and 0, otherwise. $\mathbf{1}_n$ is the n-dimensional vector whose elements are all ones. diag$\{\cdot\}$ denotes the block matrix formed in a diagonal manner of the corresponding numbers or matrices. col$\{\cdot\}$ denotes the column vector stacked by the corresponding vectors. $\otimes$ denotes the Kronecker product. $\mathcal{N}(0,\sigma^2)$, *Lap*$(0,b)$ and *Cauchy*$(0,r)$ represent Gaussian distribution with density function $\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-x^2/2\sigma^2\right)$, Laplacian distribution with density function $\frac{1}{2b}\exp\left(-|x|/b\right)$ and Cauchy distribution with density function $1\Big/\left(\pi r\left[1+(x/r)^2\right]\right)$, respectively.

II. Preliminaries and problem formulation

A. Graph preliminaries

In this paper, the communication graph is switching among topology graphs $\mathcal{G}^{(1)},\ldots,\mathcal{G}^{(M)}$, where $\mathcal{G}^{(u)} = \left(\mathcal{V},\mathcal{E}^{(u)},\mathcal{A}^{(u)}\right)$ for all $u=1,\ldots,M$. $\mathcal{V}=\{1,\ldots,N\}$ is the set of the sensors. $\mathcal{E}^{(u)} \in \{(i,j): i,j\in\mathcal{V}\}$ is the edge set. $\mathcal{A}^{(u)} = (a_{ij}^{(u)})_{N\times N}$ represents the symmetric weighted adjacency matrix of the graph whose elements are all non-negative. $a_{ij}^{(u)} > 0$ if and only if $(i,j)\in\mathcal{E}^{(u)}$. Besides, $\mathcal{N}_i^{(u)} = \{j : (i,j)\in\mathcal{E}^{(u)}\}$ is used to denote the sensor i's the neighbour set corresponding to the graph $\mathcal{G}^{(u)}$. Define Laplacian matrix as $\mathcal{L}^{(u)} = \mathcal{D}^{(u)} - \mathcal{A}^{(u)}$, where $\mathcal{D}^{(u)} = \operatorname{diag}\left(\sum_{i\in\mathcal{N}_1} a_{i1}^{(u)},\ldots,\sum_{i\in\mathcal{N}_N} a_{iN}^{(u)}\right)$.

The union of $\mathcal{G}^{(1)},\ldots,\mathcal{G}^{(M)}$ is denoted by $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{A})$, where $\mathcal{E}=\bigcup_{r=1}^M \mathcal{E}^{(u)}$, and $\mathcal{A}=\sum_{u=1}^M \mathcal{A}^{(u)}$. Besides, set $\mathcal{N}_i = \{j : (i,j)\in\mathcal{E}\}$.

Assumption 1. The union graph $\mathcal{G}$ is connected.

Remark 1. The connection assumption is necessary and commonly adopted in the distributed estimation problem [1], [30].

The communication graph at time k, denoted by $\mathcal{G}_k$, is associated with a homogeneous Markovian chain $\{m_k : k\in\mathbb{N}\}$ with a state space $\{1,\ldots,M\}$, transition probability $p_{uv} = \mathbb{P}\{m_k = v | m_{k-1} = u\}$, and stationary distribution $\pi_u = \lim_{k\to\infty}\mathbb{P}\{m_k = u\}$. If $m_k = u$, then $\mathcal{G}_k = \mathcal{G}^{(u)}$. Denote $q_{ij,k} = \mathbb{P}\{(i,j)\in\mathcal{E}^{(m_k)}\}$. For convenience, $\mathcal{E}^{(m_k)}$, $\mathcal{A}^{(m_k)}$, $a_{ij}^{(m_k)}$, $\mathcal{N}_i^{(m_k)}$, $\mathcal{L}^{(m_k)}$ and $\mathcal{D}^{(m_k)}$ are abbreviated as $\mathcal{E}_k$, $\mathcal{A}_k$, $a_{ij,k}$, $\mathcal{N}_{i,k}$, $\mathcal{L}_k$ and $\mathcal{D}_k$, respectively, in the rest of paper.

Remark 2. Markovian switching graphs can be used to model the link failures [30], [34]. $a_{ij,k} > 0$ implies that the communication link between the sensors i and j is normal. $a_{ij,k} = 0$ implies that the communication link fails.

Remark 3. Given $p_{u,1} = \mathbb{P}\{\mathcal{G}_1 = \mathcal{G}^{(u)}\}$, $q_{ij,k}$ can be recursively obtained by

$$q_{ij,k} = \sum_{u\in\mathbb{G}_{ij}} p_{u,k},\; p_{u,k+1} = \mathbb{P}\{\mathcal{G}_{k+1} = \mathcal{G}^{(u)}\} = \sum_{v=1}^{M} p_{v,k}p_{vu},$$

where $\mathbb{G}_{ij} = \{u : (i,j)\in\mathcal{E}^{(u)}\}$. By Theorem 1.2 of [35], we have $q_{ij,k} = \sum_{u\in\mathbb{G}_{ij}} \pi_u + O\left(\lambda_p^k\right)$ for some $\lambda_p\in(0,1)$. Especially when the initial distribution $\{p_{u,1} : u = 1,\ldots,M\}$ is the stationary distribution $\{\pi_u : u=1,\ldots,M\}$, we have $q_{ij,k} = \sum_{u\in\mathbb{G}_{ij}}\pi_u$.

B. Observation model

In the multi-sensor system coupled by the Markovian switching graphs, the sensor i observes the unknown parameter $\theta\in\mathbb{R}^n$ from the observation model

$$y_{i,k} = H_{i,k}\theta + w_{i,k},\; i=1,\ldots,N,\; k\in\mathbb{N}, \tag{1}$$

where θ is the unknown parameter, k is the time index, $w_{i,k}\in\mathbb{R}^{m_i}$ is the observation noise, and $y_{i,k}\in\mathbb{R}^{m_i}$ is the observation. $H_{i,k}\in\mathbb{R}^{m_i\times n}$ is the random measurement matrix.

Assumptions for the observation model (1) are given as follows.

Assumption 2. The random measurement matrix $H_{i,k}$ is not necessarily available, but its mean value $\bar{H}_i$ is known. $\sum_{i=1}^N \bar{H}_i^\top \bar{H}_i$ is invertible.

Remark 4. Assumption 2 can be used to describe possible sensor failures [1], [30]. If the sensor i fails with a probability of p_i^f, then $H_{i,k} = \delta_{i,k}\hat{H}_i$, where $\hat{H}_i$ is the deterministic measurement matrix without considering sensor failures, and $\delta_{i,k}$ is a Bernoulli variable with $\mathbb{P}\{\delta_{i,k}=1\} = 1-p_i^f$. For this model of sensor failures, $\bar{H}_i = (1-p_i^f)\hat{H}_i$. Then, when $p_i^f < 1$ for all i, $\sum_{i=1}^N \bar{H}_i^\top \bar{H}_i$ is invertible if and only if $\sum_{i=1}^N \hat{H}_i^\top \hat{H}_i$ is invertible.

Remark 5. The invertibility condition on $\sum_{i=1}^N \bar{H}_i^\top \bar{H}_i$ is the global observability assumption, which is commonly adopted in distributed estimation problem [1], [36], [37].

Assumption 3. $\{w_{i,k}, H_{i,k} : i\in\mathcal{V}, k\in\mathbb{N}\}$ is an independent sequence such that

$$\sup_{i\in\mathcal{V},\; k\in\mathbb{N}} \mathbb{E}\left\|w_{i,k}\right\|^\rho < \infty, \tag{2}$$

$$\sup_{i\in\mathcal{V},\; k\in\mathbb{N}} \mathbb{E}\left\|H_{i,k} - \bar{H}_i\right\|^\rho < \infty, \tag{3}$$

for some $\rho > 2$, and independent of the graph sequence $\{\mathcal{G}_k : k\in\mathbb{N}\}$.

Remark 6. If ρ in (2) and (3) takes different values, for example ρ_1 and ρ_2, respectively, then by Lyapunov inequality [39], (2) and (3) still hold for $\rho = \min\{\rho_1,\rho_2\}$.

C. Dynamically enhanced privacy

This section will formulate the privacy-preserving distributed estimation problem, where the observation $y_{i,k}$ is the sensitive information.

The set containing all the information transmitted in network is denoted as $\mathcal{S} = \{s_{ij,k} : (i,j)\in\mathcal{E}_k, k\in\mathbb{N}\}$, where $s_{ij,k}$ is the signal that the sensor i transmits to the sensor j at time k. Then, we introduce Fisher information as a privacy metric to quantify the privacy-preserving capability.

Definition 1 (Fisher information, [31]). Fisher information of $\mathcal{S}$ with respect to sensitive information y is defined as

$$\mathcal{I}_{\mathcal{S}}(y)=\mathbb{E}\left[\left[\frac{\partial\ln(\mathbb{P}(\mathcal{S}|y))}{\partial y}\right]\left[\frac{\partial\ln(\mathbb{P}(\mathcal{S}|y))}{\partial y}\right]^{\top}\middle|y\right].$$

Given a random variable x, the conditional Fisher information is defined as

$$\mathcal{I}_{\mathcal{S}}(y|x)=\mathbb{E}\left[\left[\frac{\partial\ln(\mathbb{P}(\mathcal{S}|x,y))}{\partial y}\right]\left[\frac{\partial\ln(\mathbb{P}(\mathcal{S}|x,y))}{\partial y}\right]^{\top}\middle|y\right].$$

Fisher information can be used to quantify the privacy-preserving capability because of the following proposition.

Proposition 1 (Cramér-Rao lower bound, [31]). If $\mathcal{I}_{\mathcal{S}}(y)$ is invertible, then for any unbiased estimator $\hat{y}=\hat{y}(\mathcal{S})$ of y, $\mathbb{E}(\hat{y}-y)(\hat{y}-y)^{\top}\geq\mathcal{I}_{\mathcal{S}}^{-1}(y)$.

By Proposition 1, smaller $\mathcal{I}_{\mathcal{S}}(y)$ implies less information leaks, and vice versa. Therefore, $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y)$ is a natural privacy metric.

Our goal is to design a privacy-preserving distributed estimation algorithm with the dynamically enhanced privacy as defined below.

Definition 2. If the privacy-preserving capability of an algorithm is said to be dynamically enhanced, then given any $i\in\mathcal{V}$ and k with $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})>0$, there exists $T>k$ such that $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,t})<\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ for all $t\geq T$.

Remark 7. By Lemma A.1 in Appendix A, $\lim_{k\to\infty}\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=0$ is a sufficient condition for the dynamically enhanced privacy.

D. Problem of interest

This paper mainly seeks to develop a new privacy-preserving distributed estimation algorithm which can simultaneously achieve

1) The privacy-preserving capability is dynamically enhanced over time;
2) The sensor i transmits only 1 bit of information to its neighbour j at each time;
3) And, the estimates for all sensors converge to the true value of the unknown parameter almost surely.

III. PRIVACY-PRESERVING ALGORITHM DESIGN

This subsection will firstly give the BQB method, and then propose BQB-PPDEA.

The traditional consensus+innovations type distributed estimation algorithms [1], [38] fuse the observations through the transmission of estimates $\hat{\theta}_{i,k-1}$, which would lead to sensitive information leakage. For the privacy issue, the following BQB method is designed to transform them into binary-valued signals before transmission. Firstly, if $k=nq+l$ for some $q\in\mathbb{N}$ and $l\in\{1,\dots,n\}$, then the sensor i generates φ_k as the n-dimensional vector whose l-th element is 1 and the others are 0. The sensor i uses φ_k to compress the previous local estimate $\hat{\theta}_{i,k-1}$ into the scalar

$$x_{i,k}=\varphi_k^{\top}\hat{\theta}_{i,k-1}. \tag{4}$$

Secondly, the sensor i generates the privacy noise $d_{ij,k}$ with distribution $F_{ij,k}(\cdot)$ for all $j\in\mathcal{N}_{i,k}$. Then, given the threshold C_{ij}, the sensor i generates the binary-valued signal

$$s_{ij,k}=\begin{cases}1, & \text{if } x_{i,k}+d_{ij,k}\leq C_{ij};\\ -1, & \text{otherwise.}\end{cases} \tag{5}$$

By using the BQB method (4)-(5), BQB-PPDEA is proposed as in Algorithm 1.

Algorithm 1 BQB-PPDEA.

```
Input: initial estimate sequence $\{\hat{\theta}_{i,0}\}$, threshold sequence $\{C_{ij}\}$ with $C_{ij}=C_{ji}$, noise distribution sequence $\{F_{ij,k}(\cdot)\}$ with $F_{ij,k}(\cdot)=F_{ji,k}(\cdot)$, step-size sequences $\{\alpha_{ij,k}\}$ with $\alpha_{ij,k}=\alpha_{ji,k}>0$ and $\{\beta_{i,k}\}$ with $\beta_{i,k}>0$.
Output: estimate sequence $\{\hat{\theta}_{i,k}\}$.
for $k=1,2,\dots,$ do
  Privacy preservation: Use the BQB method (4)-(5) to transform the previous local estimate $\hat{\theta}_{i,k-1}$ into the binary-valued signal $s_{ij,k}$, and send the binary-valued signal $s_{ij,k}$ to the neighbour $j$.
  Information fusion: Fuse the neighbourhood information.
    $\check{\theta}_{i,k}=\hat{\theta}_{i,k-1}+\varphi_k\sum_{j\in\mathcal{N}_{i,k}}\alpha_{ij,k}a_{ij,k}\left(s_{ij,k}-s_{ji,k}\right).$
  Estimate update: Use the observation $y_{i,k}$ to update the local estimate.
    $\hat{\theta}_{i,k}=\check{\theta}_{i,k}+\beta_{i,k}\bar{H}_i^{\top}\left(y_{i,k}-\bar{H}_i\hat{\theta}_{i,k-1}\right),$
  where $\bar{H}_i$ is given in Assumption 2.
end for
```

Algorithm 1 has advantages of strong privacy, low communication costs and wide applicability.

For the privacy, binary-valued quantizers improve the privacy-preserving capability of Algorithm 1. For example, in the Gaussian privacy noise case, by Proposition B.3 in Appendix B, $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})\leq\frac{2}{\pi}\mathbb{E}\mathcal{I}_{\mathcal{Z}}(y_{i,k})$, where $\mathcal{Z}=\{x_{i,k}+d_{ij,k}:i\in\mathcal{V},(i,j)\in\mathcal{E}_k,k\in\mathbb{N}\}$. This quantifies the improvement in privacy-preserving capability brought about by binary-valued quantizers.

For the communication costs, in Algorithm 1, each sensor transmits only 1 bit of information to its neighbours at each time. The data rate does not increase with the dimension of the estimates. This is mainly due to the compressing step in Algorithm 1. For comparison, [15], [17] consumes at least $\lceil\log_2(3n)\rceil$ bits to transmit an n-dimensional optimization variable, $\lceil\cdot\rceil$ is the ceiling function.

For the applicability, any *a priori* knowledge on the upper bounds of $|x_{i,k}|$ and $\|\hat{\theta}_{i,k-1}\|$ is not required in Algorithm 1 to decode binary-valued signal $s_{ij,k}$. Therefore, our BQB method is easy to implement. For comparison, such *a priori* knowledge is required in the algorithms in [16], [17].

Remark 8. The BQB method can also be applied to many other privacy-preserving problems, such as privacy-preserving distributed optimization [28], privacy-preserving distributed games [9], privacy-preserving distributed consensus [20] and so on. For example, in privacy-preserving distributed optimization [28], the decision variable x_i^k contains the sensitive information about the cost functions. Then, one can adopt the BQB method for transforming the decision variable x_i^k into

$$s'_{ij,k} = \begin{cases} 1, & \text{if } \varphi_k^\top x_i^k + d_{ij,k} \le C_{ij}; \\ -1, & \text{otherwise.} \end{cases}$$

to achieve privacy preservation, where φ_k, $d_{ij,k}$ and C_{ij} are the same as in (4) and (5).

Assumptions for the settings of Algorithm 1 are given as follows.

Assumption 4. The privacy noise sequence $\{d_{ij,k} : (i,j) \in \mathcal{E}, k \in \mathbb{N}\}$ satisfies

i) The density function $f_{ij,k}(\cdot)$ of the privacy noise $d_{ij,k}$ exists;

ii) $\eta_{ij,k} = \sup_{x\in\mathbb{R}} \frac{f_{ij,k}^2(x)}{F_{ij,k}(x)(1-F_{ij,k}(x))} < \infty$;

iii) There exists a sequence $\{\zeta_{ij,k}\}$ such that for all compact set $\mathcal{X}$, $\inf_{(i,j)\in\mathcal{E},k\in\mathbb{N},x\in\mathcal{X}} \frac{f_{ij,k}(x)}{\zeta_{ij,k}} > 0$;

iv) $\{d_{ij,k} : (i,j) \in \mathcal{E}, k \in \mathbb{N}\}$ is an independent sequence, and independent of $\{w_{i,k}, \mathcal{G}_k, H_{i,k} : i \in \mathcal{V},\ k \in \mathbb{N}\}$.

Remark 9. The Assumption 4 ii) is for the privacy analysis, and iii) is for the convergence analysis.

Remark 10. There are many types of privacy noises suitable for our BQB-PPDEA, including Guassian noises [11], [21], Laplacian noises [9], [15] and heavy-tailed Cauchy noises [40]. By Lemma 5.3 of [41] and Lemmas A.2-A.4 in Appendix A, if the privacy noise $d_{ij,k}$ obeys distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ (resp., $\mathit{Lap}(0,b_{ij,k})$, $\mathit{Cauchy}(0,r_{ij,k})$), then $\eta_{ij,k} = \frac{2}{\pi\sigma_{ij,k}^2}$ (resp., $\frac{1}{b_{ij,k}^2}$, $\frac{4}{\pi^2 r_{ij,k}^2}$), and $\zeta_{ij,k} = \frac{\sigma_{ij,1}}{\sigma_{ij,k}}$ (resp., $\frac{b_{ij,1}}{b_{ij,k}}$, $\frac{r_{ij,1}}{r_{ij,k}}$). With these noise distributions, Assumption 4 ii) and iii) can be greatly simplified. See Proposition B.1 in Appendix B.

Assumption 5. The step-size sequences $\{\alpha_{ij,k} : (i,j) \in \mathcal{E}, k \in \mathbb{N}\}$ and $\{\beta_{i,k} : i \in \mathcal{V}, k \in \mathbb{N}\}$ satisfy

i) $\sum_{k=1}^\infty \alpha_{ij,k}^2 < \infty$ and $\alpha_{ij,k} = O\left(\alpha_{ij,k+1}\right)$ for all $(i,j) \in \mathcal{E}$;

ii) $\sum_{k=1}^\infty \beta_{i,k}^2 < \infty$ and $\beta_{i,k} = O\left(\beta_{i,k+1}\right)$ for all $i \in \mathcal{V}$;

iii) $\sum_{k=1}^\infty z_k = \infty$ for $z_k = \min\{\alpha_{ij,k}\zeta_{ij,k} : (i,j) \in \mathcal{E}\} \cup \{\beta_{i,k} : i \in \mathcal{V}\}$.

Remark 11. Assumption 5 is the stochastic approximation condition for distributed estimation [26]. When the step-sizes are all polynomial, Assumption 5 iii) is equivalent to $\sum_{k=1}^\infty \alpha_{ij,k}\zeta_{ij,k} = \infty$ for all $(i,j) \in \mathcal{E}$ and $\sum_{k=1}^\infty \beta_{i,k} = \infty$ for all $i \in \mathcal{V}$. In Assumption 5, the step-sizes are not necessarily the same for all sensors, in contrast to the centralized step-sizes adopted in many distributed algorithms [1], [5], [9]. Therefore, the sensor operators can properly select their step-sizes based on their own requirements.

IV. Privacy analysis

The section will analyze the privacy-preserving capability of Algorithm 1. Theorem 1 below proves that $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k}) < \infty$. Then, Theorem 2 shows that the privacy-preserving capability of Algorithm 1 is dynamically enhanced over time.

Theorem 1. Suppose Assumptions 2, 3, 4 i), ii), iv) and 5 ii), iii) hold, and

i) $\beta_{i,k}\lambda_{\max}(Q_i) < 1$, where $Q_i = \bar{H}_i^\top \bar{H}_i$ and $\lambda_{\max}(Q_i)$ is the maximum eigenvalue of Q_i;

ii) $\sum_{t=1}^\infty \prod_{l=1}^t \eta_{ij,t}\left(1-\lambda_{\min}^+(Q_i)\beta_{i,l}\right)^2 < \infty$, where $\lambda_{\min}^+(Q_i)$ is the minimum positive eigenvalue of Q_i.

Then,

$$\begin{aligned} &\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k}) \\ &\le \sum_{j\in\mathcal{N}_i} \sum_{t=k+1}^{\infty} \beta_{i,k}^2 q_{ij,t}\eta_{ij,t} \left(\prod_{l=k+1}^{t-1} \left(1-\lambda_{\min}^+(Q_i)\beta_{i,l}\right)\right)^2 \bar{H}_i \bar{H}_i^\top \\ &< \infty. \end{aligned} \tag{6}$$

Proof. Firstly, we expand the sequence $\mathcal{S} = \{s_{ij,k} : (i,j) \in \mathcal{E}_k, k \in \mathbb{N}\}$ to $\breve{\mathcal{S}} = \{s_{ij,k} : (i,j) \in \mathcal{E}, k \in \mathbb{N}\}$. Note that we have expanded the noise sequence $\{d_{ij,k} : (i,j) \in \mathcal{E}_k, k \in \mathbb{N}\}$ to $\{d_{ij,k} : (i,j) \in \mathcal{E}, k \in \mathbb{N}\}$ in Assumption 4. Then, for all $(i,j) \in \mathcal{E}$, define

$$a'_{ij,k} = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{E}_k; \\ 0, & \text{otherwise,} \end{cases}$$

$$s'_{ij,k} = \begin{cases} 1, & \text{if } x_{i,k} + d_{ij,k} \le C_{ij}; \\ -1, & \text{otherwise.} \end{cases}$$

For $(i,j) \in \mathcal{E} \setminus \mathcal{E}_k$, define $s_{ij,k} = 0$. Then, $s_{ij,k} = a'_{ij,k} s'_{ij,k}$ and $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k}) = \mathbb{E}\mathcal{I}_{\breve{\mathcal{S}}}(y_{i,k})$.

Note that $\mathcal{I}_{\{y_{i,l}:l\neq k\}}(y_{i,k}) = 0$. Then, by Corollary A.1 in Appendix A,

$$\mathcal{I}_{\breve{\mathcal{S}}}(y_{i,k}) \le \mathcal{I}_{\breve{\mathcal{S}}}(y_{i,k}|\{y_{i,l} : l \neq k\}) \tag{7}$$

Note that $d_{uv,t}$ is independent of $\mathcal{M}_{i,t-1,k}^-$ and $y_{i,k}$, and $x_{u,t}$ is $\sigma(\mathcal{M}_{i,t-1,k}^- \cup \{y_{i,k}\})$-measurable, where $\sigma(\cdot)$ is the minimum σ-algebra containing the corresponding set, and $\mathcal{M}_{i,t,k}^- = \{y_{i,l} : l \neq k\} \cup \{s_{uv,l} : (u,v) \in \mathcal{E}, l \le t\}$. Then, given $\mathcal{M}_{i,t,k}^-$ and $y_{i,k}$, one can get $\{\breve{s}_{uv,t} : (u,v) \in \mathcal{E}\}$ is independent. Besides, given $\mathcal{M}_{i,t,k}^-$ and $y_{i,k}$, we have $s'_{uv,t}$ is uniquely determined by $d_{uv,t}$, and $a'_{uv,t}$ is uniquely determined by $\mathcal{G}_k$. Then, by Assumption 4, given $\mathcal{M}_{i,t-1,k}^-$ and $y_{i,k}$, one can get $\{s'_{uv,t} : (u,v) \in \mathcal{E}\}$ is independent of $\{a'_{uv,t} : (u,v) \in \mathcal{E}\}$. Therefore, by Lemma A.6 in Appendix A,

$$\begin{aligned} \mathcal{I}_{\breve{\mathcal{S}}}(y_{i,k}|\{y_{i,l} : l \neq k\}) &= \sum_{t=1}^{\infty} \sum_{(u,v)\in\mathcal{E}} \mathcal{I}_{s_{uv,t+1}}(y_{i,k}|\mathcal{M}_{i,t,k}^-) \\ &= \sum_{t=1}^{\infty} \sum_{j\in\mathcal{N}_i} \mathcal{I}_{s_{ij,t+1}}(y_{i,k}|\mathcal{M}_{i,t,k}^-), \end{aligned} \tag{8}$$

Denote $\bar{q}_{ij,t} = \mathbb{P}\{(i,j) \in \mathcal{E}|\mathcal{G}_{t-1}\}$, and note that $\{d_{ij,k} : k \in \mathbb{N}\}$ is independent. Then, we have

$$\begin{aligned} &\ln \mathbb{P}\left\{ s_{ij,t} \middle| y_{i,k}, \mathcal{M}_{i,t,k-1}^- \right\} \\ &= \ln\left(\bar{q}_{ij,t} F_{ij,t}(C_{ij} - x_{i,t})\right) \mathbb{I}_{\{s_{ij,t}=1\}} + \ln(1-\bar{q}_{ij,t}) \mathbb{I}_{\{s_{ij,t}=0\}} \\ &\quad + \ln\left(\bar{q}_{ij,t}\left(1 - F_{ij,t}(C_{ij} - x_{i,t})\right)\right) \mathbb{I}_{\{s_{ij,t}=-1\}}, \end{aligned}$$

which implies

$$\begin{aligned}
&\frac{\partial}{\partial y_{i,k}} \ln\left(\mathbb{P}\left\{s_{ij,t}\middle| y_{i,k}, \mathcal{M}_{i,t-1,k}^{-}\right\}\right)\\
=&\frac{\partial}{\partial y_{i,k}} \ln\left(\bar{q}_{ij,t} F_{ij,t}(C_{ij}-x_{i,t})\right) \mathbb{I}_{\{s_{ij,t}=1\}}\\
&+\frac{\partial}{\partial y_{i,k}} \ln\left(\bar{q}_{ij,t}\left(1-F_{ij,t}(C_{ij}-x_{i,t})\right)\right) \mathbb{I}_{\{s_{ij,t}=-1\}}\\
=&-\frac{f_{ij,t}\left(C_{ij}-x_{i,t}\right)}{F_{ij,t}(C_{ij}-x_{i,t})} \frac{\partial x_{i,t}}{\partial y_{i,k}} \mathbb{I}_{\{s_{ij,t}=1\}}\\
&+\frac{f_{ij,t}\left(C_{ij}-x_{i,t}\right)}{1-F_{ij,t}(C_{ij}-x_{i,t})} \frac{\partial x_{i,t}}{\partial y_{i,k}} \mathbb{I}_{\{s_{ij,t}=-1\}}.
\end{aligned} \tag{9}$$

Now, we calculate $\frac{\partial x_{i,t}}{\partial y_{i,k}}$. If $k \geq t$, then $\frac{\partial x_{i,t}}{\partial y_{i,k}}=0$. If $k<t$, then $\frac{\partial x_{i,t}}{\partial y_{i,k}}=\beta_{i,k} \bar{H}_i\left(\prod_{l=k+1}^{t-1}(I_n-\beta_{i,l} Q_i)\right)^{\top} \varphi_t$. Let $J_i=Q_i^{+} Q_i$. Then, by Lemma A.7 in Appendix A,

$$\begin{aligned}
\frac{\partial x_{i,t}}{\partial y_{i,k}} =&\beta_{i,k} \bar{H}_i J_i\left(\prod_{l=k+1}^{t-1}\left(I_n-\beta_{i,l} Q_i\right)\right)^{\top} \varphi_t\\
=&\beta_{i,k} \bar{H}_i\left(\prod_{l=k+1}^{t-1}\left(J_i-\beta_{i,l} Q_i\right)\right)^{\top} \varphi_t.
\end{aligned} \tag{10}$$

Hence, by (7)-(10) and Lemmas A.8 and A.9 in Appendix A,

$$\begin{aligned}
&\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=\mathbb{E}\mathcal{I}_{\breve{\mathcal{S}}}(y_{i,k})\\
=&\sum_{t=1}^{\infty} \sum_{j \in \mathcal{N}_i} \mathbb{E}\left[\left(\frac{\partial}{\partial y_{i,k}} \ln\left(\mathbb{P}\left\{s_{ij,t}\middle| y_{i,k}, \mathcal{M}_{i,t-1,k}^{-}\right\}\right)\right)\right.\\
&\left.\cdot\left(\frac{\partial}{\partial y_{i,k}} \ln\left(\mathbb{P}\left\{s_{ij,t}\middle| y_{i,k}, \mathcal{M}_{i,t-1,k}^{-}\right\}\right)\right)^{\top}\right]\\
=&\sum_{j \in \mathcal{N}_i} \sum_{t=k+1}^{\infty} \beta_{i,k}^2 \mathbb{E}\left[\frac{\bar{q}_{ij,t} f_{ij,t}^2\left(C_{ij}-x_{i,t}\right)}{F_{ij,t}(C_{ij}-x_{i,t})\left(1-F_{ij,t}(C_{ij}-x_{i,t})\right)}\right]\\
&\cdot \bar{H}_i\left(\prod_{l=k+1}^{t-1}\left(J_i-\beta_{i,l} Q_i\right)\right)^{\top} \varphi_t \varphi_t^{\top}\left(\prod_{l=k+1}^{t-1}\left(J_i-\beta_{i,l} Q_i\right)\right) \bar{H}_i^{\top}\\
\leq&\sum_{j \in \mathcal{N}_i} \sum_{t=k+1}^{\infty} \beta_{i,k}^2 q_{ij,t} \eta_{ij,t}\left(\prod_{l=k+1}^{t-1}\left(1-\lambda_{\min}^{+}(Q_i) \beta_{i,l}\right)\right)^2 \bar{H}_i \bar{H}_i^{\top}\\
<&\infty.
\end{aligned}$$

□

The following theorem shows that Algorithm 1 achieves the dynamically enhanced privacy.

Theorem 2. Suppose the condition of Theorem 1 holds, and

iv) $p_{u,1}=\mathbb{P}\{\mathcal{G}_1=\mathcal{G}^{(u)}\}=\pi_u$;

v) $\eta_{ij,k} \leq \frac{\eta_{ij,1}}{k^{2\epsilon_{ij}}}$ with $\eta_{ij,1}>0$ and $\epsilon_{ij} \geq 0$;

vi) $\beta_{i,k}=\frac{\beta_{i,1}}{k^{\delta_i}}$ if $k \geq k_{i,0}$; and 0, otherwise, where $\delta_i \in (1/2,1]$, $\beta_{i,1} \in (0, k_{i,0}^{\delta_i})$ and $2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}>1$;

then

$$\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k}) \leq \sum_{j \in \mathcal{N}_i} \sum_{u \in \mathbb{G}_{ij}} \pi_u R_{ij,k} \beta_{i,k} \eta_{ij,k} \bar{H}_i \bar{H}_i^{\top}=O\left(\frac{1}{k^{\delta_i+2\epsilon_{ij}}}\right), \tag{11}$$

where $q_{ij,k}$ is given in Subsection II-A, and

$$R_{ij,k}=\begin{cases}\frac{\beta_{i,1}}{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}-1} \frac{(k+1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}} k^{2\epsilon_{ij}}}{(k-1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}}}, \text{if } \delta_i=1;\\ \frac{\beta_{i,1}}{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}-(\delta_i-2\epsilon_{ij})k^{\delta_i-1}}, \quad \text{if } \delta_i \in (1/2,1).\end{cases}$$

Therefore, Algorithm 1 achieves the dynamically enhanced privacy.

Proof. If $k<k_{i,0}$, then $\beta_{i,k}=0$, which together with (6) implies $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=0$.

If $k \geq k_{i,0}$, then by Lemma A.2 of [42], one can get

$$\begin{aligned}
&\bar{H}_i\left(\prod_{l=k+1}^{t-1}\left(J_i-\beta_{i,l} Q_i\right)\right)^{\top} \varphi_t \varphi_t^{\top}\left(\prod_{l=k+1}^{t-1}\left(J_i-\beta_{i,l} Q_i\right)\right) \bar{H}_i^{\top}\\
\leq&\left(\prod_{l=k+1}^{t-1}\left(1-\frac{\lambda_{\min}^{+}(Q_i)\beta_{i,1}}{l^{\delta_i}}\right)\right)^2 \bar{H}_i \bar{H}_i^{\top}\\
\leq&\begin{cases}\left(\frac{k+1}{t-1}\right)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}} \bar{H}_i \bar{H}_i^{\top},\\ \qquad\qquad \text{if } \delta_i=1;\\ \exp\left(\frac{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}}{1-\delta_i}\left((k+1)^{1-\delta_i}-t^{1-\delta_i}\right)\right) \bar{H}_i \bar{H}_i^{\top},\\ \qquad\qquad \text{if } \delta_i<1.\end{cases}
\end{aligned} \tag{12}$$

Therefore, if $\delta_i=1$, then

$$\begin{aligned}
&\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})\\
\leq&\sum_{j \in \mathcal{N}_i} \sum_{t=k+1}^{\infty} \beta_{i,k}^2 q_{ij,t} \eta_{ij,t}\left(\frac{k+1}{t-1}\right)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}} \bar{H}_i \bar{H}_i^{\top}\\
\leq&\sum_{j \in \mathcal{N}_i} \beta_{i,1}^2 \eta_{ij,1}\left(\sum_{u \in \mathbb{G}_{ij}} \pi_u\right) \frac{(k+1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}}}{k^2}\\
&\cdot \sum_{t=k+1}^{\infty} \frac{\bar{H}_i \bar{H}_i^{\top}}{(t-1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}}}\\
\leq&\sum_{j \in \mathcal{N}_i} \beta_{i,1}^2 \eta_{ij,1}\left(\sum_{u \in \mathbb{G}_{ij}} \pi_u\right) \frac{(k+1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}}}{k^2}\\
&\cdot \frac{(k-1)^{1-2\lambda_{\min}^{+}(Q_i)\beta_{i,1}-2\epsilon_{ij}}}{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}-1} \bar{H}_i \bar{H}_i^{\top}\\
\leq&\sum_{j \in \mathcal{N}_i}\left(\sum_{u \in \mathbb{G}_{ij}} \pi_u\right) \frac{\beta_{i,1}}{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}-1}\\
&\cdot \frac{(k+1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}} k^{2\epsilon_{ij}}}{(k-1)^{2\lambda_{\min}^{+}(Q_i)\beta_{i,1}+2\epsilon_{ij}}} \beta_{i,k} \eta_{ij,k} \bar{H}_i \bar{H}_i^{\top}.
\end{aligned} \tag{13}$$

If $\delta_i<1$, then $2\lambda_{\min}^{+}(Q_i)\beta_{i,1}k^{1-\delta_i}>1-2\epsilon_{ij} \leq \delta_i-2\epsilon_{ij}$, which together with Lemma A.10 in Appendix A implies

$$\begin{aligned}
&\mathbb{E}\mathcal{I}_{\{s_{ij,t}:j\in\mathcal{N}_i,t\in\mathbb{N}\}}(y_{i,k})\\
\leq&\sum_{j\in\mathcal{N}_i}\sum_{t=k+1}^{\infty}\beta_{i,k}^2\eta_{ij,t}q_{ij,t}\frac{\exp\left(\frac{2\lambda_{\min}^+(Q_i)\beta_{i,1}}{1-\delta_i}(k+1)^{1-\delta_i}\right)}{\exp\left(\frac{2\lambda_{\min}^+(Q_i)\beta_{i,1}}{1-\delta_i}t^{1-\delta_i}\right)}\bar{H}_i\bar{H}_i^\top\\
\leq&\sum_{j\in\mathcal{N}_i}\left(\sum_{u\in\mathbb{G}_{ij}}\pi_u\right)\frac{\beta_{i,1}^2\eta_{ij,1}}{k^{2\delta_i}}\exp\left(\frac{2\lambda_{\min}^+(Q_i)\beta_{i,1}}{1-\delta_i}(k+1)^{1-\delta_i}\right)\\
&\cdot\sum_{t=k+1}^{\infty}\frac{\exp\left(-\frac{2\lambda_{\min}^+(Q_i)\beta_{i,1}}{1-\delta_i}t^{1-\delta_i}\right)}{t^{2\epsilon_{ij}}}\bar{H}_i\bar{H}_i^\top\\
\leq&\sum_{j\in\mathcal{N}_i}\left(\sum_{u\in\mathbb{G}_{ij}}\pi_u\right)\frac{\beta_{i,1}}{2\lambda_{\min}^+(Q_i)\beta_{i,1}-(\delta_i-2\epsilon_{ij})k^{\delta_i-1}}\\
&\cdot\beta_{i,k}\eta_{ij,k}\bar{H}_i\bar{H}_i^\top. \qquad (14)
\end{aligned}$$

Hence, by $\beta_{i,k}\eta_{ij,k}=O\left(\frac{1}{k^{\delta_i+2\epsilon_i}}\right)$, (11) is obtained. Then, by Proposition B.3, Algorithm 1 achieves the dynamically enhanced privacy. □

Remark 12. By (11), there is a linear relationship between the upper bound of $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ and $\sum_{j\in\mathcal{N}_i}\beta_{i,k}\eta_{ij,k}$. Therefore, the sensor i's operator can control the convergence rate of $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ by properly selecting the step-size $\beta_{i,k}$ and the privacy noise distributions. Additionally, the stationary distribution of Markovian switching graphs is also shown as a key factor affecting the privacy-preserving capability in (11).

The following corollary analyzes the privacy-preserving capability of Algorithm 1 with three types of privacy noise distributions, including Gaussian, Laplacian and Cauchy ones.

Corollary 1. Assume that Assumptions 2-5 hold. The privacy noise $d_{ij,k}$ obeys one of the following distributions

i) $\mathcal{N}(0,\sigma_{ij,k}^2)$ with $\sigma_{ij,k}=\sigma_{ij,1}k^{\epsilon_{ij}}$, $\sigma_{ij,1}>0$ and $\epsilon_{ij}\geq 0$;
ii) *Lap*$(0,b_{ij,k})$ with $b_{ij,k}=b_{ij,1}k^{\epsilon_{ij}}$, $b_{ij,1}>0$ and $\epsilon_{ij}\geq 0$;
iii) *Cauchy*$(0,r_{ij,k})$ with $r_{ij,k}=r_{ij,1}k^{\epsilon_{ij}}$, $r_{ij,1}>0$ and $\epsilon_{ij}\geq 0$.

The step-size $\beta_{i,k}$ is set as in Theorem 2. Then, $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=O\left(\frac{1}{k^{\delta_i+2\epsilon_{ij}}}\right)$.

Proof. Firstly, we prove that for all these three types of noise distributions, there exists positive sequence $\{\eta_{ij,1}\}$ such that $\eta_{ij,k}=\frac{\eta_{ij,1}}{k^{2\epsilon_{ij}}}$.

For Gaussian privacy noise case, by Lemma 5.3 of [29],

$$\begin{aligned}
\eta_{ij,k}&=\sup_{x\in\mathbb{R}}\frac{f_{ij,k}^2(x)}{F_{ij,k}(x)\left(1-F_{ij,k}(x)\right)}\\
&=\frac{f_{ij,k}^2(0)}{F_{ij,k}(0)\left(1-F_{ij,k}(0)\right)}\\
&=\frac{2}{\pi\sigma_{ij,k}^2}=\frac{2}{\pi\sigma_{ij,1}^2k^{2\epsilon_{ij}}}.
\end{aligned}$$

Set $\eta_{ij,1}=\frac{2}{\pi\sigma_{ij,1}^2}$. Then, $\eta_{ij,k}=\frac{\eta_{ij,1}}{k^{2\epsilon_{ij}}}$.

For the Laplacian privacy noise case, by Lemma A.3, $\eta_{ij,k}=\frac{1}{b_{ij,k}^2}=\frac{1}{b_{ij,1}^2k^{2\epsilon_{ij}}}$. Set $\eta_{ij,1}=\frac{1}{b_{ij,1}^2}$. Then, $\eta_{ij,k}=\frac{\eta_{ij,1}}{k^{2\epsilon_{ij}}}$.

For the Cauchy privacy noise case, by Lemma A.4, $\eta_{ij,k}=\frac{4}{\pi^2r_{ij,k}^2}=\frac{4}{\pi^2r_{ij,1}^2k^{2\epsilon_{ij}}}$. Set $\eta_{ij,1}=\frac{4}{\pi^2r_{ij,1}^2}$. Then, $\eta_{ij,k}=\frac{\eta_{ij,1}}{k^{2\epsilon_{ij}}}$.

By Theorem 1.2 of [35], we have $q_{ij,k}-\sum_{u\in\mathbb{G}_{ij}}\pi_u=O\left(\lambda_p^k\right)$ for some $\lambda_p\in(0,1)$. Hence, similar to (12)-(14), one can get $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=O\left(\frac{1}{k^{\delta_i+2\epsilon_{ij}}}\right)$. □

Remark 13. Note that the variance of Gaussian distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ is $\sigma_{ij,k}^2$, and the variance of Laplacian distribution *Lap*$(0,b_{ij,k})$ is $2b_{ij,k}^2$. Then, by Corollary 1, for these two types of noise distributions, the upper bound of $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ is inversely proportional to the noise variance. Therefore, when the noise variance grows at a faster rate, the privacy of Algorithm 1 will also be enhanced at a faster rate.

Remark 14. Cauchy distribution is heavy-tailed with infinite variance. Therefore, the variance cannot be used to describe the scale of the Cauchy privacy noise $d_{ij,k}$. Instead, the interquartile range, i.e., $\mathbb{P}\left\{\frac{1}{4}<d_{ij,k}<\frac{3}{4}\right\}$, is considered here. For Cauchy distribution *Cauchy*$(0,r_{ij,k})$, the interquartile range $\mathbb{P}\left\{\frac{1}{4}<d_{ij,k}<\frac{3}{4}\right\}=2r_{ij,k}$. Then, by Corollary 1, for Cauchy distribution, the upper bound of $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ is inversely proportional to the square of the interquartile range. Note that the variances of Gaussian distribution and Laplacian distribution are directly proportional to their interquartile ranges, respectively. Then, by Remark 13, the relationship between the privacy level $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ and the Cauchy privacy noises is similar to the Gaussian privacy noise and Laplacian privacy noise cases. Besides, [40] uses heavy-tailed noises to preserve the outlier of sensitive information. For the mechanism in [40], when the heavy-tailed privacy noise is larger, the privacy-preserving capability will be better. This is consistent with our results.

V. Convergence analysis

This section will focus on the convergence properties of Algorithm 1. Firstly, the almost sure convergence will be proved. Then, the almost sure convergence rate will be obtained.

For convenience, denote

$$\begin{aligned}
\tilde{\theta}_{i,k}&=\hat{\theta}_{i,k}-\theta,\ \tilde{\Theta}_k=\mathrm{col}\{\tilde{\theta}_{1,k},\ldots,\tilde{\theta}_{N,k}\},\ \bar{a}_{ij}=\sum_{r=1}^{M}\pi_ra_{ij}^{(r)},\\
\bar{\mathbb{H}}&=\mathrm{diag}\{\bar{H}_1^\top\bar{H}_1,\ldots,\bar{H}_N^\top\bar{H}_N\},\\
\bar{\mathbb{H}}_{\beta,k}&=\mathrm{diag}\{\beta_{1,k}\bar{H}_1^\top\bar{H}_1,\ldots,\beta_{N,k}\bar{H}_N^\top\bar{H}_N\},\\
\Phi_{i,k}&=\varphi_k\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}(a_{ij,k}-\bar{a}_{ij})\left(s_{ij,k}-s_{ji,k}\right),\\
\Phi'_{i,k}&=\varphi_k\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}\bar{a}_{ij}\left((s_{ij,k}-s_{ji,k})-2(\hat{F}_{ij,k}-\hat{F}_{ji,k})\right),\\
W_k&=\mathrm{col}\{\beta_{1,k}\left(y_{1,k}-\bar{H}_1\theta\right),\ldots,\beta_{N,k}\left(y_{N,k}-\bar{H}_N\theta\right)\},\\
&\quad+\left[\Phi_{1,k}^\top,\ldots,\Phi_{N,k}^\top\right]^\top+\left[(\Phi'_{1,k})^\top,\ldots,(\Phi'_{N,k})^\top\right]^\top,\\
\mathcal{F}_k&=\sigma(\{w_{i,t},\mathcal{G}_t,H_{i,t},d_{ij,t}:i\in\mathcal{V},(i,j)\in\mathcal{E}_t,1\leq t\leq k\}).
\end{aligned}$$

Then, $\tilde{\Theta}_k$ is $\mathcal{F}_k$-measurable.

The following theorem proves the almost sure convergence of Algorithm 1.

Theorem 3. Suppose Assumptions 1, 2, 3, 4 i), iii), iv) and 5 hold. Then, the estimate $\hat{\theta}_{i,k}$ in Algorithm 1 converges to the true value θ almost surely.

Proof. By Theorem 1.2 of [35], there exists $\lambda_a \in (0,1)$ such that $\mathbb{E}a_{ij,k} = \bar{a}_{ij} + O\left(\lambda_a^k\right)$. Then, by Assumptions 3 and 4 iv), we have $\mathbb{E}\left[a_{ij,k}s_{ij,k}|\mathcal{F}_{k-1}\right] = \bar{a}_{ij}F(C_{ij}-x_{i,k}) + O\left(\lambda_a^k\right)$. Therefore, one can get

$$\begin{aligned}
&\mathbb{E}\left[\|\tilde{\theta}_{i,k}\|^2\middle|\mathcal{F}_{k-1}\right]\\
=&\|\tilde{\theta}_{i,k-1}\|^2 - 2\beta_{i,k}\left(\bar{H}_i\tilde{\theta}_{i,k}\right)^2\\
&+ 2\varphi_k^\top\tilde{\theta}_{i,k-1}\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}\bar{a}_{ij}\left(\hat{F}_{ij,k} - \hat{F}_{ji,k}\right)\\
&+ O\left(\beta_{i,k}^2\left(\|\tilde{\theta}_{i,k-1}\|^2 + 1\right) + \sum_{j\in\mathcal{N}_i}\alpha_{ij,k}^2 + \lambda_a^k\right),
\end{aligned}$$

where $\hat{F}_{ij,k} = F_{ij,k}(C_{ij} - x_{i,k})$. Define $\tilde{x}_{i,k} = \varphi_k^\top\tilde{\theta}_{i,k-1}$. By $x_{i,k} = \varphi_k^\top\hat{\theta}_{i,k-1} = \tilde{x}_{i,k} + \varphi_k^\top\theta$, we have $x_{i,k} - x_{j,k} = \tilde{x}_{i,k} - \tilde{x}_{j,k}$. Then,

$$\begin{aligned}
&\sum_{i\in\mathcal{V}}\varphi_k^\top\tilde{\theta}_{i,k-1}\sum_{j\in\mathcal{N}_i}\alpha_{ij,k}\bar{a}_{ij}\left(\hat{F}_{ij,k} - \hat{F}_{ji,k}\right)\\
=&2\sum_{(i,j)\in\mathcal{E}}\alpha_{ij,k}\bar{a}_{ij}\left(x_{i,k} - x_{j,k}\right)\left(\hat{F}_{ij,k} - \hat{F}_{ji,k}\right) \le 0,
\end{aligned}$$

which implies

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i\in\mathcal{V}}\|\tilde{\theta}_{i,k}\|^2\middle|\mathcal{F}_{k-1}\right] \le \sum_{i\in\mathcal{V}}\|\tilde{\theta}_{i,k-1}\|^2\\
&+ O\left(\sum_{i\in\mathcal{V}}\beta_{i,k}^2\left(\|\tilde{\theta}_{i,k-1}\|^2 + 1\right) + \sum_{(i,j)\in\mathcal{E}}\alpha_{ij,k}^2 + \lambda_a^k\right).
\end{aligned}$$

Hence, by Theorem 1 of [44], $\sum_{i\in\mathcal{V}}\|\tilde{\theta}_{i,k}\|^2$ converges to a finite value almost surely. Therefore, $\tilde{\theta}_{i,k}$, $\hat{\theta}_{i,k}$, and $x_{i,k}$ are all bounded almost surely.

By the Lagrange mean value theorem [43], there exists $\xi_{ij,k}$ between $C_{ij} - x_{i,k}$ and $C_{ij} - x_{j,k}$ such that

$$\begin{aligned}
\hat{F}_{ij,k} - \hat{F}_{ji,k} =& f_{ij,k}(\xi_{ij,k})\left(x_{j,k} - x_{i,k}\right)\\
=& f_{ij,k}(\xi_{ij,k})\left(\tilde{x}_{j,k} - \tilde{x}_{i,k}\right).
\end{aligned}$$

For convenience, set $\check{f}_{ij,k} = f_{ij,k}(\xi_{ij,k})$. By the almost sure boundedness of $x_{i,k}$ and Assumption 4, there exists $\underline{f} > 0$ such that $\check{f}_{ij,k} \ge \underline{f}\zeta_{ij,k}$ almost surely.

Define $\mathcal{L}_{F,k}$ as a Laplacian matrix whose element in the i-th row and j-th column is $-\alpha_{ij,k}\bar{a}_{ij}\check{f}_{ij,k}$ if $i \neq j$, and $\sum_{l\in\mathcal{N}_i}\alpha_{1j,k}\bar{a}_{il}\check{f}_{il,k}$ if $i = j$. Then,

$$\tilde{\Theta}_k = \left(I_{N\times n} - \mathbb{H}_{\beta,k} - \mathcal{L}_{F,k}\otimes\varphi_k\varphi_k^\top\right)\tilde{\Theta}_{k-1} + W_k, \quad (15)$$

and $\mathcal{L}_{F,k} \ge z_k\underline{f}\bar{\mathcal{L}}$, where z_k is given in Assumption 5 and $\bar{\mathcal{L}} = \sum_{r=1}^M \pi_r\mathcal{L}^{(r)}$.

In addition, by Lemma 5.4 in [45], one can get

$$\begin{aligned}
&\sum_{t=k-n+1}^{k}\frac{1}{z_t}\left(\bar{\mathbb{H}}_{\beta,t} + \mathcal{L}_{F,t}\otimes\varphi_t\varphi_t^\top\right)\\
\ge&\sum_{t=k-n+1}^{k}\left(\bar{\mathbb{H}} + \underline{f}\bar{\mathcal{L}}\otimes\varphi_t\varphi_t^\top\right) \ge n\bar{\mathbb{H}} + \underline{f}\bar{\mathcal{L}}\otimes I_n > 0. \quad (16)
\end{aligned}$$

Hence, by Corollary A.2 in Appendix A, $\tilde{\Theta}_k$ and then $\tilde{\theta}_{i,k}$ converge to 0 almost surely. □

Remark 15. Note that in Algorithm 1, each sensor transmits 1 bit of information to its neighbours at each time, and as analyzed in Proposition B.2, the privacy noises are allowed to be increased. Then, by Theorem 3, the estimates of Algorithm 1 can converge to the true value θ even under 1 communication data rate and increasing privacy noises, which is the first to be achieved among existing privacy-preserving distributed algorithms [9], [11], [28].

Remark 16. In Assumption 4, the privacy noise can be heavy-tailed. Therefore, the results in Theorem 3 can also be applied to the heavy-tailed communication noise case [36], [37]. For Algorithm 1, the key to achieving convergence with heavy-tailed noises lies in the binary-valued quantizer, which transmits noisy signals with probably infinite variances to binary-valued signals with uniformly bounded variances.

Then, the following theorem calculates the almost sure convergence rate of Algorithm 1.

Theorem 4. Suppose Assumptions 1-5 hold, $\rho > 4$ and the distribution of privacy noise $d_{ij,k}$ is $\mathcal{N}(0,\sigma_{ij,k}^2)$. Given $k_{i,0}$, set $\alpha_{ij,k} = \frac{\alpha_{ij,1}}{k^{\gamma_{ij}}}$, $\beta_{i,k} = \frac{\beta_{i,1}}{k^{\delta_i}}$ if $k \ge k_{i,0}$; and 0, otherwise, and $\sigma_{ij,k} = \sigma_{ij,1}k^{\epsilon_{ij}}$, where

i) $\alpha_{ij,1} = \alpha_{ji,1} > 0$, $\sigma_{ij,1} = \sigma_{ji,1} > 0$, $\gamma_{ij} = \gamma_{ji} > \frac{1}{2}$ and $\epsilon_{ij} = \epsilon_{ji} \ge 0$ for all $(i,j)\in\mathcal{E}$, and $\beta_{i,1} > 0$ for all $i\in\mathcal{V}$;
ii) $\max_{(i,j)\in\mathcal{E}}\gamma_{ij} + \epsilon_{ij} < \min_{i\in\mathcal{V}}\delta_i \le \max_{i\in\mathcal{V}}\delta_i \le 1$.

Then, the almost sure convergence rate of the estimation error for the sensor i is

$$\tilde{\theta}_{i,k} = \begin{cases}
O\left(1\Big/k^{\frac{\lambda_H\min_{i\in\mathcal{V}}\beta_{i,1}}{N}}\right), & \\
\qquad \text{if } \bar{b} = 1,\ 2\underline{b} - \frac{2\lambda_H\min_{i\in\mathcal{V}}\beta_{i,1}}{N} > 1; & \\
O\left(\ln k\big/k^{\underline{b}-1/2}\right), & \\
\qquad \text{if } \bar{b} = 1,\ 2\underline{b} - \frac{2\lambda_H\min_{i\in\mathcal{V}}\beta_{i,1}}{N} \le 1; & \\
O\left(1\Big/k^{\underline{b}-\bar{b}/2}\right), & \\
\qquad \text{if } \bar{b} < 1, &
\end{cases} \quad \text{a.s.},$$

where $\lambda_H = \lambda_{\min}\left(\sum_{i=1}^N\bar{H}_i^\top\bar{H}_i\right)$, $\underline{b} = \min_{(i,j)\in\mathcal{E}}\gamma_{ij}$ and $\bar{b} = \max_{i\in\mathcal{V}}\delta_i$.

Proof. By Lemma A.2, $\zeta_{ij,k}$ in Assumption 4 can be $\frac{1}{k^{\epsilon_{ij}}}$. In this case, z_k in Assumption 5 is $\min\left\{\frac{\alpha_{ij,1}}{k^{\gamma_{ij}+\epsilon_{ij}}} : (i,j)\in\mathcal{E}\right\}\cup\left\{\frac{\beta_{i,1}}{k^{\delta_i}} : i\in\mathcal{V}\right\}$.

If $\bar{b} < 1$, then the theorem can be proved by (15), (16) and Corollary A.3 in Appendix A.

If $\bar{b}=1$, then $k\sum_{i=1}^N \beta_{i,k}\bar{H}_i^\top \bar{H}_i \geq \lambda_H \min_{i\in\mathcal{V}} \beta_{i,1}$. Hence, by Lemma 5.4 of [45],

$$\begin{aligned}&\frac{1}{n}\sum_{t=k-n+1}^{k} t\left(\bar{\mathbb{H}}_{\beta,t}+\mathcal{L}_{F,t}\otimes\varphi_t\varphi_t^\top\right)\\ \geq&\frac{\lambda_H \min_{i\in\mathcal{V}}\beta_{i,1}}{N}I_{nN}+O\left(\frac{1}{k^\tau}\right)\end{aligned}$$

for some $\tau>0$, which together with (15) and Corollary A.3 implies the theorem. □

Remark 17. For all $\upsilon\in(0,\frac{1}{2})$, when $\delta_i=1$, $\gamma_{ij}>\upsilon+\frac{1}{2}$ and $\beta_{i,1}$ is sufficiently large, by Theorem 4, Algorithm 1 can achieve an almost sure convergence rate of $o(1/k^\upsilon)$. The convergence rate is consistent with the classical one [38] of distributed estimation without considering the quantized communications and privacy issues.

Remark 18. By Theorems 2 and 4, the best privacy level and convergence rate will be achieved simultaneously when $\delta_i=1$.

Remark 19. If the distribution of $d_{ij,k}$ is *Lap*$(0,b_{ij,1}k^{\epsilon_{ij}})$ and *Cauchy*$(0,r_{ij,1}k^{\epsilon_{ij}})$, then by Lemma A.2, $\zeta_{ij,k}$ in Assumption 4 can also be $\frac{1}{k^{\epsilon_{ij}}}$. Therefore, the convergence rate in Theorem 4 can also be achieved in the Laplacian noise and Cauchy noise cases.

VI. Trade-off between privacy and convergence rate

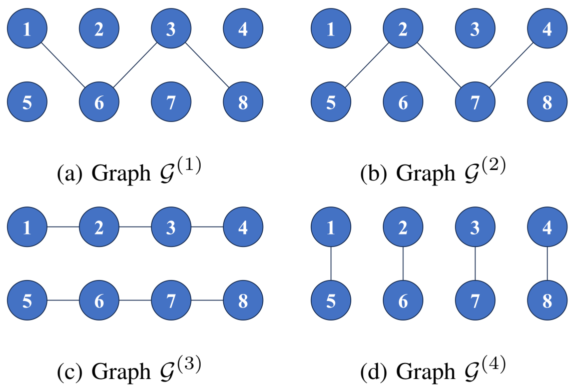
Based on the privacy and convergence analysis in Theorems 1-4, this section will establish the trade-off between the privacy level and the convergence rate of Algorithm 1.

Theorem 5. Suppose Assumptions 1-5 hold. Then, given $\nu\in(\frac{1}{2},1)$, there exist step-size sequences $\{\alpha_{ij,k}:(i,j)\in\mathcal{E}_k,k\in\mathbb{N}\}$, $\{\beta_{i,k}:i\in\mathcal{V},k\in\mathbb{N}\}$ and the privacy noise distribution sequence $\{F_{ij,k}(\cdot):(i,j)\in\mathcal{E}_k,k\in\mathbb{N}\}$ such that $\mathbb{E}\mathcal{I}_\mathcal{S}(y_{i,k})=O\left(\frac{1}{k^\chi}\right)$ and $\tilde{\theta}_{i,k}=O\left(\frac{1}{k^{\nu-\chi/2}}\right)$ almost surely for all $i\in\mathcal{V}$ and $\chi\in[1,2\nu)$.

Proof. Consider the privacy noises obeying the Gaussian distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ with $\sigma_{ij,k}=\sigma_{ij,1}k^{\epsilon_{ij}}$, $\sigma_{ij,1}=\sigma_{ji,1}>0$ and $\epsilon_{ij}=\epsilon_{ji}\geq 0$ as in Corollary 1 and Theorem 4.

Set $k_{i,0}=\exp\left(\left\lfloor\frac{1}{\delta_i}\ln\beta_{i,1}\right\rfloor+1\right)$, $\delta_i=1$, $\epsilon_{ij}=\frac{\chi-1}{2}$, $\gamma_{ij}=\frac{2+\nu-\chi}{2}$, and $\beta_{i,1}$ be any number bigger than $\frac{2-\chi}{2\lambda_{\min}^+(Q_i)}$, where $\lfloor\cdot\rfloor$ is the floor function. The step-size $\alpha_{ij,k}=\frac{\alpha_{ij,1}}{k^{\gamma_{ij}}}$, $\beta_{i,k}=\frac{\beta_{i,1}}{k^{\delta_i}}$ if $k\geq k_{i,0}$; and 0, otherwise. Then, the step-size conditions in Theorems 2 and 4 are achieved simultaneously. By Corollary 1, $\mathbb{E}\mathcal{I}_\mathcal{S}(y_{i,k})=O\left(\frac{1}{k^{\delta_i+2\epsilon_{ij}}}\right)=O\left(\frac{1}{k^\chi}\right)$. By Theorem 4, $\tilde{\theta}_{i,k}=O\left(\ln k/k^{(1+\nu-\chi)/2}\right)=O\left(\frac{1}{k^{\nu-\chi/2}}\right)$ almost surely. The theorem is proved. □

Remark 20. By Theorem 5, better privacy implies a slower convergence rate, and vice versa. The sensor operators can determine their preferences by properly selecting step-sizes and privacy noise parameters.



(a) Graph $\mathcal{G}^{(1)}$ (b) Graph $\mathcal{G}^{(2)}$

(c) Graph $\mathcal{G}^{(3)}$ (d) Graph $\mathcal{G}^{(4)}$

Fig. 1: Communication graphs

VII. Simulation

This section will demonstrate the main results of the paper by a simulation example.

Consider an 8 sensor system. The communication graph sequence $\{\mathcal{G}_k:k\in\mathbb{N}\}$ is switching among $\mathcal{G}^{(1)}$, $\mathcal{G}^{(2)}$, $\mathcal{G}^{(3)}$ and $\mathcal{G}^{(4)}$ as shown in Figure 1. For all $u=1,2,3,4$, $a_{ij}^{(u)}=1$ if $(i,j)\in\mathcal{E}^{(u)}$; and 0, otherwise. The communication graph sequence $\{\mathcal{G}_k:k\in\mathbb{N}\}$ is associated with a Markovian chain $\{m_k:k\in\mathbb{N}\}$. The initial probability $p_{u,1}=\mathbb{P}\{\mathcal{G}_1=\mathcal{G}^{(u)}\}=\frac{1}{4}$. The transition probability matrix

$$P=(p_{uv})_{4\times 4}=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}&0&0\\0&\frac{1}{2}&\frac{1}{2}&0\\0&0&\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&0&0&\frac{1}{2}\end{bmatrix},$$

where $p_{uv}=\mathbb{P}\{m_k=v|m_{k-1}=u\}$. Therefore, the stationary distribution $\pi_u=\frac{1}{4}$ for all $u=1,2,3,4$.

In the observation model, the unknown parameter $\theta=\begin{bmatrix}1&-1\end{bmatrix}^\top$. Sensors fail with probability $\frac{1}{2}$. When the sensor i does not fail at time k, the measurement matrix $H_{i,k}=\begin{bmatrix}2&0\end{bmatrix}$ if i is odd, and $\begin{bmatrix}0&2\end{bmatrix}$ if i is even. When the sensor i fails, $H_{i,k}=0$. Therefore, $\bar{H}_i=\begin{bmatrix}1&0\end{bmatrix}$ if i is odd, and $\begin{bmatrix}0&1\end{bmatrix}$ if i is even. The observation noise $w_{i,k}$ is i.i.d. Gaussian with zero mean and standard deviation 0.1.

In Algorithm 1, the threshold $C_{ij}=0$. The step-sizes $\alpha_{i,k}=\frac{3}{k^{0.8}}$, and $\beta_{i,k}=\frac{3}{k}$ if $k\geq 8$; and 0, otherwise. Three types of privacy noise distributions are considered, including Gaussian distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ with $\sigma_{ij,k}=k^{0.15}$, Laplacian distribution *Lap*$(0,b_{ij,k})$ with $b_{ij,k}=k^{0.15}$ and Cauchy distribution *Cauchy*$(0,r_{ij,k})$ with $r_{ij,k}=k^{0.15}$. Figure 2 draws the upper bounds of the non-zero elements in $\mathbb{E}\mathcal{I}_\mathcal{S}(y_{i,k})$ for each sensor i given by Theorem 2. The figure indicates that the privacy-preserving capability of Algorithm 1 is dynamically enhanced under the three types of privacy noise distributions.

Remark 21. Under our setting, $\bar{H}_i\bar{H}_i^\top=\begin{bmatrix}1&0\\0&0\end{bmatrix}$ if i is odd, and $\bar{H}_i\bar{H}_i^\top=\begin{bmatrix}0&0\\0&1\end{bmatrix}$. Then, by Theorem 2, there is only one element in the matrix $\mathcal{I}_\mathcal{S}(y_{i,k})$ is non-zero. Therefore, it is sufficient to depict the trajectory of non-zero element in the matrix $\mathbb{E}\mathcal{I}_\mathcal{S}(y_{i,k})$ in Figure 2.

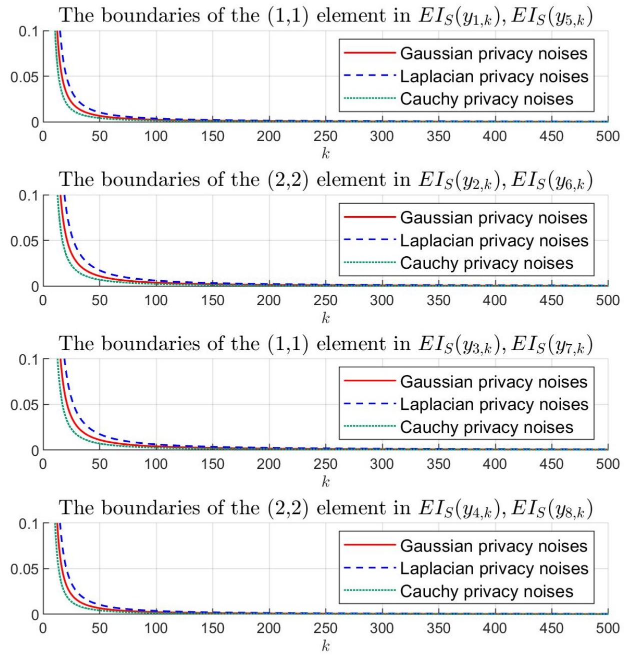


Fig. 2: The upper boundaries of the non-zero elements in $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ for each sensor i

Then, we repeat the simulation 100 times, and Figure 3 illustrates the trajectories of $\frac{1}{100N}\sum_{i=1}^{N}\sum_{\varsigma=1}^{100}\|\tilde{\theta}_{i,k}^{\varsigma}\|^2$, where $\tilde{\theta}_{i,k}^{\varsigma}$ is the estimate of θ by sensor i at time k in the ς-th run. The figure demonstrates that the estimates can converge the true value θ even under increasing noises and 1 communication data rate.

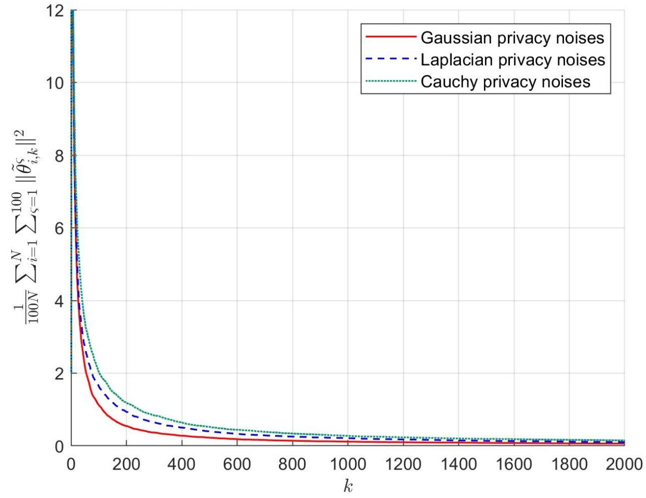


Fig. 3: The trajectories of $\frac{1}{100N}\sum_{i=1}^{N}\sum_{\varsigma=1}^{100}\|\tilde{\theta}_{i,k}^{\varsigma}\|^2$

To show the trade-off between privacy and convergence rate, in Algorithm 1, the step-size $\alpha_k = \frac{3}{k^{(2.9-\chi)/2}}$, and the privacy noises is Cauchy distributed with $r_{ij,k} = k^{\frac{\chi-1}{2}}$, where $\chi = 1.3,\ 1.6$ and 1.9. Figure 4 depicts the log-log plot for the boundaries of $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{1,k})$. It is observed that a better privacy level is achieved with a larger χ. Figure 5 shows the log-log plot for the trajectories of $\frac{1}{100N}\sum_{i=1}^{N}\sum_{\varsigma=1}^{100}\|\tilde{\theta}_{i,k}^{\varsigma}\|^2$. It is observed that a better convergence rate is achieved with a smaller χ. Therefore, Figures 4 and 5 jointly demonstrate the trade-off between privacy and convergence rate for Algorithm 1.

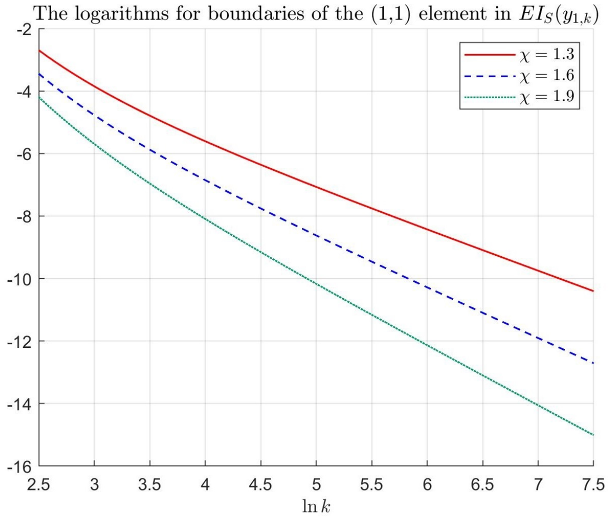


Fig. 4: The log-log plot for the boundaries of the non-zero elements in $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{1,k})$ with different χ

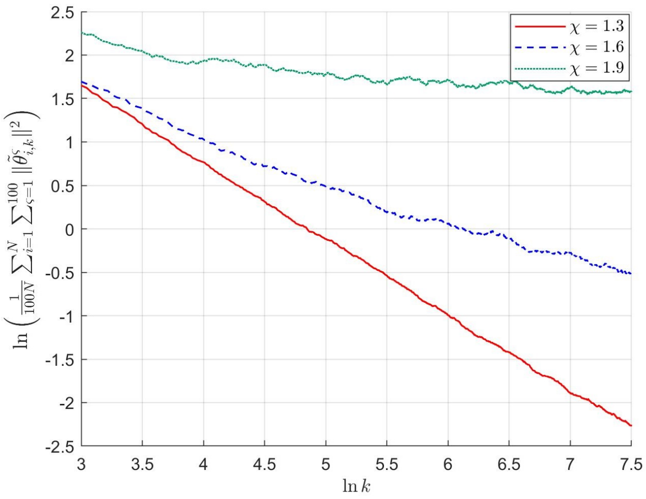


Fig. 5: The log-log plot for $\frac{1}{100N}\sum_{i=1}^{N}\sum_{\varsigma=1}^{100}\|\tilde{\theta}_{i,k}^{\varsigma}\|^2$ with different χ

VIII. Conclusion

This paper proposes BQB-PPDEA, which has multiple advantages. For the privacy, the proposed algorithm achieves the dynamically enhanced privacy, and the Fisher information-based privacy metric $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$ is proved to converge to 0 at a polynomial rate. For the communication costs, each sensor transmits only 1 bit of information to its neighbours at each time. Besides, the information receiver does not require any *a priori* knowledge on the upper bounds of the estimates' norms to decode the quantized information. For the effectiveness, the proposed algorithm can achieve almost sure convergence even with increasing privacy noises. A polynomial convergence rate is also obtained. Besides, the trade-off between privacy and convergence rate is established. When the step-sizes and privacy noise distributions are properly selected, the better privacy-preserving capability implies a slower convergence rate, and vice versa.

There are still many interesting topics deserving further investigation, For example, how to apply the proposed method to distributed optimization problems to achieve the dynamically enhanced privacy and a limited data rate, and how to extend our results to the time-varying unknown parameter case.

Appendix A
Lemmas and Corollaries

Lemma A.1. If $\lim_{k\to\infty}\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=0$, then the privacy-preserving capability is dynamically enhanced.

Proof. Since $\lim_{k\to\infty}\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})=0$, for any $A>0$, there exists $T\in\mathbb{N}$ such that $\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,t})\leq A$ for all $t\geq T$. Then, the lemma can be proved by setting $A=\mathbb{E}\mathcal{I}_{\mathcal{S}}(y_{i,k})$. □

Lemma A.2. a) If the noise $d_{ij,k}$ obeys the distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ with $\inf_k\sigma_{ij,k}>0$, then $\zeta_{ij,k}$ in iii) of Assumption 4 can be $\frac{\sigma_{ij,1}}{\sigma_{ij,k}}$;

b) If the noise $d_{ij,k}$ obeys the distribution *Lap*$(0,b_{ij,k}^2)$ with $\inf_k b_{ij,k}>0$, then $\zeta_{ij,k}$ in iii) of Assumption 4 can be $\frac{b_{ij,1}}{b_{ij,k}}$;

c) If the noise $d_{ij,k}$ obeys the distribution *Cauchy*$(0,r_{ij,k}^2)$ with $\inf_k r_{ij,k}>0$, then $\zeta_{ij,k}$ in iii) of Assumption 4 can be $\frac{r_{ij,1}}{r_{ij,k}}$;

Proof. Consider the Gaussian distribution case. Denote $f_G^\star(\cdot)$ as the density function of the standard Gaussian distribution. Then,

$$f_{ij,k}(x)=\frac{1}{\sigma_{ij,k}}f_G^\star\left(\frac{x}{\sigma_{ij,k}}\right).$$

Since $\inf_k\sigma_{ij,k}>0$, there exists a compact set $\mathcal{X}'$ such that $\frac{x}{\sigma_{ij,k}}\in\mathcal{X}'$ for all $(i,j)\in\mathcal{E}$, $k\in\mathbb{N}$ and $x\in\mathcal{X}'$. Therefore, when $\zeta_{ij,k}=\frac{1}{\sigma_{ij,k}}$,

$$\inf_{(i,j)\in\mathcal{E},k\in\mathbb{N},x\in\mathcal{X}}\frac{f_{ij,k}(x)}{\zeta_{ij,k}}\geq\inf_{z\in\mathcal{X}'}\frac{f_G^\star(z)}{\sigma_{ij,1}}>0.$$

Then, a) of Lemma A.2 holds.

The proofs of b) and c) are similar to that of a), and hence, omitted here. □

Lemma A.3. Given $b>0$, if $F_L(\cdot;b)$ and $f_L(\cdot;b)$ are the distribution function and the density function of the distribution *Lap*$(0,b)$, respectively, then

$$\sup_{x\in\mathbb{R}}\frac{f_L^2(x;b)}{F_L(x;b)(1-F_L(x;b))}=\frac{1}{b^2}.$$

Proof. Since $f_L(x;b)=\frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$, one can get $F_L(x;b)=\frac{1}{2}\exp\left(\frac{x}{b}\right)$ if $x<0$; and $1-\frac{1}{2}\exp\left(-\frac{x}{b}\right)$, otherwise.

By symmetry,

$$\sup_{x\in\mathbb{R}}\frac{f_L^2(x;b)}{F_L(x;b)(1-F_L(x;b))}=\sup_{x\geq 0}\frac{f_L^2(x;b)}{F_L(x;b)(1-F_L(x;b))}$$
$$=\sup_{x\geq 0}\frac{1}{2b^2\left(\exp\left(\frac{x}{b}\right)-\frac{1}{2}\right)}=\frac{1}{b^2}.$$

The lemma is thereby proved. □

Lemma A.4. Given $r>0$, if $F_C(\cdot;r)$ and $f_C(\cdot;r)$ are the distribution function and the density function of the distribution *Cauchy*$(0,r)$, respectively, then

$$\sup_{x\in\mathbb{R}}\frac{f_C^2(x;r)}{F_C(x;r)(1-F_C(x;r))}=\frac{4}{\pi^2r^2}.$$

Proof. Since $f_C(x;r)=\frac{1}{\pi r\left[1+(x/r)^2\right]}$, one can get $F_C(x;r)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{x}{r}\right)$.

Firstly, consider the case of $r=1$. In this case, $F_C(x;1)$ and and $f_C(\cdot;r)$ are abbreviated as $F_C(x)$ and $f_C(x)$, respectively. Denote

$$h_{C,1}(x)=\frac{F_C(x)(1-F_C(x))}{f_C^2(x)}=(1+x^2)^2\left(\frac{\pi^2}{4}-\arctan^2x\right).$$

Then, $h_{C,1}'(x)=(1+x^2)\left(\pi^2x-2\arctan x-4x\arctan^2x\right)$.

Furthermore, denote

$$h_{C,2}(x)=\pi^2x-2\arctan x-4x\arctan^2x.$$

Then, $h_{C,1}'(x)=(1+x^2)h_{C,2}(x)$, and

$$\begin{aligned}h_{C,2}'(x)=&\pi^2-\frac{2}{1+x^2}-4\arctan^2x-\frac{8x\arctan x}{1+x^2},\\ h_{C,2}''(x)=&\frac{-4x-16\arctan x}{(1+x^2)^2}.\end{aligned}$$

Note that $h_{C,2}''(x)>0$ when $x<0$; $h_{C,2}''(x)<0$ when $x>0$; and $\lim_{x\to\infty}h_{C,2}'(x)=\lim_{x\to-\infty}h_{C,2}'(x)=0$. Then, $h_{C,2}'(x)>0$, which implies that $h_{C,2}(x)$ is strictly monotonously increasing. Furthermore, by $h_{C,2}(0)=0$, we have $h_{C,2}(x)<0$ when $x<0$; and $h_{C,2}(x)>0$ when $x>0$.

Note that $h_{C,1}'(x)=(1+x^2)h_{C,2}(x)$. Then, $h_{C,1}'(x)<0$ when $x<0$; and $h_{C,1}'(x)>0$ when $x>0$. Therefore,

$$\sup_{x\in\mathbb{R}}\frac{f_C^2(x)}{F_C(x)(1-F_C(x))}=\frac{1}{\inf_{x\in\mathbb{R}}h_{C,1}(x)}=\frac{1}{h_{C,1}(0)}=\frac{4}{\pi^2}.$$

Now, consider the case of $r\neq 1$. In this case, we have

$$F_C(x;r)=F_C\left(\frac{x}{r};1\right),\quad f_C(x;r)=\frac{1}{r}f_C\left(\frac{x}{r};1\right).$$

Therefore,

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\frac{f_C^2(x;r)}{F_C(x;r)(1-F_C(x;r))}\\ =&\sup_{x\in\mathbb{R}}\frac{f_C^2\left(\frac{x}{r};1\right)}{r^2F_C\left(\frac{x}{r};1\right)\left(1-F_C\left(\frac{x}{r};1\right)\right)}=\frac{4}{\pi^2r^2}.\end{aligned}$$ □

Lemma A.5 ([31]). For random variables X,Y,ϑ, we have $\mathcal{I}_{X,Y}(\vartheta)=\mathcal{I}_X(\vartheta)+\mathcal{I}_Y(\vartheta|X)\geq\mathcal{I}_X(\vartheta)$.

Remark 22. Lemma A.5 is the chain rule for Fisher information matrices.

Corollary A.1. For random variables X,Y,Z,ϑ, we have

a) $\mathcal{I}_{X,Y}(\vartheta|Z)=\mathcal{I}_X(\vartheta|Z)+\mathcal{I}_Y(\vartheta|X,Z)$;
b) If $\mathcal{I}_Y(\vartheta|X)=0$, then $\mathcal{I}_X(\vartheta|Y)\le\mathcal{I}_{X,Y}(\vartheta)=\mathcal{I}_X(\vartheta)$;
c) If $\mathcal{I}_X(\vartheta|Z)=0$, then $\mathcal{I}_Y(\vartheta|Z)\le\mathcal{I}_Y(\vartheta|X,Z)$.

Proof. a) By Lemma A.5, we have

$$\begin{aligned}\mathcal{I}_{X,Y}(\vartheta|Z)=&\mathcal{I}_{X,Y,Z}(\vartheta)-\mathcal{I}_Z(\vartheta)\\=&\mathcal{I}_X(\vartheta|Y,Z)+\mathcal{I}_{X,Z}(\vartheta|Y)-\mathcal{I}_Z(\vartheta)\\=&\mathcal{I}_X(\vartheta|Z)+\mathcal{I}_Y(\vartheta|X,Z).\end{aligned}$$

b) By Lemma A.5, we have

$$\begin{aligned}\mathcal{I}_X(\vartheta|Y)=&\mathcal{I}_{X,Y}(\vartheta)-\mathcal{I}_Y(\vartheta)\le\mathcal{I}_{X,Y}(\vartheta)\\=&\mathcal{I}_X(\vartheta)+\mathcal{I}_Y(\vartheta|X)=\mathcal{I}_X(\vartheta).\end{aligned}$$

c) By a), we have

$$\begin{aligned}\mathcal{I}_Y(\vartheta|Z)=&\mathcal{I}_{X,Y}(\vartheta|Z)-\mathcal{I}_X(\vartheta|Y,Z)\le\mathcal{I}_{X,Y}(\vartheta|Z)\\=&\mathcal{I}_X(\vartheta|Z)+\mathcal{I}_Y(\vartheta|X,Z)=\mathcal{I}_Y(\vartheta|X,Z).\end{aligned}$$ □

Lemma A.6. For random variables X,ϑ, and random variable sequences $\mathcal{Y}_k=\{Y_{i,k}:i=1,\dots,N\}$, $\mathcal{Z}_k=\{Z_{i,k}:i=1,\dots,N\}$ for all $k\in\mathbb{N}$, if

i) $Y_{1,k},\dots,Y_{N,k}\neq0$, $Z_{1,k},\dots,Z_{N,k}\in\{0,1\}$;
ii) Given ϑ, X and $\breve{\mathcal{Z}}_{k-1}$, the sequence $\mathcal{Y}_k$ is independent, where $\breve{\mathcal{Z}}_k=\bigcup_{t=1}^k\hat{\mathcal{Z}}_t$ and $\hat{\mathcal{Z}}_k=\{Z_{i,k}Y_{i,k},i=1,\dots,N\}$;
iii) Given ϑ, X and $\breve{\mathcal{Z}}_{k-1}$, the sequence $\mathcal{Y}_k$ is independent of $\mathcal{Z}_k$;
iv) $\mathcal{I}_{\mathcal{Z}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})=0$,

then

$$\mathcal{I}_{\breve{\mathcal{Z}}_\infty}(\vartheta|X)=\sum_{k=1}^\infty\sum_{i=1}^N\mathcal{I}_{Z_{i,k}Y_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1}).$$

Proof. By Corollary A.1,

$$\mathcal{I}_{\breve{\mathcal{Z}}_\infty}(\vartheta|X)=\sum_{k=1}^\infty\mathcal{I}_{\hat{\mathcal{Z}}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1}).\tag{A.1}$$

Note that by i), we have $Z_{i,k}$ can be uniquely determined by $Z_{i,k}Y_{i,k}$. Then, by iii), given ϑ, X, $\breve{\mathcal{Z}}_{k-1}$ and $\mathcal{Z}_k$, we have $\hat{\mathcal{Z}}_k$ is independent. Hence, by Corollary A.1,

$$\begin{aligned}&\mathcal{I}_{\hat{\mathcal{Z}}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})=\mathcal{I}_{\hat{\mathcal{Z}}_k,\mathcal{Z}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})\\=&\mathcal{I}_{\hat{\mathcal{Z}}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1},\mathcal{Z}_k)+\mathcal{I}_{\mathcal{Z}_k}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})\\=&\sum_{i=1}^N\mathcal{I}_{Z_{i,k}Y_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1},\mathcal{Z}_k).\end{aligned}\tag{A.2}$$

By iv), given ϑ, X, $\breve{\mathcal{Z}}_{k-1}$ and $Z_{i,k}$, we have $Y_{i,k}$ is independent of $Z_{j,k}$ for all $j\neq i$. Therefore, by Corollary A.1,

$$\begin{aligned}&\mathcal{I}_{Z_{i,k}Y_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1},\mathcal{Z}_k)\\=&\mathcal{I}_{Z_{i,k}Y_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1},Z_{i,k})\\=&\mathcal{I}_{Z_{i,k}Y_{i,k},Z_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})-\mathcal{I}_{Z_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1})\\=&\mathcal{I}_{Z_{i,k}Y_{i,k}}(\vartheta|X,\breve{\mathcal{Z}}_{k-1}),\end{aligned}$$

which together with (A.1) and (A.2) implies the lemma. □

Lemma A.7. For a matrix H, set $Q=H^\top H$ and $J=Q^+Q$. Then, $HJ=H$.

Proof. By Theorem 1 of [46],

$$\begin{aligned}HJ=&(H^\top)^+H^\top HQ^+Q=(H^\top)^+QQ^+Q\\=&(H^\top)^+Q=(H^\top)^+H^\top H=H.\end{aligned}$$ □

Lemma A.8. For a positive semi-definite matrix Q, set $J=Q^+Q$. Then, $\lambda_{\max}(J-\beta Q)=1-\beta\lambda_{\min}^+(Q)$, where $\beta\in\left[0,\frac{1}{\lambda_{\min}^+(Q)}\right]$, and $\lambda_{\max}(\cdot)$, $\lambda_{\min}^+(\cdot)$ are defined in Theorem 1.

Proof. By Theorem 5 of [47], all the eigenvectors v for Q are eigenvectors for $J-\beta Q$. If $Qv=0$, then $(J-\beta Q)v=0$. If $Qv=\lambda v$ for some $\lambda>0$, then $(J-\beta Q)v=(1-\beta\lambda)v$. The lemma is thereby proved. □

Lemma A.9. If sequences $\{a_k:k\in\mathbb{N}\}$, $\{b_k:k\in\mathbb{N}\}$ and $\{\eta_k:k\in\mathbb{N}\}$ satisfy

i) $a_k\in[0,\bar{a}]$ for some $\bar{a}<1$;
ii) $\eta_k>0$;
iii) $\sum_{t=1}^\infty\prod_{l=1}^t\eta_t(1-a_l)^p<\infty$ for some positive integer p;
iv) $b_k>0$ and $\sum_{k=1}^\infty b_k<\infty$,

then $\sum_{t=k}^\infty\prod_{l=k}^t\eta_t(1-a_l+b_l)^p<\infty$ for all positive integer k.

Proof. Firstly, we have

$$\begin{aligned}\sum_{t=k}^\infty\prod_{l=k}^t\eta_t(1-a_l)^p=&\frac{\sum_{t=k}^\infty\prod_{l=1}^t\eta_t(1-a_l)^p}{\prod_{l=1}^{k-1}(1-a_l)^p}\\\le&\frac{\sum_{t=1}^\infty\prod_{l=1}^t\eta_t(1-a_l)^p}{\prod_{l=1}^{k-1}(1-a_l)^p}<\infty.\end{aligned}$$

Then, one can get

$$\begin{aligned}&\sum_{t=k}^\infty\prod_{l=k}^t\eta_t(1-a_l+b_l)^p\\\le&\sum_{t=k}^\infty\prod_{l=k}^t\eta_t(1-a_l)^p\left(1+\frac{b_l}{1-\bar{a}}\right)^p\\\le&\left(\sum_{t=k}^\infty\prod_{l=k}^t\eta_t(1-a_l)^p\right)\left(\prod_{t=1}^\infty\left(1+\frac{b_t}{1-\bar{a}}\right)\right)^p<\infty.\end{aligned}$$ □

Lemma A.10. If $c,k_0>0$, $g\ge0$ and $p\in(0,1]$ satisfy $cpk_0^p\ge1-p-g$, then

$$\begin{aligned}&\sum_{t=1}^k\frac{\exp\left(-c(t+k_0)^p\right)}{(t+k_0)^g}\\\le&\frac{k_0^{1-p-g}\exp(-ck_0^p)-(k+k_0)^{1-p-g}\exp(-c(k+k_0)^p)}{cp-(1-p-g)k_0^{-p}}.\end{aligned}$$

Proof. From the condition of the lemma, we have

$$\begin{aligned}&\frac{\sum_{t=1}^k\exp\left(-c(t+k_0)^p\right)}{(t+k_0)^g}\le\int_{k_0}^{k+k_0}\frac{\exp\left(-ct^p\right)}{t^g}\mathrm{d}t\\\le&\int_{k_0}^{k+k_0}\frac{cp-(1-p-g)t^{-p}}{cp-(1-p-g)k_0^{-p}}\frac{\exp\left(-ct^p\right)}{t^g}\mathrm{d}t\\=&\frac{\int_{k_0}^{k+k_0}cpt^{-g}\exp\left(-ct^p\right)-(1-p-g)t^{-p-g}\exp\left(-ct^p\right)\mathrm{d}t}{cp-(1-p-g)k_0^{-p}}\\=&\frac{k_0^{1-p-g}\exp(-ck_0^p)-(k+k_0)^{1-p-g}\exp(-c(k+k_0)^p)}{cp-(1-p-g)k_0^{-p}}.\end{aligned}$$

The lemma is thereby proved. □

Lemma A.11. Assume that

i) $\{\alpha_k : k \in \mathbb{N}\}$, $\{\beta_k : k \in \mathbb{N}\}$ and $\{\gamma_k : k \in \mathbb{N}\}$ are positive sequences satisfying $\sum_{k=1}^{\infty} \alpha_k = \infty$, $\sum_{k=1}^{\infty} \beta_k^2 < \infty$ and $\sum_{k=1}^{\infty} \gamma_k^2 < \infty$;

ii) $\{\mathcal{F}_k : k \in \mathbb{N}\}$ is a σ-algebra sequence with $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$ for all k;

iii) $\{W_k, \mathcal{F}_k : k \in \mathbb{N}\}$ is a sequence of adaptive random variables satisfying $\sum_{k=1}^{\infty} \|\mathbb{E}[W_k|\mathcal{F}_{k-1}]\| < \infty$ and $\mathbb{E}[\|W_k - \mathbb{E}[W_k|\mathcal{F}_{k-1}]\|^{\rho}|\mathcal{F}_{k-1}] = O(\beta_k^{\rho})$ almost surely for some $\rho > 2$;

iv) $\{U_k : k \in \mathbb{N}\}$ is a sequence with $\sum_{k=1}^{\infty} \alpha_k^2 \|U_k\|^2 < \infty$. And, U_k is $\mathcal{F}_{k-1}$-measurable;

v) $U_k + U_k^{\top} \geq 2aI_n$ for some $p \in \mathbb{N}$, $a > 0$ and all $k \in \mathbb{N}$ almost surely;

vi) $\{X_k, \mathcal{F}_k : k \in \mathbb{N}\}$ is a sequence of adaptive random variables with

$$X_k = (I_n - \alpha_k U_k + O(\gamma_k)) X_{k-1} + W_k, \text{ a.s.} \tag{A.3}$$

Then, X_k converges to 0 almost surely.

Proof. Consider $X'_k = X_k - Y_k$, where $Y_0 = 0$ and $Y_k = (I_n - \alpha_k U_k + O(\gamma_k)) Y_{k-1} + \mathbb{E}[W_k|\mathcal{F}_{k-1}]$. Since $\|Y_k\| \leq (1 - a\alpha_k + O(\alpha_k^2 \|U_k\|^2 + \gamma_k)) \|Y_{k-1}\| + \|\mathbb{E}[W_k|\mathcal{F}_{k-1}]\|$, by Lemma 2 of [28], Y_k converges to 0. Therefore, it suffices to prove the convergence of X'_k, which satisfies

$$X'_k = (I_n - \alpha_k U_k + O(\gamma_k)) X'_{k-1} + W_k - \mathbb{E}[W_k|\mathcal{F}_{k-1}]. \tag{A.4}$$

By (A.4), one can get

$$\begin{aligned}&\mathbb{E}\left[\|X'_k\|^2 \middle| \mathcal{F}_{k-1}\right] \\ \leq &\left(1 - 2a\alpha_k + \alpha_k^2 \|U_k\|^2 + O(\gamma_k)\right) \|X'_{k-1}\|^2 + O\left(\beta_k^2\right).\end{aligned} \tag{A.5}$$

Then, by Lemma 2 of [28], we have $\lim_{k\to\infty} \|X_k\|^2 = 0$ almost surely. □

Corollary A.2. If i)-iv) and vi) in Lemma A.11 hold, $\alpha_k = O(\alpha_{k-1})$ and

$$\frac{1}{p} \sum_{t=k-p+1}^{k} \left(U_t + U_t^{\top}\right) \geq 2aI_n \tag{A.6}$$

for some $p \in \mathbb{N}$, $a > 0$ and all $k \in \mathbb{N}$ almost surely, then X_k converges to 0 almost surely.

Proof. By (A.3), one can get

$$\begin{aligned}X_k = &\prod_{t=k-p+1}^{k} (I_n - \alpha_t U_t + O(\gamma_t)) X_{k-p} \\ &+ \sum_{t=k-p+1}^{k} \prod_{l=t+1}^{k} (I_n - \alpha_l U_l) W_t.\end{aligned}$$

In the above recursive function,

$$\begin{aligned}&\prod_{t=k-p+1}^{k} (I_n - \alpha_t U_t + O(\gamma_t)) \\ =&I_n - \sum_{t=k-p+1}^{k} \alpha_t U_t + O\left(\sum_{t=k-p+1}^{k} \left(\gamma_t + \alpha_k^2 \|U_k\|^2\right)\right),\end{aligned}$$

Note that by (A.6),

$$\begin{aligned}&\frac{1}{p} \sum_{t=k-p+1}^{k} \alpha_t \left(U_t + U_t^{\top}\right) \\ \geq &\left(\min_{k-p<t\leq k} \alpha_t\right) \frac{1}{p} \sum_{t=k-p+1}^{k} \left(U_t + U_t^{\top}\right) \\ \geq &2a \left(\min_{k-p<t\leq k} \alpha_t\right) I_n,\end{aligned}$$

and by Lemma A.2 of [26], $\sum_{k=p+1}^{\infty} \min_{k-p<t\leq k} \alpha_t = \infty$. Then, the corollary can be proved by Lemma A.11. □

Lemma A.12. If an adaptive sequence $\{V_k, \mathcal{F}_k : k \in \mathbb{N}\}$ satisfies

$$\mathbb{E}[V_k|\mathcal{F}_{k-1}] \leq \left(1 - \frac{a}{k} + \gamma_k\right) V_{k-1} + O\left(\frac{1}{k^b}\right)$$

with $a > 0$, $b > 1$ and $\sum_{k=1}^{\infty} \gamma_k < \infty$, then

$$V_k = \begin{cases} O\left(\frac{1}{k^a}\right), & \text{if } b - a > 1; \\ O\left(\frac{(\ln k)^2}{k^{b-1}}\right), & \text{if } b - a \leq 1. \end{cases}$$

Proof. If $b - a > 1$, then

$$\begin{aligned}&\mathbb{E}[k^a V_k|\mathcal{F}_{k-1}] \\ \leq &\left(1 - \frac{a}{k} + \gamma_k\right) \left(1 + \frac{a}{k} + O\left(\frac{1}{k^2}\right)\right) (k-1)^a V_{k-1} \\ &+ O\left(\frac{1}{k^{b-a}}\right) \\ \leq &\left(1 + \gamma_k + O\left(\frac{1}{k^2}\right)\right) (k-1)^a V_{k-1} + O\left(\frac{1}{k^{b-a}}\right),\end{aligned}$$

which together with Theorem 1 of [44] implies that $V_k = O\left(\frac{1}{k^a}\right)$ almost surely.

If $b - a \leq 1$, then

$$\begin{aligned}&\mathbb{E}\left[\frac{k^{b-1}}{(\ln k)^2} V_k \middle| \mathcal{F}_{k-1}\right] \\ \leq &\left(1 - \frac{a}{k} + \gamma_k\right) \left(1 + \frac{b-1}{k} + O\left(\frac{1}{k^2}\right)\right) \frac{(b-1)^{b-1}}{(\ln(k-1))^2} V_{k-1} \\ &+ O\left(\frac{1}{k(\ln k)^2}\right) \\ \leq &\left(1 + \gamma_k + O\left(\frac{1}{k^2}\right)\right) \frac{(b-1)^{b-1}}{(\ln(k-1))^2} V_{k-1} + O\left(\frac{1}{k(\ln k)^2}\right),\end{aligned}$$

which together with Theorem 1 of [44] implies that $V_k = O\left(\frac{(\ln k)^2}{k^{b-1}}\right)$ almost surely. The lemma is thereby proved. □

Lemma A.13. If sequences $\{V_k : k \in \mathbb{N}\}$, $\{\xi_k : k \in \mathbb{N}\}$, $\{\eta_k : k \in \mathbb{N}\}$ and $\{\gamma_k : k \in \mathbb{N}\}$ satisfy

i) $\xi_k \geq 0$, $\overline{\lim}_{k\to\infty} \xi_k < 1$;

ii) $\sum_{k=1}^{\infty} \eta_k < \infty$, $\sum_{k=1}^{\infty} |\gamma_k| < \infty$;
iii) $V_k \leq (1 - \xi_k + \gamma_k) V_{k-1} + \eta_k + O(\xi_k)$,

then V_k is uniformly upper bounded.

Proof. Without loss of generality, assume $\gamma_k \geq 0$. Besides, by $\sum_{k=1}^{\infty} \gamma_k < \infty$, there exists k_0 such that $\gamma_k < \frac{1}{3}$ and $\xi_k < 1 + \gamma_k$ for all $k \geq k_0$. Set $U_k = \prod_{t=k_0}^{k} \left(1 - \gamma_t - \frac{|\gamma_t|}{2}\right) \left(V_k - \sum_{t=k_0}^{k} \eta_t\right)$. Then, there exists $M > 0$ such that

$$\begin{aligned}
U_k &= \prod_{t=k_0}^{k} \left(1 - \gamma_t - \frac{|\gamma_t|}{2}\right) \left(V_k - \sum_{t=k_0}^{k} \eta_t\right) \\
&\leq \prod_{t=k_0}^{k} \left(1 - \gamma_t - \frac{|\gamma_t|}{2}\right) \\
&\quad \cdot \left((1 - \xi_k + \gamma_k) \left(V_{k-1} - \sum_{t=k_0}^{k-1} \eta_t\right) + O(\xi_k + |\gamma_k|)\right) \\
&= \left(1 - \gamma_k - \frac{|\gamma_k|}{2}\right) (1 - \xi_k + \gamma_k) U_{k-1} + M(\xi_k + |\gamma_k|).
\end{aligned}$$

If $U_{k-1} < 2M$, then

$$\begin{aligned}
U_k &< \left(1 - \gamma_k - \frac{|\gamma_k|}{2}\right) (1 - \xi_k + \gamma_k) 2M + M(\xi_k + |\gamma_k|) \\
&\leq \left(1 - \frac{1}{2}(\xi_k + |\gamma_k|)\right) 2M + M(\xi_k + |\gamma_k|) \leq 2M.
\end{aligned}$$

If $U_{k-1} \geq 2M$, then

$$U_k \leq \left(1 - \frac{1}{2}(\xi_k + |\gamma_k|)\right) U_{k-1} + M(\xi_k + |\gamma_k|) \leq U_{k-1}.$$

Therefore, $U_k \leq \max\{U_{k-1}, 2M\}$, which implies the uniformly boundedness of U_k and further V_k. □

Lemma A.14. If i)-vi) in Lemma A.11 hold, $\rho > 4$, $\alpha_k = \frac{1}{k^c}$, $\beta_k = \frac{1}{k^b}$ for $c \in (\frac{1}{2}, 1]$ and $b > 1$, and $\|\mathbb{E}[W_k | \mathcal{F}_{k-1}]\| \leq \lambda^k$ for some $\lambda \in (0, 1)$, then

$$X_k = \begin{cases} O\left(\frac{1}{k^a}\right), & \text{if } c = 1, 2b - 2a > 1; \\ O\left(\frac{\ln k}{k^{b-1/2}}\right), & \text{if } c = 1, 2b - 2a \leq 1; \\ O\left(\frac{1}{k^{b-c/2}}\right), & \text{if } c \in (\frac{1}{2}, 1), \end{cases} \text{ a.s.} \quad \text{(A.7)}$$

Proof. Consider X'_k and Y_k in the proof of Lemma A.11. One can get

$$\begin{aligned}
&\frac{\|Y_k\|}{\prod_{t=1}^{k} \left(1 - \frac{a}{k^c}\right)} \\
\leq &(1 + O(\alpha_k^2 \|U_k\|^2 + \gamma_k)) \frac{\|Y_{k-1}\|}{\prod_{t=1}^{k-1} \left(1 - \frac{a}{k^c}\right)} + \frac{\lambda^k}{\prod_{t=1}^{k} \left(1 - \frac{a}{k^c}\right)}.
\end{aligned}$$

By Lemma A.2 of [42] and Lemma A.13,

$$Y_k = \begin{cases} O\left(\frac{1}{k^a}\right), & \text{if } c = 1; \\ O\left(\exp\left(\frac{a}{1-c} k^{1-c}\right)\right), & \text{if } c \in (\frac{1}{2}, 1). \end{cases} \quad \text{(A.8)}$$

Therefore, it suffices to calculate the convergence rate of X'_k.

If $c = 1$, then the lemma can be proved by (A.5) and Lemma A.12. Then, it suffices to analyze the case of $c < 1$.

For convenience, denote $\tilde{W}_k = W_k - \mathbb{E}[W_k | \mathcal{F}_{k-1}]$. By (A.4), one can get

$$\begin{aligned}
&k^{2b-c} \|X'_k\|^2 \\
\leq &\left(1 - \frac{a}{k^c} + \alpha_k^2 \|U_k\|^2 + O(\gamma_k)\right) (k-1)^{2b-c} \|X'_{k-1}\|^2 \\
&+ 2k^{2b-c} \tilde{W}_k^{\top} (I_n - \alpha_k U_k + O(\gamma_k)) X'_{k-1} \\
&+ k^{2b-c} \|\tilde{W}_k\|^2, \text{ a.s.} \quad \text{(A.9)}
\end{aligned}$$

Then, by Lemma 2 of [48],

$$\begin{aligned}
&\sum_{t=1}^{k} 2t^{2b-c} \tilde{W}_t^{\top} (I_n - \alpha_t U_t + O(\gamma_t)) X'_{t-1} \\
\leq &\sum_{t=1}^{k} (2t^b \tilde{W}_t)^{\top} \left(t^{b-c} (I_n - \alpha_t U_t + O(\gamma_t)) X'_{t-1}\right) \\
= &O(1) + o\left(\sum_{t=1}^{k} t^{2b-2c} \|X'_{t-1}\|^2\right), \text{ a.s.,} \quad \text{(A.10)}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{t=1}^{k} t^{2b-c} \left(\|\tilde{W}_t\|^2 - \mathbb{E}\left[\|\tilde{W}_t\|^2 \middle| \mathcal{F}_{k-1}\right]\right) \\
\leq &\sum_{t=1}^{k} t^{2b} \left(\|\tilde{W}_t\|^2 - \mathbb{E}\left[\|\tilde{W}_t\|^2 \middle| \mathcal{F}_{k-1}\right]\right) \cdot \frac{1}{k^c} = O(1), \text{ a.s.}
\end{aligned}$$

Therefore, $X_k = O\left(\frac{1}{k^b}\right)$ almost surely, which together with (A.10) implies

$$\sum_{t=1}^{k} 2t^{2b-c} \tilde{W}_t^{\top} (I_n - \alpha_t U_t + O(\gamma_t)) X'_{t-1} = O(1).$$

Then, the lemma can be proved by (A.8), (A.9) and Lemma A.13. □

Corollary A.3. Suppose i)-iv), vi) in Lemma A.11 and (A.6) in Corollary A.2 hold. ρ, α_k and β_k are set as in Lemma A.14. Then, X_k can achieve the almost sure convergence rate as in (A.7).

The proof of Corollary A.3 is similar to Corollary A.2, and thereby omitted here.

Appendix B
Examples for privacy noises

The following two propositions gives sufficient conditions on privacy noises satisfying Assumptions 4 and 5, when the privacy noises are Gaussian, Laplacian and Cauchy.

Proposition B.1. For the noise distribution $\mathcal{N}(0, \sigma_{ij,k}^2)$ (resp., *Lap*$(0, b_{ij,k})$, *Cauchy*$(0, r_{ij,k})$), Assumption 4 ii) holds when $\sigma_{ij,k} > 0$ (resp., $b_{ij,k} > 0$, $r_{ij,k} > 0$), and Assumption 4 iii) holds when $\inf_{k \in \mathbb{N}} \sigma_{ij,k} > 0$ (resp., $\inf_{k \in \mathbb{N}} b_{ij,k} > 0$, $\inf_{k \in \mathbb{N}} r_{ij,k} > 0$).

Proof. Consider the Gaussian noise case. By Lemma 5.3 of [41], $\eta_{ij,k} = \frac{2}{\pi \sigma_{ij,k}^2}$. Therefore, when $\sigma_{ij,k} > 0$, $\eta_{ij,k} < \infty$. Besides, Lemma A.2 in Appendix A implies that $\inf_{k \in \mathbb{N}} \sigma_{ij,k} > 0$ is sufficient to achieve Assumption 4 ii).

The analysis for the Laplacian and Cauchy noise cases is similar, and thereby omitted here. □

Remark B.1. By Proposition B.1, for Gaussian and Laplacian privacy noises, Assumption 4 ii) and iii) can be replaced with the condition that there is a uniform positive lower bound of the noise variances. The reasons to adopt the assumption are twofold. For the privacy, sufficient privacy noises can ensure the privacy-preserving capability of the algorithm. For the effectiveness, the privacy noises are also necessary dithered noises in the quantizers [26]. The lack of sufficient dithered noises in the quantizers will result in the algorithm failing to converge.

Proposition B.2. For the noise distribution $\mathcal{N}(0,\sigma_{ij,k}^2)$ (resp., *Lap*$(0,b_{ij,k})$, *Cauchy*$(0,r_{ij,k})$) with $\sigma_{ij,k}=\sigma_{ij,1}k^{\epsilon_{ij}}$ (resp., $b_{ij,k}=b_{ij,1}k^{\epsilon_{ij}}$, $r_{ij,k}=r_{ij,1}k^{\epsilon_{ij}}$) and $\zeta_{ij,k}=k^{-\epsilon_{ij}}$, there exists step-size sequences $\{\alpha_{ij,k}:(i,j)\in\mathcal{E},k\in\mathbb{N}\}$ and $\{\beta_{i,k}:i\in\mathcal{V},k\in\mathbb{N}\}$ satisfying Assumption 5 if $\epsilon_{ij}<\frac{1}{2}$.

Proof. Set $\alpha_{ij,k}=\frac{\alpha_{ij,1}}{k^{\gamma_{ij}}}$ and $\beta_{i,k}=\frac{\beta_{i,1}}{k^{\delta_i}}$ with $\gamma_{ij}>\frac{1}{2}$ and $\gamma_{ij}+\epsilon_{ij}<\delta_i\le 1$. Then, Assumption 5 holds. □

Remark 23. $\zeta_{ij,k}$ in Proposition B.2 is the consistent with the one given in Lemma A.2.

The following proposition takes Gaussian noise for example to analyze the improvement of privacy-preserving capability using BQB method compared to unquantized methods.

Proposition B.3. Assume that random variables z_k,d_k,l_k, random variables y,ϕ_k and real number C_k satisfy

i) $z_k=\phi_k^\top y+d_k+l_k$, $s_k=2\mathbb{I}_{\{z_k\le C_k\}}-1$;
ii) Given $s_1,\dots,s_{k-1}$, ϕ_k and l_k are independent of y;
iii) d_k is Gaussian distributed with zero mean, and independent of $d_1,\dots,d_{k-1},y,\phi_k,l_k$.

Then, $\mathbb{E}\mathcal{I}_{s_{1:k}}(y)\le\frac{2}{\pi}\mathbb{E}\mathcal{I}_{z_{1:k}}(y)$, where $s_{1:k}=\{s_1,\dots,s_k\}$ and $z_{1:k}=\{z_1,\dots,z_k\}$

Proof. By Lemma A.5, $\mathbb{E}\mathcal{I}_{s_{1:k}}(y)=\sum_{t=1}^k\mathbb{E}\mathcal{I}_{s_t}(y|s_{1:t-1})$ and $\mathbb{E}\mathcal{I}_{z_{1:k}}(y)=\sum_{t=1}^k\mathbb{E}\mathcal{I}_{z_t}(y|z_{1:t-1})$. Then, by Proposition 1 of [22] and Lemma 5.3 of [41], $\mathbb{E}\mathcal{I}_{s_t}(y|s_{1:t-1})\le\frac{2}{\pi}\mathbb{E}\mathcal{I}_{z_t}(y|z_{1:t-1})$. The proposition is thereby proved. □

References

[1] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[2] A. H. Sayed, S. Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior," *IEEE Signal Proc. Mag.*, vol. 30, no. 3, pp. 155-171, 2013.

[3] C. Li, P. Zhou, L. Xiong, Q. Wang, and T. Wang, "Differentially private distributed online learning," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 8, pp. 1440-1453, 2018.

[4] A. Smith, "Privacy-preserving statistical estimation with optimal convergence rates," in *Proc. 43rd Annu. ACM Symp. Theory Comput. (STOC)*, San Jose, CA, USA, pp. 813-821, Jun. 6-8, 2011.

[5] Y. Lu and M. H. Zhu, "Privacy preserving distributed optimization using homomorphic encryption," *Automatica*, vol. 96, pp. 314-325, 2018.

[6] C. N. Hadjicostis and A. D. Domínguez-García, "Privacy-preserving distributed averaging via homomorphically encrypted ratio consensus," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3887–3894, 2020

[7] C. Regueiro, I. Seco, S. de Diego, O. Lage, and L. Etxebarria, "Privacy-enhancing distributed protocol for data aggregation based on blockchain and homomorphic encryption," *Inf. Process. Manage.*, vol. 58, no. 6, 2021, Art. no. 102745.

[8] X. Li, N. Wang, L. Zhu, S. Yuan, and Z. Guan, "FUSE: a federated learning and U-shape split learning-based electricity theft detection framework," *Sci. China Inf. Sci.*, vol. 67, no. 4, 2024, Art. no. 149302.

[9] M. J. Ye, G. Q. Hu, L. H. Xie, and S. Y. Xu, "Differentially private distributed Nash equilibrium seeking for aggregative games," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2451-2458, 2021.

[10] F. Farokhi and H. Sandberg, "Ensuring privacy with constrained additive noise by minimizing Fisher information," *Automatica*, vol. 99, pp. 275-288, 2019.

[11] C. Gratton, N. K. D. Venkategowda, R. Arablouei, and S. Werner, "Privacy-preserved distributed learning with zeroth-order optimization," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 265-279, 2021.

[12] K. Wei, J. Li, C. Ma, M. Ding, F. Shu, H. T. Zhao, W. Chen, and H. B. Zhu, "Gradient sparsification for efficient wireless federated learning with differential privacy," *Sci. China Inf. Sci.*, vol. 67, no. 4, 2024, Art. no. 142303.

[13] Y. Q. Wang, "Privacy-preserving average consensus via state decomposition," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4711-4716, 2019.

[14] C. Altafini, "A system-theoretic framework for privacy preservation in continuous-time multiagent dynamics," *Automatica*, vol. 122, 2020, Art. no. 109253.

[15] N. Lang, E. Sofer, T. Shaked, and N. Shlezinger, "Joint privacy enhancement and quantization in federated learning," *IEEE Trans. Signal Process.*, vol. 71, pp. 295-310, 2023.

[16] W. Chen, L. Liu, and G. P. Liu, "Privacy-preserving distributed economic dispatch of microgrids: A dynamic quantization-based consensus scheme with homomorphic encryption," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 701-713, 2023.

[17] Y. Q. Wang and T. Başar, "Quantization enabled privacy protection in decentralized stochastic optimization," *IEEE Trans. Autom. Control*, vol. 68, no. 7, pp. 4038-4052, 2023.

[18] L. Liu, Y. Kawano, and M. Cao, "Design of stochastic quantizers for privacy preservation," 2024, arXiv:2403.03048.

[19] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. Commun. Technol.*, vol. 12, no. 4, pp. 162–165, 1964.

[20] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private average consensus: obstructions, trade-offs, and optimal algorithm design," *Automatica*, vol. 81, no. 7, pp. 221–231, 2017.

[21] J. M. Ke, J. M. Wang, and J. F. Zhang, "Differentiated output-based privacy-preserving average consensus," *IEEE Control Syst. Lett.*, vol. 7, pp. 1369-1347, 2023.

[22] Y. Wang, Y. L. Zhao, and J. F. Zhang, "Asymptotically efficient Quasi-Newton type identification with quantized observations under bounded persistent excitations," *Automatica*, vol. 166, 2024, Art. no. 111722.

[23] Z. Q. Chen and Y. Q. Wang, "Locally differentially private distributed online learning with guaranteed optimality," 2023, arXiv:2306.14094.

[24] B. Jayaraman, L. X. Wang, D. Evans, and Q. Q. Gu, "Distributed learning without distress: Privacy-preserving empirical risk minimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, Montreal, QC, Canada, pp. 1–12, Dec. 2018.

[25] J. M. Wang, J. F. Zhang, and X. K. Liu, "Differentially private resilient distributed cooperative online estimation over digraphs," *Int. J. Robust Nonlin. Control*, vol. 32, no. 15, pp. 8670-8688, 2022.

[26] J. M. Ke, X. D. Lu, Y. L. Zhao, and J. F. Zhang, "Signal-comparison-based distributed estimation under decaying average bit rate communications," 2024, arXiv:2405.18694.

[27] Y. Liu, J. Liu, and T. Başar, "Differentially private gossip gradient descent," in *Proc. 57th IEEE Conf. Decis. Control*, Miami, FL, USA, pp. 2777-2782, Dec. 17-19, 2018.

[28] Y. Q. Wang and A. Nedić, "Tailoring gradient methods for differentially private distributed optimization," *IEEE Trans. Autom. Control*, vol. 29, no. 2, pp. 872 - 887, 2024.

[29] J. M. Wang, J. W. Tan, and J. F. Zhang, "Differentially private distributed parameter estimation," *J. Syst. Sci. Complex.* vol. 36, no. 1, pp. 187-204, 2023.

[30] Q. Zhang and J. F. Zhang, "Distributed parameter estimation over unreliable networks with Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2545-2560, 2012.

[31] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246-1250, 1998.

[32] R. Nassif, S. Vlaski, M. Carpentiero, V. Matta, M. Antonini, and A. H. Sayed "Quantization for decentralized learning under subspace constraints," *IEEE Trans. Signal Process.*, vol. 2320-2335, 2023.

[33] M. Carpentiero, V. Matta, and A. H. Sayed, "Distributed adaptive learning under communication constraints," *IEEE Open J. Signal Process.*, vol. 5, pp. 321-358, 2023.

[34] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355-369, 2009.

[35] E. Seneta, *Non-negative Matrices and Markov Chains (2nd ed.)*, Springer: NY, USA, 2006.

[36] D. Jakovetic, M. Vukovic, D. Bajovic, A. K. Sahu, and S. Kar, "Distributed recursive estimation under heavy-tail communication noise," *SIAM J. Control Optim.*, vol. 61, no. 3, 1582-1609, 2023.

[37] M. Vukovic, D. Jakovetic, D. Bajovic, and S. Kar, "Nonlinear consensus+innovations under correlated heavy-tailed noises: Mean square convergence rate and asymptotics," *SIAM J. Control Optim.*, vol. 62, no. 1, pp. 376-399, 2024.

[38] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topic Signal Process.*, vol. 5, no. 4, pp. 674-690, 2011.

[39] A. N. Shiryaev, *Probability (2nd ed.)*, Springer: NY, USA, 1996.

[40] K. Ito, Y. Kawano, and K. Kashima, "Privacy protection with heavy-tailed noise for linear dynamical systems," *Automatica*, vol. 131, 2021, Art. no. 109732.

[41] Y. Wang, X. Li, Y. L. Zhao, and J. F. Zhang, "Threshold selection and resource allocation for quantized identification," *J. Syst. Sci. Complex.*, vol. 37, no. 1, pp. 204-229, 2024.

[42] J. M. Wang, J. M. Ke and J. F. Zhang, "Differentially private bipartite consensus over signed networks with time-varying noises," *IEEE Trans. Automa. Control*, 2024, doi: 10.1109/TAC.2024.3351869.

[43] V. A. Zorich, *Mathematical Analysis I (2nd ed.)*, Springer: NY, USA, 2015.

[44] H. Robbins and D. Siegmund, "A convergence theorem for non negative almost supermartingales and some applications," in *Optimizing Methods in Statistics*, Academic Press: New York, pp. 233–257, 1971.

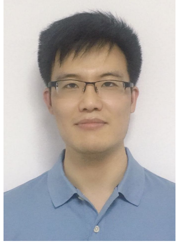
[45] S. Y. Xie and L. Guo, "Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition," *SIAM J. Control Optim.*, vol. 56, pp. 3404-3431, 2018.

[46] T. N. E. Greville, "Note on the generalized inverse of a matrix product," *SIAM Rev.*, vol. 8, no. 4, pp. 518-521, 1966.

[47] J. E. Scroggs and P. L. Odell, "An alternate definition of a pseudoinverse of a matrix," *SIAM J. Appl. Math.*, vol. 14, no. 4, pp. 796-810, 1966.

[48] C. Z. Wei, "Asymptotic properties of least-squares estimates in stochastic regression models," *Ann. Statist.*, vol. 13, pp. 1498–1508, 1985.

Jieming Ke (IEEE Student Member) received the B.S. degree in Mathematics from University of Chinese Academy of Science, Beijing, China, in 2020. He is currently working toward the Ph.D. degree majoring in system theory at Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), Beijing, China. His research interests include privacy and security in stochastic systems, and identification and control of quantized systems.



Jimin Wang (IEEE Member) received the B.S. degree in mathematics from Shandong Normal University, China, in 2012 and the Ph.D. degree from School of Mathematics, Shandong University, China, in 2018. From May 2017 to May 2018, he was a joint Ph.D. student with the School of Electrical Engineering and Computing, The University of Newcastle, Australia. From July 2018 to December 2020, he was a postdoctoral researcher in the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), China. He is currently an associate professor in the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His current research interests include privacy and security in cyber-physical systems, stochastic systems and networked control systems. He is a member of the IEEE CSS Technical Committee on Security and Privacy, the IEEE CSS Technical Committee on Networks and Communication Systems, the IFAC Technical Committee 1.5 on Networked Systems.

Ji-Feng Zhang (IEEE Fellow) received the B.S. degree in mathematics from Shandong University, China, in 1985, and the Ph.D. degree from the Institute of Systems Science (ISS), Chinese Academy of Sciences (CAS), China, in 1991. Since 1985, he has been with the ISS, CAS. Now he is also with the School of Automation and Electrical Engineering, Zhongyuan University of Technology. His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems.

He is an IEEE Fellow, IFAC Fellow, CAA Fellow, SIAM Fellow, member of the European Academy of Sciences and Arts, and Academician of the International Academy for Systems and Cybernetic Sciences. He received the Second Prize of the State Natural Science Award of China in 2010 and 2015, respectively. He was a Vice-President of the Chinese Association of Automation, the Chinese Mathematical Society and the Systems Engineering Society of China. He was a Vice-Chair of the IFAC Technical Board, member of the Board of Governors, IEEE Control Systems Society; Convenor of Systems Science Discipline, Academic Degree Committee of the State Council of China. He served as Editor-in-Chief, Deputy Editor-in-Chief or Associate Editor for more than 10 journals, including *Science China Information Sciences*, *IEEE Transactions on Automatic Control* and *SIAM Journal on Control and Optimization* etc.