J Syst Sci Complex (2023) 36: 187–204

Differentially Private Distributed Parameter Estimation*

WANG Jimin · TAN Jianwei · ZHANG Ji-Feng

DOI: 10.1007/s11424-022-2012-9
Received: 5 January 2022 / Revised: 17 May 2022

Abstract Data privacy is an important issue in control systems, especially when datasets contain sensitive information about individuals. In this paper, the authors are concerned with the differentially private distributed parameter estimation problem, that is, we estimate an unknown parameter while protecting the sensitive information of each agent. First, the authors propose a distributed stochastic approximation estimation algorithm in the form of the differentially private consensus+innovations (DP-CI), and establish the privacy and convergence property of the proposed algorithm. Specifically, it is shown that the proposed algorithm asymptotically unbiased converges in mean-square to the unknown parameter while differential privacy-preserving holds for finite number of iterations. Then, the exponentially damping step-size and privacy noise for DP-CI algorithm is given. The estimate approximately converges to the unknown parameter with an error proportional to the step-size parameter while differential privacy-preserving holds for all iterations. The tradeoff between accuracy and privacy of the algorithm is effectively shown. Finally, a simulation example is provided to verify the effectiveness of the proposed algorithm.

Keywords Differential privacy, distributed parameter estimation, stochastic approximation.

1 Introduction

When estimating an unknown signal/parameter in a distributed sensor network, each sensor can produce a local estimate based on its own noisy measurements and the information gathered from other sensors. In the centralized estimation scenario, all the sensors transmit data to a fusion center. With the fast development of sensor networks and wireless communications, the

WANG Jimin
School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083*, China.* Email: jimwang@ustb.edu.cn.
TAN Jianwei · ZHANG Ji-Feng (Corresponding author)
Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China; School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing* 100049*, China.* Email: jwtan@amss.ac.cn; jif@iss.ac.cn.
*The work is supported by the National Key R&D Program of China under Grant No. 2018YFA0703800, the National Natural Science Foundation of China under Grant No. 61877057, and China Post-Doctoral Science Foundation under Grant No. 2018M641506.
◇ *This paper was recommended for publication by Editor WU Zhengguang.*

scale of system is becoming increasingly large, the computation and communication burden increases rapidly with the system's size. On the other hand, in the centralized processing, collecting measurements from all other distributed sensors over the network may be infeasible in many practical situations due to limited communication capabilities, energy consumptions or packet losses. Thus, distributed estimation algorithms are developed, which will be more robust, need fewer communications and allow parallel information processing. Since sensors need only to exchange information with their own neighbors, distributed estimation has attracted a great deal of attention[1–12], and has widely applications in collaborative spectral sensing in cognitive radio systems, target localization in biological networks, fish schooling, bee swarming, and bird flight in mobile adaptive networks, etc.

However, the communication among agents in such a distributed manner brings about privacy concerns if the local agent's training data contains sensitive information such as salary, medical records, initial states. For example, in biological networks, animals are interested in moving toward a target (such as a nutrition source). The "home" of the animals is sensitive for hunters/predators. In traffic networks, the initial location of each vehicle is the driver's house, which is sensitive information. Therefore, it is of great importance to protect sensitive information in multi-agent systems[13]. In the realm of control systems, some privacy-preserving approaches have recently been proposed[14–19], such as homomorphic encryption[14, 15], adding artificial noise[16–19], etc.

Among others, differential privacy is a well-known privacy-preserving method and has applications in many domains such as data mining[20], distributed multiset intersection and union[21], machine learning[22, 23], distributed optimization[24–27, 39] and so on. Roughly speaking, differential privacy deliberately releases data and ensures that the participation of a single agent in a database does not affect the output of data processing substantially. In this case, it is unlikely in the sense of probability that an eavesdropper could learn each agent's sensitive information. The basic idea used by differential privacy is to "perturb" the exact data before releasing them, which will compromise the system performances[28].

Many important works on differential privacy based control and estimation have been presented[29–38]. Specifically, [34, 35] studied the differentially private estimation in the centralized processing. [36] considered a differentially private distributed stochastic gradient algorithm. However, the convergence analysis for the proposed algorithm in [36] was not presented. [37] gave a differentially private distributed stochastic gradient algorithm with connected gossiping agents, where the input perturbation (adding noise to the sensitive information) is used for achieving the goal of privacy protection. [38] studied a differentially private algorithm for linear regression learning in a decentralized fashion, where the t-step privacy-preserving analysis and estimation error bound were given. However, the estimation error bound is given by $O(t)$ or $O(\exp(t^\alpha)), 0 \leq \alpha < 1$, which is not reasonable in practice.

In this paper, we study the differentially private distributed parameter estimation problem, that is, we estimate the unknown parameter and at the same time protect the sensitive information of each agent. A new differentially private distributed parameter estimation algorithm is given, i.e., the DP-CI algorithm. First, a distributed stochastic approximation algorithm is

provided to simultaneously handle differential privacy requirements and distributed parameter estimation. The estimate converges to the true parameter in mean-square and the differential privacy-preserving holds for finite number of iterations. Then, the exponentially damping step-size and privacy noise for DP-CI algorithm is given to preserve the differential privacy of all iterations, but the estimate approximately converges to the unknown parameter with an error proportional to the step-size parameter. The main contributions are given as follows.

1) We introduce the differential privacy into distributed parameter estimation problem, where each agent's sensitive information is protected from eavesdropper or the honest-but-curious agent. The protection level is measured in the sense of ε-differential privacy. For two different step-sizes and privacy noise forms, the privacy and convergence analysis of the algorithm are studied, respectively. The existing literature on distributed estimation[4–6, 10, 11] involves only parameter estimate but without privacy protection mechanisms. In contrast to [37], here we not only establish the estimate convergence in mean-square but also the mean-square convergence rate of the proposed estimation algorithm.

2) For a class of step-size with the form of $t^{-\gamma}$, $0 < \gamma \leq 1$, we give the convergence rate of the DP-CI algorithm when the $T\varepsilon$-differential privacy holds. In practice, it is more reasonable than the case where the estimation error bound is given by $O(t)$ or $O(\exp(t^{\alpha})), 0 \leq \alpha < 1$ in [38].

3) The form of the added noise variances for privacy-preserving is general in this paper. The exponentially damping Laplacian noise given in advance is a special case of this paper[25, 29, 30, 32, 33]. Moreover, different from the added noise σ_t being $O(t^{-1})$ in [26, 27], the added noise σ_t in this paper can be $O(t^{-\gamma})$, $0 < \gamma \leq 1$.

2 Preliminaries

2.1 Notations

Throughout this paper, the following standard notations are used. $Z \geq 0$ ($Z > 0$) means that the symmetric matrix Z is semi-positive definite (positive definite). $\mathbf{1}_N$ stands for the N-dimensional vector with all elements being one. $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote, respectively, the n-dimensional Euclidean space, and the set of all $m \times n$ real matrices. $\|x\|$ refers to Euclidean norm of the vector x. I, 0 are identity matrix and zero matrix with appropriate dimensions, respectively. In addition, diag $\{A_1, A_2, \cdots, A_n\}$ stands for a (block) diagonal matrix with $A_1, A_2, \cdots, A_n$ in order on the diagonal. The expectation of a random variable X is denoted by $\mathbb{E}[X]$. $\|x\|_1$ denotes the 1-norm of the vector $x \in \mathbb{R}^n$, i.e., $\|x\|_1 = \sum_{i=1}^{n} |x_i|$. $\otimes$ denotes the Kronecker product.

2.2 Graph Theory

In this paper, the communication among agents of a network is modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where consists of a non-empty node set $\mathcal{V} = \{1, 2, \cdots, N\}$ and an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. $A = [a_{i,j}]$ is the adjacency matrix of $\mathcal{G}$, where $a_{i,j} = 1$ if $(i, j) \in \mathcal{E}$ and $a_{i,j} = 0$, otherwise. $\mathcal{N}_i = \{j \in \mathcal{V}, (j, i) \in \mathcal{E}\}$ denotes the neighborhood of agent i. Here, we assume the

self-edge (i,i) does not exist. $\mathcal{G}$ is called connected if for any pair agents (i_1,i_l), there exists a path from i_1 to i_l consisting of edges $(i_1,i_2),(i_2,i_3),\cdots,(i_{l-1},i_l)$. The Laplacian matrix of $\mathcal{G}$ is defined as $\mathcal{L}=\mathcal{D}-A$, where the degree matrix $\mathcal{D}=\mathrm{diag}\{\sum_{j=1}^{N}a_{1j},\sum_{j=1}^{N}a_{2j},\cdots,\sum_{j=1}^{N}a_{Nj}\}$.

3 Problem Formulation

3.1 Observation Model

Here we consider a linear observation model in a network given by

$$y_i(t)=H_i(t)\theta^*+\omega_i(t),\quad i\in\mathcal{V},\tag{1}$$

where $y_i(t)\in\mathbb{R}^{m_i}$ is the measurement vector, $\omega_i(t)\in\mathbb{R}^{m_i}$ is the zero-mean i.i.d. measurement noise, and $H_i(t)\in\mathbb{R}^{m_i\times n}$ represents the time-varying measurement matrix of agent i. $\theta^*\in\mathbb{R}^n$ is an unknown parameter vector.

As shown in [4–7], to identify system parameter θ^* in (1), each agent updates its estimate as follows: $x_i(t+1)=x_i(t)-\alpha(t)\sum_{j\in\mathcal{N}_i}(x_i(t)-x_j(t))+\alpha(t)H_i^{\mathrm{T}}(t)(y_i(t)-H_i(t)x_i(t))$. During the iterative process, each agent in the network needs to exchange the estimate $x_i(t)$ with its neighbors. Although there is no need for each agent to share its own data, the risk of information leakage still exists if the local data contain sensitive information like medical or financial records. The sensitive information in our setup is $\{y_i(0),y_i(1),\cdots,y_i(T)\}$ for some $T>0,i\in\mathcal{V}$. Generally, the initial state and the states near that are important for each agent, for example, the initial position of a vehicle. The adversary could be an outsider who eavesdrops the exchanging information, or the honest-but-curious agent who follows the iterative process honestly but tends to infer the sensitive information. We assume that adversary have the following information: (i) The communication topology of the network; (ii) The exchanging information among agents. In this case, adversary may steal the sensitive information of the agents, e.g., model inversion attack method.

In the following, we will design a new distributed parameter estimation algorithm to protect the sensitive information from potential adversary.

3.2 Differential Privacy

In order to protect the sensitive information from potential adversary, inspired by [34], we introduce the concepts about differential privacy.

Definition 3.1 (δ-adjacency) Given $\delta>0$, two vectors $Y(t)=[y_1^{\mathrm{T}}(t),y_2^{\mathrm{T}}(t),\cdots,y_N^{\mathrm{T}}(t)]^{\mathrm{T}}$ and $Y'(t)=[y_1'^{\mathrm{T}}(t),y_2'^{\mathrm{T}}(t),\cdots,y_N'^{\mathrm{T}}(t)]^{\mathrm{T}}$, $Y(t)$ and $Y'(t)$ are δ-adjacent if there exists some $i_0\in\mathcal{V}$ such that

$$y_i(t)=y_i'(t),\quad\forall i\neq i_0,\quad\|y_{i_0}(t)-y_{i_0}'(t)\|_1\leq\delta.\tag{2}$$

Remark 3.2 Definition 3.1 implies that two signal sets are adjacent if only one agent changes its measurement vector, which is a key component of any private implementation as it specifies which pieces of sensitive data must be made approximately indistinguishable to potential adversary.

Definition 3.3 (Differential privacy) Given $\varepsilon > 0$, a randomized mechanism $\mathcal{M}$ is ε-differential privacy if for any δ-adjacent vectors $Y(t)$ and $Y'(t)$, and any set of outputs $\Upsilon \subseteq \text{Range}(\mathcal{M})$, such that

$$\mathbb{P}\{\mathcal{M}(Y(t)) \in \Upsilon\} \leq \mathrm{e}^{\varepsilon}\mathbb{P}\{\mathcal{M}(Y'(t)) \in \Upsilon\}.$$

Remark 3.4 Note that the constant ε measures the privacy level of the randomized mechanism $\mathcal{M}$, i.e., a smaller ε implies a higher privacy level. As pointed out in [34], ε is taken to be a small constant, e.g., $\varepsilon \approx 0.1$.

Lemma 3.5 (see [26]) (*Adaptive sequential composition*) *Consider a sequence of mechanisms* $\{\mathcal{M}_t\}_{t=1}^{\mathrm{T}}$, *in which the output of* $\mathcal{M}_t$ *may depend on* $\mathcal{M}_1$, $\mathcal{M}_2$,$\cdots$, $\mathcal{M}_{t-1}$ *as described below:*

$$\mathcal{M}_t(D) = \mathcal{M}_t(D, \mathcal{M}_1(D), \mathcal{M}_2(D), \cdots, \mathcal{M}_{t-1}(D)).$$

Suppose $\mathcal{M}_t(\cdot, a_1, a_2, \cdots, a_{t-1})$ *preserves* ε_t*-differential privacy for any* $a_1 \in \text{range}(\mathcal{M}_1)$,$\cdots$, $a_{t-1} \in \text{range}(\mathcal{M}_{t-1})$. *Then, the* T*-tuple mechanism* $\mathcal{M} := (\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_T)$ *preserves* ε*-differential privacy for* $\varepsilon = \sum_{t=1}^{T} \varepsilon_t$.

Remark 3.6 The worst privacy happens when an adversary applies many queries to the same sensitive data. If the distributed algorithm runs in an offline way (i.e., each update depends on all data points over the T iterations), then the offline algorithm suffers the "sum" of the differential privacy described in Lemma 3.5.

The objective in this paper: We design a fully distributed algorithm to preserve the privacy of each measurement vector $y_i(t)$ and estimate the unknown parameter θ^*, and further analyze the tradeoff between accuracy and privacy of the algorithm.

4 The DP-CI Algorithm

In this section, we propose a differentially private distributed parameter estimation algorithm, which we term as DP-CI algorithm. The key steps of the DP-CI algorithm are summarized as: (i) Transmit to its neighbors a message that is corrupted with noise (rather than its internal estimate as in the traditional CI algorithm); (ii) Use the (noisy) messages from its neighbors to update its internal state. In the following, we give these two steps for the algorithm: Message passing and estimation update.

Message passing In every iteration of the DP-CI algorithm, each agent i transmits its current noisy estimate $\widetilde{x}_i(t)$ to each of its neighbors $j \in \mathcal{N}_i$. Specifically, agents broadcast $\widetilde{x}_i(t) = x_i(t) + n_i(t)$, where $x_i(t)$ is the estimate of agent i at time $t \geq 1$ for the parameter vector θ^*, and each element of $n_i(t) \in \mathbb{R}^n$ is the zero-mean i.i.d. Laplacian noise with the variance of $2\sigma_t^2$, i.e., $n_i^l(t) \sim \text{Lap}(0, \sigma_t), l = 1, 2, \cdots, n$.

Estimation update

$$x_i(t+1) = \widetilde{x}_i(t) - \alpha(t) \sum_{j \in \mathcal{N}_i} (\widetilde{x}_i(t) - \widetilde{x}_j(t)) + \alpha(t) H_i^{\mathrm{T}}(t)(y_i(t) - H_i(t)\widetilde{x}_i(t)), \tag{3}$$

where $\alpha(t)$ is the step-size of the proposed algorithm. The selection of $\alpha(t)$ to ensure the convergence and differential privacy of the proposed algorithm will be discussed in more details later.

Remark 4.1 Different from the traditional distributed parameter estimation algorithm (see [4–6, 10, 11]), to ensure the ε-DP of the DP-CI algorithm, we add the Laplacian noise to the local estimation $x_i(t)$ when broadcast it.

4.1 Privacy Analysis

In this subsection, we will show the ε-differential privacy for each iteration of the DP-CI algorithm. In the context of differential privacy, the corresponding mechanism for the DP-CI algorithm maps $D = \{Y(t),\ t \in \mathbb{N}\}$ to a sequence of messages $\{\widetilde{X}(t),\ t \geq 1\}$. This method guarantees differential privacy and is known as output perturbation[34]. Next, we derive conditions on the noise variances under which each iteration of the proposed algorithm satisfies ε-differential privacy.

In differential privacy, a key quantity determines how much noise to be added at each iteration, which is referred to as the sensitivity of the proposed algorithm.

Definition 4.2 (see [23]) (Sensitivity) For given δ-adjacency relation (2), the sensitivity of an output map g at each iteration t is defined as

$$\Delta(t) = \sup_{\{Y(t),Y'(t):\mathrm{Adj}(Y(t),Y'(t))\}} \|g(Y(t)) - g(Y'(t))\|_1.$$

Remark 4.3 The sensitivity of an output map g captures the magnitude by which a single agent's data can change the output map g in the worst case. For our proposed algorithm, g refers to the output map from $Y(t)$ to $X(t) = [x_1^{\mathrm{T}}(t), x_2^{\mathrm{T}}(t), \cdots, x_N^{\mathrm{T}}(t)]^{\mathrm{T}}$.

Lemma 4.4 *For given δ-adjacency relation (2), the sensitivity of the DP-CI algorithm at each iteration t satisfies $\Delta(t) \leq \alpha(t-1)\delta H_{\max}(t-1)$, where $H_{\max}(t-1) = \max_{i\in\mathcal{V}}\{S(H_i(t-1))\}, S(H_i(t-1)) := \sum_{j=1}^{m_i}\sum_{k=1}^{n}|(H_i(t-1))_{jk}|$.*

Proof For given δ-adjacency relation (2), from (3) it follows that $\|x_i(t) - x_i'(t)\|_1 = \|\alpha(t-1)H_i^{\mathrm{T}}(t-1)(y_i(t-1) - y_i'(t-1))\|_1 \leq \alpha(t-1)\delta S(H_i(t-1))$. By Definition 4.2, for two δ-adjacent data $Y(t-1)$ and $Y'(t-1)$, it follows that $\Delta(t) = \sup_{\{Y(t-1),Y'(t-1):\mathrm{Adj}(Y(t-1),Y'(t-1))\}}\|X(t) - X'(t)\|_1 \leq \max_{i\in\mathcal{V}}\|x_i(t) - x_i'(t)\|_1 \leq \alpha(t-1)\delta H_{\max}(t-1)$. ∎

Theorem 4.5 *For given $\varepsilon > 0$, each iteration of the DP-CI algorithm is ε-DP if σ_t satisfies*

$$\sigma_t \geq \frac{\alpha(t-1)}{\varepsilon}\delta H_{\max}(t-1). \tag{4}$$

Proof Let $Y(t), Y'(t) \in \mathbb{R}^m, m = \sum_{i=1}^{N} m_i$ be the δ-adjacent measurement vector, $X(t) \in \mathbb{R}^{nN}$ be the estimate vector at time t, $\widetilde{X}(t) = [\widetilde{x}_1^{\mathrm{T}}(t), \widetilde{x}_2^{\mathrm{T}}(t), \cdots, \widetilde{x}_N^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{nN}$ be the broadcast vector at time t, respectively. $\mathcal{M}(\cdot)$ denotes the process from sensitive data $Y(t)$ to broadcast data $\widetilde{X}(t+1)$. In order to preserve ε-DP, $Y(t)$ and $Y'(t)$ generate identical observation, i.e., for any $t \geq 1$, $\widetilde{X}(t)$ and $\widetilde{X}'(t)$ are in the same observation set.

First, for any given $t > 0$, we compute the probability of $\mathcal{M}(Y(t)) = \widetilde{X}(t)$,

$$\begin{aligned}\Pr\left(\mathcal{M}(Y(t)) = \widetilde{X}(t)\right) &= \Pr(\widetilde{X}(t+1)|\widetilde{X}(t), Y(t)) \\ &= \Pr(n(t) = \widetilde{X}(t) - X(t)),\end{aligned} \tag{5}$$

where $n(t) = [n_1^{\mathrm{T}}(t), n_2^{\mathrm{T}}(t), \cdots, n_N^{\mathrm{T}}(t)]^{\mathrm{T}} \in \mathbb{R}^{nN}$.

Similarly,

$$\Pr\left(\mathcal{M}(Y'(t)) = \widetilde{X}(t)\right) = \Pr(n(t) = \widetilde{X}(t) - X'(t)). \tag{6}$$

Then, as each element of $n_i(t)$ is also mutually independent, from (5) and (6) it follows that the ratio of corresponding probability density functions satisfies

$$\begin{aligned}\frac{f(\mathcal{M}(Y(t)) = \widetilde{X}(t))}{f(\mathcal{M}(Y'(t)) = \widetilde{X}(t))} &= \frac{f(n(t) = \widetilde{X}(t) - X(t))}{f(n(t) = \widetilde{X}(t) - X'(t))} \\ &= \frac{\exp\left(||\widetilde{X}(t) - X(t)||_1/\sigma_t\right)}{\exp\left(||\widetilde{X}(t) - X'(t)||_1/\sigma_t\right)} \\ &\leq \exp\left(||X(t) - X'(t)||_1/\sigma_t\right) \\ &\leq \exp\left(\frac{\Delta(t)}{\sigma_t}\right).\end{aligned}$$

Since $\varepsilon \geq \frac{\Delta(t)}{\sigma_t}$, we have $\frac{f(\mathcal{M}(Y(t))=\widetilde{X}(t))}{f(\mathcal{M}(Y'(t))=\widetilde{X}(t)}) \leq \mathrm{e}^{\varepsilon}$. Thus, for any measurable set of $\Upsilon \subseteq \mathit{Range}(\mathcal{M}(Y(t)))$, it holds

$$\begin{aligned}\Pr(\mathcal{M}(Y(t)) \in \Upsilon) &= \int_{\Upsilon} f(\mathcal{M}(Y(t)) = Z)\mathrm{d}Z \\ &\leq \mathrm{e}^{\varepsilon} \int_{\Upsilon} f(\mathcal{M}(Y'(t)) = Z)\mathrm{d}Z \\ &= \mathrm{e}^{\varepsilon} \Pr(\mathcal{M}(Y'(t)) \in \Upsilon).\end{aligned}$$

Therefore, according to Definition 3.3, the theorem is obtained. ∎

Remark 4.6 Different from the exponentially damping Laplacian noise given in advance (see [25, 29, 30, 32, 33]), the form of the privacy noise given in this paper is more general and only needs to satisfy $\sigma_t = \frac{\alpha(t-1)}{\varepsilon}\delta H_{\max}(t-1)$, e.g., $\sigma_t = \frac{1}{t^{\gamma}}$, $\frac{1}{2} < \gamma \leq 1$. Specifically, if we set the zero-mean i.i.d. Laplacian noise $n(t)$ with the variance of $2\sigma_t^2$, where $\sigma_t = cq^t$, $0 < c < 1$, $0 < q < 1$, the results in Theorem 4.5 still hold. However, as shown below, the step-size in this form cannot guarantee the mean-square convergence of the DP-CI algorithm.

Remark 4.7 In Theorem 4.5, we give the relationship between the scale parameter σ_t of the added noise, the step-size $\alpha(t)$ and the privacy index ε. From (4) it follows that the scalar parameter σ_t of the added noise is inversely proportional to privacy index ε. In other words, each agent preserves stronger privacy when the added noise is more dispersive. In order to preserve ε-DP and meanwhile make parameter estimation as accurate as possible, we will take $\sigma_t = \frac{\alpha(t-1)}{\varepsilon}\delta H_{\max}(t-1)$ in the following.

To facilitate convergence analysis of (3), we define the stacked vectors and matrices as follows

$$\begin{aligned}&\Theta^*=\mathbf{1}_N\otimes\theta^*,\quad H(t)=\mathrm{diag}\{H_1^{\mathrm{T}}(t),H_2^{\mathrm{T}}(t),\cdots,H_N^{\mathrm{T}}(t)\},\\&X(t)=\big[x_1^{\mathrm{T}}(t),x_2^{\mathrm{T}}(t),\cdots,x_N^{\mathrm{T}}(t)\big]^{\mathrm{T}},\quad Y(t)=\big[y_1^{\mathrm{T}}(t),y_2^{\mathrm{T}}(t),\cdots,y_N^{\mathrm{T}}(t)\big]^{\mathrm{T}},\\&\omega(t)=\big[\omega_1^{\mathrm{T}}(t),\omega_2^{\mathrm{T}}(t),\cdots,\omega_N^{\mathrm{T}}(t)\big]^{\mathrm{T}},\quad n(t)=\big[n_1^{\mathrm{T}}(t),n_2^{\mathrm{T}}(t),\cdots,n_N^{\mathrm{T}}(t)\big]^{\mathrm{T}}.\end{aligned}$$

Now, we rewrite (3) in the following compact form:

$$\begin{aligned}X(t+1)=&X(t)-\alpha(t)(\mathcal{L}\otimes I_n)X(t)+\alpha(t)H(t)\omega(t)\\&+\alpha(t)H(t)H^{\mathrm{T}}(t)(\Theta^*-X(t))+n(t)\\&-\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))n(t).\end{aligned}\tag{7}$$

In order to obtain the performance analysis of the DP-CI algorithm, we make the following assumptions:

A1) The graph is connected.

A2) There exist positive numbers $\gamma_0>0,\eta_0>0$ such that $\eta_0 I\le G(t):=\sum_{i=1}^N H_i^{\mathrm{T}}(t)H_i(t)\le\gamma_0 I$ hold for all $t\in\mathbb{N}$.

A3) $\{\omega(t),t\ge0\}$ is with $\sigma_\omega^2\triangleq\sup_{t\ge0}\mathbb{E}[\|\omega(t)\|^2]<\infty$.

Next, we will show the tradeoff between accuracy and privacy of the algorithm with different forms of step-size and privacy noise.

4.2 Convergence Analysis with Stochastic Approximation-Type Step-Size

In this subsection, we adopt the stochastic approximation-type step-size in the DP-CI algorithm, i.e., the step-size sequence $\{\alpha(t),t\ge0\}$ satisfies the following assumption.

A4) The step-size $\{\alpha(t),t\ge0\}$ satisfies

$$\alpha(t)>0,\quad\sum_{t=0}^{\infty}\alpha(t)=\infty,\quad\lim_{t\to\infty}\alpha(t)=0,\quad\alpha(t+1)=O(\alpha(t)).$$

Remark 4.8 One example of the step-size sequence $\{\alpha(t),t\ge0\}$, which satisfies Assumption A4), can be chosen as $\alpha(t)=\frac{1}{t^\gamma},0<\gamma\le1$.

Before giving the main results, we first introduce two lemmas.

Lemma 4.9 (see [4]) *Under Assumptions* A1) *and* A2), $\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t)$ *is a positive definite symmetry matrix for all* $t\in\mathbb{N}$. *Furthermore, there exist a positive definite matrix* $M\in\mathbb{R}^{Nn\times Nn}$ *and a sufficiently large integer* T, *such that* $\alpha(t)M<\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))<I$, *for any* $t>T$.

Lemma 4.10 (see [40]) *Let* $\{V(t),t=0,1,\cdots\}$, $\{\varrho(t),t=0,1,\cdots\}$, *and* $\{q(t),t=0,1,\cdots\}$ *be real sequences, satisfying* $0<q(t)\le1$, $\varrho(t)\ge0$, $t=0,1,\cdots$, $\sum_{t=0}^{\infty}q(t)=\infty$, $\lim_{t\to\infty}\frac{\varrho(t)}{q(t)}=0$, *and* $V(t+1)\le(1-q(t))V(t)+\varrho(t)$. *Then,* $\limsup_{t\to\infty}V(t)\le0$. *Particularly, if* $V(t)\ge0$, $t=0,1,\cdots$, *then* $\lim_{t\to\infty}V(t)=0$.

In the following theorem, we will show how the privacy index ε affect the convergence rate of the DP-CI algorithm when $\alpha(t)=\frac{1}{t^\gamma},0<\gamma\le1$.

Theorem 4.11 *Suppose Assumptions* A1)–A3) *hold,* $\alpha(t)=\frac{1}{t^\gamma}$, $\sigma_t=\frac{\delta H_{\max}(t-1)}{\varepsilon(t-1)^\gamma}, 0<\gamma\leq 1$. *Then, the convergence rate of the DP-CI algorithm is given as follows.*

(i) *When* $0<\gamma<1$, *there holds* $\sum_{i\in\mathcal{V}}\mathbb{E}\|x_i(t)-\theta^*\|^2=O\big(\frac{1}{\varepsilon^2 t^\gamma}\big)$.

(ii) *When* $\gamma=1$, *there holds*

$$\sum_{i\in\mathcal{V}}\mathbb{E}\|x_i(t)-\theta^*\|^2=\begin{cases}O\Big(\dfrac{1}{\varepsilon^2 t^{b_0}}\Big), & b_0<1,\\ O\Big(\dfrac{\ln t}{\varepsilon^2 t}\Big), & b_0=1,\\ O\Big(\dfrac{1}{\varepsilon^2 t}\Big), & b_0>1,\end{cases}$$

where b_0 *is a constant such that* $0<b_0\leq\min\{\lambda_2(\mathcal{L}),\eta_0\}$.

Proof Let $\delta(t)=X(t)-\Theta^*$ be the parameter estimation error. Then, from $(\mathcal{L}\otimes I_n)\Theta^*=0$ and (7), we have

$$\begin{aligned}\delta(t+1)=&\big[I-\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))\big]\delta(t)+n(t)\\&+\alpha(t)H(t)\omega(t)-\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))n(t).\end{aligned}\tag{8}$$

Let $V(t)=\|\delta(t)\|^2$. Then, from (8) it follows that

$$V(t+1)=\delta^{\mathrm{T}}(t)\Gamma_0^{\mathrm{T}}\Gamma_0\delta(t)+2\delta^{\mathrm{T}}(t)\Gamma_0^{\mathrm{T}}\Gamma_1+\|\Gamma_1\|^2,\tag{9}$$

where $\Gamma_0=\big[I-\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))\big]$, $\Gamma_1=n(t)+\alpha(t)H(t)\omega(t)-\alpha(t)(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))n(t)$.

By [41], Theorem 2.8, from Assumption A1) if follows that $\mathcal{L}$ has a unique eigenvalue zero, and $\lambda_2(\mathcal{L})>0$. Then, from Assumption A2) and Lemma 4.9, there exists a constant $b_0>0$ such that $\mathbb{E}\big[\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t)\big]>b_0 I$ and

$$\mathbb{E}\big[\delta^{\mathrm{T}}(t)\Gamma_0^{\mathrm{T}}\Gamma_0\delta(t)\big]\leq\big[1-2b_0\alpha(t)+\Gamma_2\alpha^2(t)\big]V(t),\tag{10}$$

where $\Gamma_2=2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2$. Since $\omega(t)$ and $n(t)$ are the zero-mean noise, we have $\mathbb{E}\big[\omega(t)\big]=\mathbb{E}\big[n(t)\big]=0$, which further implies that

$$\mathbb{E}\big[\delta^{\mathrm{T}}(t)\Gamma_0^{\mathrm{T}}\Gamma_1\big]=0.\tag{11}$$

In addition, since the distribution of $\omega(t)$ and $n(t)$ is independent, by the Cauchy-Schwarz inequality we have

$$\begin{aligned}\mathbb{E}\big[\|\Gamma_1\|^2\big]\leq&\,3\mathbb{E}[\|n(t)\|^2]+3\alpha^2(t)\|H(t)\|^2\mathbb{E}[\|\omega(t)\|^2]\\&+3\alpha^2(t)\|(\mathcal{L}\otimes I_n+H(t)H^{\mathrm{T}}(t))\|^2\mathbb{E}[\|n(t)\|^2]\\\leq&\,3\mathbb{E}[\|n(t)\|^2]+3\gamma_0\alpha^2(t)\mathbb{E}[\|\omega(t)\|^2]\\&+3\alpha^2(t)\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2\mathbb{E}[\|n(t)\|^2].\end{aligned}$$

This together with (9)–(11) leads to

$$\mathbb{E}\big[V(t+1)\big]\leq\big[1-2b_0\alpha(t)+\Gamma_2\alpha^2(t)\big]\mathbb{E}\big[V(t)\big]$$

$$+3\mathbb{E}[\|n(t)\|^2]+3\gamma_0\alpha^2(t)\mathbb{E}[\|\omega(t)\|^2]$$
$$+3\alpha^2(t)\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2\mathbb{E}[\|n(t)\|^2]. \tag{12}$$

When $\alpha(t)=\frac{1}{t^\gamma}, 0<\gamma\le 1$, by (12) there exist $t>t_0$ and $\beta>0$ such that $-2b_0\alpha(t)+\Gamma_2\alpha^2(t)\le -\frac{b_0}{t^\gamma}$, $3(1+\alpha^2(t)\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2)\mathbb{E}[\|n(t)\|^2]+3\gamma_0\alpha^2(t)\mathbb{E}[\|\omega(t)\|^2]\le\frac{\beta}{\varepsilon^2 t^{2\gamma}}$. Then, we have $\mathbb{E}\big[V(t+1)\big]\le\big[1-\frac{b_0}{t^\gamma}\big]\mathbb{E}\big[V(t)\big]+\frac{\beta}{\varepsilon^2 t^{2\gamma}}$, as $t>t_0$. Thus, iteratively we have

$$\mathbb{E}\big[V(t+1)\big]\le\prod_{k=t_0}^{t}\left[1-\frac{b_0}{k^\gamma}\right]\mathbb{E}\big[V(t_0)\big]+\sum_{l=t_0}^{t-1}\prod_{k=l+1}^{t}\left(1-\frac{b_0}{k^\gamma}\right)\frac{\beta}{\varepsilon^2 l^{2\gamma}}+\frac{\beta}{\varepsilon^2 t^{2\gamma}}. \tag{13}$$

Note that $\prod_{k=t_0}^{t}\big[1-\frac{b_0}{k^\gamma}\big]=\exp\big(\sum_{k=t_0}^{t}\log(1-\frac{b_0}{k^\gamma})\big)=O\big(\exp\big(-\sum_{k=t_0}^{t}\frac{b_0}{k^\gamma}\big)\big)$. Then, we will discuss the above equation for $\gamma=1$ and $0<\gamma<1$, respectively.

When $\gamma=1$, it is obtained that

$$\prod_{k=t_0}^{t}\left[1-\frac{b_0}{k^\gamma}\right]=O\bigg(\exp\bigg(-\sum_{k=t_0}^{t}\frac{b_0}{k}\bigg)\bigg)$$
$$=O\bigg(\exp\bigg(-b_0\log\frac{t}{t_0}\bigg)\bigg)=O\bigg(\frac{1}{t^{b_0}}\bigg). \tag{14}$$

From (13) and (14) it follows that $\mathbb{E}\big[V(t+1)\big]=O\big(\frac{1}{t^{b_0}}\big)+O\big(\sum_{l=t_0}^{t-1}(\frac{l}{t})^{b_0}\frac{\beta}{\varepsilon^2 l^2}\big)+O(\frac{\beta}{\varepsilon^2 t^2})=O\big(\exp\big(-b_0\log\frac{t}{t_0}\big)\big)+O\big(\frac{1}{t^{b_0}}\sum_{l=t_0}^{t-1}\frac{\beta}{\varepsilon^2 l^{2-b_0}}\big)+O(\frac{\beta}{\varepsilon^2 t^2})=O\big(\frac{1}{t^{b_0}}\big)+O\big(\frac{1}{t^{b_0}}\sum_{l=t_0}^{t-1}\frac{\beta}{\varepsilon^2 l^{2-b_0}}\big)+O(\frac{\beta}{\varepsilon^2 t^2})$. By $\sum_{l=t_0}^{t-1}\frac{\beta}{\varepsilon^2 l^{2-b_0}}\le\int_{t_0-1}^{t}\frac{\beta}{\varepsilon^2 x^{2-b_0}}dx$, we have

$$\sum_{l=t_0}^{t-1}\frac{\beta}{\varepsilon^2 l^{2-b_0}}=\begin{cases}O\Big(\dfrac{1}{\varepsilon^2}\Big), & b_0<1,\\ O\Big(\dfrac{\ln t}{\varepsilon^2}\Big), & b_0=1,\\ O\Big(\dfrac{1}{\varepsilon^2 t^{1-b_0}}\Big), & b_0>1.\end{cases}$$

Thus, the results hold for $\gamma=1$.

When $0<\gamma<1$, we have $\prod_{k=t_0}^{t}\big[1-\frac{b_0}{k^\gamma}\big]=O\big(\exp\big(-\sum_{k=t_0}^{t}\frac{b_0}{k^\gamma}\big)\big)=O\big(\exp\big(-\frac{b_0}{1-\gamma}[(t+1)^{1-\gamma}-t_0^{1-\gamma}]\big)\big)$. Note that for large enough t_0 and $l\ge t_0$, $(1-\frac{b_0}{l^\gamma})^{-1}\le 2$. Then, from (13) it follows that

$$\begin{aligned}&\mathbb{E}\big[V(t+1)\big]\\ &\le\prod_{k=t_0}^{t}\left[1-\frac{b_0}{k^\gamma}\right]\mathbb{E}\big[V(t_0)\big]+\sum_{l=t_0}^{t-1}\prod_{k=l+1}^{t}\left(1-\frac{b_0}{k^\gamma}\right)\frac{\beta}{\varepsilon^2 l^{2\gamma}}+\frac{\beta}{\varepsilon^2 t^{2\gamma}}\\ &\le\prod_{k=t_0}^{t}\left[1-\frac{b_0}{k^\gamma}\right]\mathbb{E}\big[V(t_0)\big]+2\sum_{l=t_0}^{t-1}\prod_{k=l}^{t}\left(1-\frac{b_0}{k^\gamma}\right)\frac{\beta}{\varepsilon^2 l^{2\gamma}}+\frac{\beta}{\varepsilon^2 t^{2\gamma}}\\ &=O\bigg(\exp\bigg(-\frac{b_0}{1-\gamma}(t+1)^{1-\gamma}\bigg)\bigg)+O\bigg(\frac{1}{\varepsilon^2 t^{2\gamma}}\bigg)+O\bigg(\sum_{l=t_0}^{t-1}\exp\bigg(-\frac{b_0}{1-\gamma}(t+1)^{1-\gamma}\bigg)\\ &\quad\times\exp\bigg(\frac{b_0}{1-\gamma}l^{1-\gamma}\bigg)\frac{\beta}{\varepsilon^2 l^{2\gamma}}\bigg).\end{aligned} \tag{15}$$

Note that for large t_0, $\frac{\gamma}{b_0 t_0^{1-\gamma}} < \frac{1}{2}$. Then, we have

$$\begin{aligned}
&\sum_{l=t_0}^{t-1} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) \frac{\beta}{\varepsilon^2 l^{2\gamma}} \\
\leq& \int_{t_0}^{t} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) \frac{\beta}{\varepsilon^2 l^{2\gamma}} dl \\
=& \frac{1}{b_0} \int_{t_0}^{t} \frac{\beta}{\varepsilon^2 l^{\gamma}} d\left(\exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right)\right) \\
=& \frac{1}{b_0} \frac{\beta}{\varepsilon^2 l^{\gamma}} \left(\exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right)\right)\Big|_{t_0}^{t} - \frac{\beta}{\varepsilon^2 b_0} \int_{t_0}^{t} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) d\left(\frac{1}{l^{\gamma}}\right) \\
=& \frac{1}{b_0} \frac{\beta}{\varepsilon^2 l^{\gamma}} \left(\exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right)\right)\Big|_{t_0}^{t} + \frac{\gamma}{b_0} \int_{t_0}^{t} \frac{1}{l^{1-\gamma}} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) \frac{\beta}{\varepsilon^2 l^{2\gamma}} dl \\
\leq& \frac{1}{b_0} \frac{\beta}{\varepsilon^2 t^{\gamma}} \left(\exp\left(\frac{b_0}{1-\gamma} t^{1-\gamma}\right)\right) + \frac{\gamma}{b_0 t_0^{1-\gamma}} \int_{t_0}^{t} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) \frac{\beta}{\varepsilon^2 l^{2\gamma}} dl \\
\leq& \frac{1}{b_0} \frac{\beta}{\varepsilon^2 t^{\gamma}} \left(\exp\left(\frac{b_0}{1-\gamma} t^{1-\gamma}\right)\right) + \frac{1}{2} \int_{t_0}^{t} \exp\left(\frac{b_0}{1-\gamma} l^{1-\gamma}\right) \frac{\beta}{\varepsilon^2 l^{2\gamma}} dl.
\end{aligned}$$

Hence, $\sum_{l=t_0}^{t-1} \exp(\frac{b_0}{1-\gamma} l^{1-\gamma}) \frac{\beta}{\varepsilon^2 l^{2\gamma}} = O(\frac{1}{\varepsilon^2 t^{\gamma}} (\exp(\frac{b_0}{1-\gamma} t^{1-\gamma})))$. From (15) it follows that $\mathbb{E}[V(t+1)] = O(\frac{1}{\varepsilon^2 t^{\gamma}})$. Thus, the proof is completed. ∎

Remark 4.12 A differentially private decentralized algorithm for linear regression learning is studied in [38], where the t-step privacy-preserving analysis and estimation error bound were given. However, the estimation error bound is given by $O(t)$ or $O(\exp(t^{\alpha})), 0 \leq \alpha < 1$, which is not reasonable in practice. Moreover, Theorem 4.11 shows that the privacy index ε affects the convergence rate of the algorithm in the form of $O(\frac{1}{\varepsilon^2})$, and the convergence rate of the proposed algorithm is invariant with respect to the amount of added noise as goes to infinity.

Next, we will establish the asymptotic unbiasedness mean-square convergence of the DP-CI algorithm by using the stochastic approximation-type step-size conditions[5], which is given as follows.

Theorem 4.13 *If Assumptions* A1)–A4) *hold, and* $\sigma_t = \frac{\alpha(t-1)}{\varepsilon} \delta H_{\max}(t-1)$, *then the estimate sequence* $\{x_i(t)\}$ *given by* (3) *is asymptotically unbiased mean-square convergence to the true parameter* θ^*.

Proof By taking expectation on both sides of (8), and $\omega(t)$, $n(t)$ are zero-mean noise, it is obtained that

$$\mathbb{E}[\delta(t+1)] = [I - \alpha(t)(\mathcal{L} \otimes I_n + H(t)H^{\mathrm{T}}(t))]\mathbb{E}[\delta(t)]. \tag{16}$$

By Lemma 4.9 and taking norm operator on both sides of (16), it is obtained that

$$\|\mathbb{E}\big[\delta(t+1)\big]\| \leq (1 - \alpha(t)m)\|\mathbb{E}\big[\delta(t)\big]\|, \quad \forall t > T, \tag{17}$$

where $m = \lambda_{\min}(M)$. Since $\alpha(t) \to 0$ (Assumption A4)), there exists t_0 such that $\alpha(t_0) \leq \frac{1}{\lambda_{\max}(M)}$, $\forall t > t_0$. Continuing the recursion in (17), we have $\|\mathbb{E}[\delta(t)]\| \leq (\prod_{j=t_0}^{t-1}(1 - \alpha(j)m))\|\mathbb{E}[\delta(t_0)]\|$,

$t \geq t_0$. Then, from the inequality $1-a \leq \mathrm{e}^{-a}$, $0 \leq a \leq 1$, it is obtained that $\|\mathbb{E}[\delta(t)]\| \leq \mathrm{e}^{-m\sum_{j=t_0}^{t-1}\alpha(j)}\|\mathbb{E}[\delta(t_0)]\|$. From $m>0$ and the sum of the step-size to infinity, it follows that $\lim_{t\to\infty}\|\mathbb{E}[\delta(t)]\|=0$. Thus, the estimate sequence $\{x_i(t)\}$ given by (3) is asymptotically unbiased with respect to the true parameter θ^*, i.e., $\lim_{t\to\infty}\mathbb{E}[x_i(t)]=\theta^*$, $i\in\mathcal{V}$.

By Assumption A3) and $\sigma_t=\frac{\alpha(t-1)}{\varepsilon}\delta H_{\max}(t-1)$, from (12) it follows that

$$\begin{aligned}
&\mathbb{E}\big[V(t+1)\big]\\
\leq\ &\big[1-2b_0\alpha(t)+\Gamma_2\alpha^2(t)\big]\mathbb{E}\big[V(t)\big]\\
&+\alpha^2(t-1)\bigg(\frac{6Nn\delta^2H_{\max}^2(t-1)}{\varepsilon^2}\bigg)\\
&+3\alpha^2(t)\bigg(\gamma_0\sigma_\omega^2+\alpha^2(t-1)\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2\frac{2Nn\delta^2H_{\max}^2(t-1)}{\varepsilon^2}\bigg).
\end{aligned}\tag{18}$$

Noting that $b_0>0$, $\lim_{t\to\infty}\alpha(t)=0$, there exists $t_0>0$, such that $\Gamma_2\alpha(t)\leq b_0$, and $2b_0\alpha(t)\leq 1$, $\forall t>t_0$. Then, from Assumption A4) it follows that

$$\begin{aligned}
&0\leq 1-2b_0\alpha(t)+\Gamma_2\alpha^2(t)<1,\quad \forall t>t_0,\\
&\sum_{t=t_0}^{\infty}[2b_0\alpha(t)-\Gamma_2\alpha^2(t)]\geq b_0\sum_{t=t_0}^{\infty}\alpha(t)=\infty,\\
&\lim_{t\to\infty}\frac{1}{2b_0\alpha(t)-\Gamma_2\alpha^2(t)}\bigg(\frac{6\alpha^2(t-1)Nn\delta^2H_{\max}^2(t-1)}{\varepsilon^2}\\
&\quad+3\alpha^2(t)\bigg(\gamma_0\sigma_\omega^2+\alpha^2(t-1)\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2\times\frac{2Nn\delta^2H_{\max}^2(t-1)}{\varepsilon^2}\bigg)\bigg)=0.
\end{aligned}$$

This together with (18) and Lemma 4.10 implies the results of the theorem. ∎

Remark 4.14 From Theorem 4.13 it follows that the proposed algorithm is noise-resilient and provably convergent. However, as shown in [25, 26, 29, 30, 32, 33], the accuracy of the system performance is compromised by the added noise due to differential privacy.

According to Lemma 3.5, the privacy guarantee becomes weaker when t becomes larger. In particular, if the privacy-preserving in all iterations is considered, then the total privacy index will be ∞, which means that there is no privacy guarantee anymore. Therefore, when stochastic approximation-type step-size is used, the proposed algorithm is effective for differential privacy in finite number of iterations, which is consistent with Theorem 1 of [23]. The details are given in the following theorem.

Theorem 4.15 (Total privacy leakage) *For given $\varepsilon, T>0$, the DP-CI algorithm is εT-DP after T-times iteration.*

In the following, we will properly choose the step-size and added privacy noise to achieve the privacy-preserving for all iterations.

4.3 Convergence Analysis with an Exponentially Damping Step-Size

In this subsection, we suppose the form of the step-size is $\alpha(t)=cq^{t+1}$, where $c>0, 0<q<1$[29–33]. A similar result to Theorem 4.13 on the convergence property can also be established. However, instead of achieving exact mean-square convergence, the estimate of each agent i

approximately converges to the unknown parameter with an error proportional to the step-size parameters c, q. The following theorem formally states the result.

Theorem 4.16 *Suppose Assumptions* A1)–A3) *hold, and* $\sigma_t = cp^t\delta H_{\max}(t-1)$, $\alpha(t) = cq^{t+1}$, *where* $0 < c < \frac{\|\mathcal{L}\otimes I_n\|^2+\gamma_0^2}{b_0}$, $p \in (q,1)$. *Then, the estimate sequence* $x_i(t)$ *of the DP-CI algorithm satisfies*

$$\lim_{t\to\infty}\mathbb{E}\big[V(t)\big] \le \mathrm{e}^{\frac{2b_0c(1+q)-(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2}{1-q^2}}\big[\|\delta(0)\|^2\big] + \frac{c^2(6Nn\delta^2H_{\max}^2(t-1))}{1-p^2} + \frac{3c^2q^2\gamma_0\sigma_\omega^2}{1-q^2} + \frac{3c^4q^2(\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2(2Nn\delta^2H_{\max}^2(t-1))}{1-q^2p^2}.$$

Proof Following the same arguments as in Theorem 4.13. From (12) it follows that

$$\begin{aligned}&\mathbb{E}\big[V(t+1)\big]\\ &\le \big[1-2b_0cq^{t+1}+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2t+2}\big]\mathbb{E}\big[V(t)\big]\\ &\quad +c^2p^{2t}(6Nn\delta^2H_{\max}^2(t-1)) + 3c^2q^{2t+2}(\gamma_0\sigma_\omega^2\\ &\quad +c^2p^{2t}\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2(2Nn\delta^2H_{\max}^2(t-1)).\end{aligned}$$

Iterating the above inequality, it is obtained that

$$\begin{aligned}&\mathbb{E}\big[V(t)\big]\\ &\le \prod_{\iota=0}^{t}\big[1-2b_0cq^{\iota}+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2\iota}\big]V(0)\\ &\quad +\sum_{\varrho=0}^{t-2}\Big(c^2p^{2\varrho}(6Nn\delta^2H_{\max}^2(t-1)) + 3c^2q^{2\varrho+2}(\gamma_0\sigma_\omega^2\\ &\qquad +c^2p^{2\varrho}\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2(2Nn\delta^2H_{\max}^2(t-1))\Big)\\ &\quad \times\prod_{\iota=\varrho+1}^{t-1}\big[1-2b_0cq^{\iota+1}+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2\iota+2}\big] + c^2p^{2t-2}(6Nn\delta^2H_{\max}^2(t-1))\\ &\quad +3c^2q^{2t}(\gamma_0\sigma_\omega^2 + c^2p^{2t-2}\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2(2Nn\delta^2H_{\max}^2(t-1)).\end{aligned} \tag{19}$$

By utilizing $1-x \le \mathrm{e}^{-x}$, and when $t\to\infty$, we have

$$\begin{aligned}\prod_{\iota=0}^{\infty}\big[1-2b_0cq^{\iota}+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2\iota}\big] &\le \mathrm{e}^{\sum_{\iota=0}^{\infty}(2b_0cq^{\iota}-(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2\iota})}\\ &= \mathrm{e}^{\frac{2b_0c(1+q)-(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2}{1-q^2}}.\end{aligned} \tag{20}$$

From $c < \frac{\|\mathcal{L}\otimes I_n\|^2+\gamma_0^2}{b_0}$ it follows that $0 < 1-2b_0cq^t+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2t} < 1$. Therefore, $\prod_{\iota=\varrho+1}^{t-1}\big[1-2b_0cq^{\iota}+(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2q^{2\iota}\big] < 1$ holds. Furthermore, from (19) and (20) it follows that

$$\lim_{t\to\infty}\mathbb{E}\big[V(t)\big] \le \mathrm{e}^{\frac{2b_0c(1+q)-(2\|\mathcal{L}\otimes I_n\|^2+2\gamma_0^2)c^2}{1-q^2}}\big[\|\delta(0)\|^2\big] + \frac{c^2(6Nn\delta^2H_{\max}^2(t-1))}{1-p^2} + \frac{3c^2q^2\gamma_0\sigma_\omega^2}{1-q^2} + \frac{3c^4q^2(\|(\mathcal{L}\otimes I_n+\gamma_0 I)\|^2(2Nn\delta^2H_{\max}^2(t-1))}{1-q^2p^2}.$$

The proof is completed. ∎

Theorem 4.17 (Total privacy leakage) *For given $0 < q < p < 1$, the DP-CI algorithm is $\frac{p}{p-q}$-DP after all iterations.*

Proof Note that the privacy index of the DP-CI algorithm at each iteration is $\varepsilon_t = (\frac{q}{p})^t$. Then, according to Lemma 3.5, the results can be obtained. ∎

5 Simulation Example

In this section, we provide a numerical simulation to testify the effectiveness of distributed estimator based on the DP-CI algorithm proposed in this paper. Let $N = 5$, the adjacency matrix of the network is $A = [0, 1, 0, 1, 0; 1, 0, 1, 0, 1; 0, 1, 0, 1, 0; 1, 0, 1, 0, 1; 0, 1, 0, 1, 0]$, see Figure 1. The true parameter vector is $\theta^* = [-1, 1]^{\mathrm{T}}$, the observation matrices and the initial parameter estimates of these agents are chosen as: $H_1(t) = [1 + \sin(t),\ 0]$, $H_2(t) = [1 - \cos(t),\ 0]$, $H_3(t) = [0,\ 1 + \cos(t)]$, $H_4(t) = [1,\ 1 - \sin(t)]$, $H_5(t) = [1,\ 0.5\sin(t)]$, $x_i(0) = [0,\ 0.4]$, $i = 1, 2, \cdots, 5$.

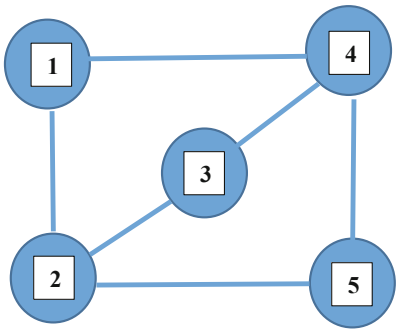


Figure 1 undirected interaction topology

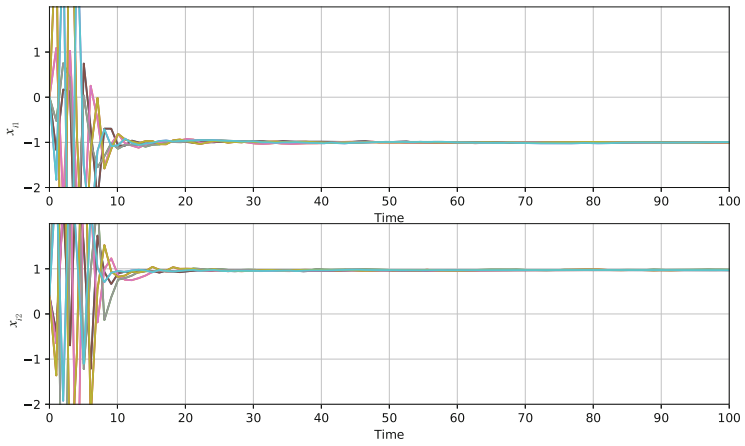
The noises $\omega_i(t)$ of agents are spatially independent white noises with uniform distribution on $[-0.2, 0.2]$.

First, we take the step-size $\alpha(t) = \frac{2}{t+2}$, the privacy index $\varepsilon = 0.8$ and $\delta = 0.2$. From (4) it follows that the scalar parameter σ_t of the added noises $n_i(t)$ is $\frac{3}{2(t+1)}$. Further, if we set the privacy index $\varepsilon = 0.4$, then from (4) it follows that the scalar parameter σ_t of the added noises $n_i(t)$ is $\frac{3}{t+1}$. Under the above setting, simulation result of the proposed algorithm with two privacy indices is given in Figure 2. As shown in Figure 2, the asymptotically unbiased estimate of the unknown parameter in mean-square is achieved by using the stochastic approximation-type step-size, but differential privacy-preserving holds for finite number of iterations. In addition, comparing (a) and (b) in Figure 2, we find that the smaller the privacy index (higher the privacy level), the slower the convergence rate of the algorithm.

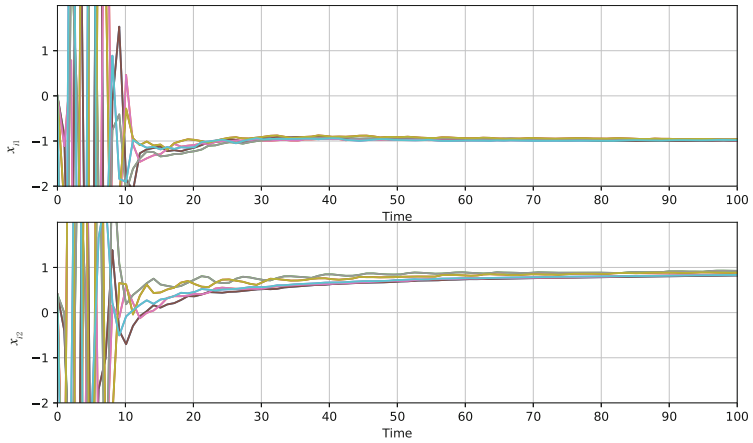
Second, we take the step-size $\alpha(t) = 0.4^{t+1}$, $\delta = 0.2$, and the scalar parameter σ_t of the added noises $n_i(t)$ is 0.6×0.8^t. From Theorem 4.17 it follows that the total privacy index is 2. As shown in Figure 3, the estimate sequence does not converge in mean-square to a

common value by using the exponentially damping step-size and privacy noises, but differential privacy-preserving holds for all iterations.

Based on the above analysis, the tradeoff between accuracy and privacy of the algorithm is shown, which is consistent with the theoretical analysis.



(a) $\varepsilon = 0.8$



(b) $\varepsilon = 0.4$

Figure 2 Trajectories of the estimates by a stochastic approximation-type step-size

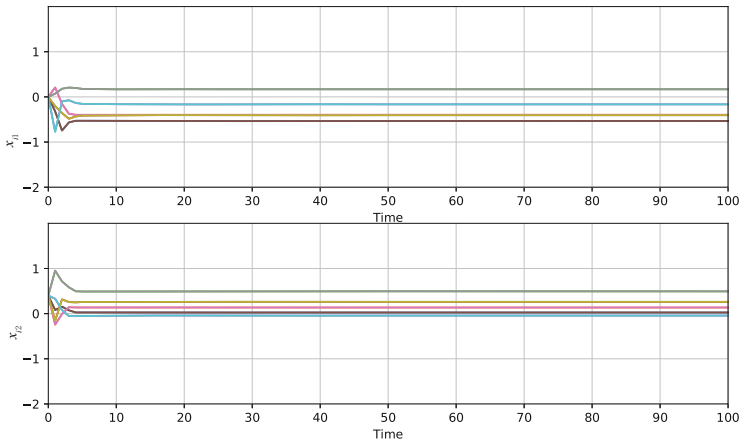


Figure 3 Trajectories of the estimates by an exponentially damping step-size

6 Conclusion

In this paper, the differentially private distributed parameter estimation has been studied. To protect the sensitive information of each agent and estimate the unknown parameter, a DP-CI algorithm has been designed. First, by using the stochastic approximation-type step-size conditions, the estimate sequence convergence in mean-square is guaranteed. The $T\varepsilon$-differential privacy and mean square convergence rate of the proposed algorithm are characterized. Then, the exponentially damping step-size and privacy noise for the DP-CI algorithm is given, which can ensure that the differential privacy-preserving holds for all iterations. The tradeoff between accuracy and privacy of the algorithm has been shown. Finally, a simulation example is given to verify the effectiveness of the proposed algorithm. We have found that it is not straightforward to extend the same techniques to differentially private resilient distributed parameter estimation when some agents are faulty, which makes designing differentially private resilient distributed parameter estimation algorithm a valuable and interesting future research direction.

References

[1] Wang Y, Zhao Y L, and Zhang J F, Distributed recursive projection identification with binary-valued observations, *Journal of Systems Science and Complexity*, 2021, **34**(5): 2048–2068.

[2] Lourenço I, Mattila R, Rojas C R, et al., Hidden Markov models: Inverse filtering, belief estimation and privacy protections, *Journal of Systems Science and Complexity*, 2021, **34**(5): 2048–2068.

[3] Ma X, Yi P, and Chen J, Distributed gradient tracking methods with finite data rates, *Journal of Systems Science and Complexity*, 2021, **34**(5): 1927–1952.

[4] Kar S and Moura J M F, Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs, *IEEE Journal of Selected Topics in Signal Processing*, 2021, **5**(4): 674–690.

[5] Zhang Q and Zhang J F, Distributed parameter estimation over unreliable networks with Markovian switching topologies, *IEEE Transactions on Automatic Control*, 2012, **57**(10): 2545–2560.

[6] Kar S, Moura J M F, and Ramanan K, Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication, *IEEE Transactions on Information Theory*, 2012, **58**(6): 3575–3605.

[7] Kar S, Moura J M F, and Poor H V, Distributed linear parameter estimation: Asymptotically efficient adaptive strategies, *SIAM Journal on Control and Optimization*, 2013, **51**(3): 2200–2229.

[8] You K, Xie L, and Song S, Asymptotically optimal parameter estimation with scheduled measurements, *IEEE Transactions on Signal Processing*, 2013, **61**(14): 3521–3531.

[9] Lin P and Qi H, Distributed gradient-based sampling algorithm for least-squares in switching multi-agent networks, *Science China Information Sciences*, 2020, **63**(9): 199203:1–199203:3.

[10] Chen Y, Kar S, and Moura J M F, Resilient distributed estimation: Exponential convergence under sensor attacks, *IEEE Conference on Decision and Control*, 2018, 7275–7282,

[11] Chen Y, Kar S, and Moura J M F, Resilient distributed estimation through adversary detection, *IEEE Transactions on Signal Processing*, 2018, **66**(9): 2455–2469.

[12] Xiao H C, Ding D R, Dong H L, et al., Adaptive event-triggered state estimation for large-scale systems subject to deception attacks, *Science China Information Sciences*, 2022, **65**: 122207:1–122207:16.

[13] Zhang J F, Tan J W, and Wang J M, Privacy security in control systems, *Science China Information Sciences*, 2021, **64**: 176201:1–176201:3.

[14] Farokhi F, Shames I, and Batterham N, Secure and private control using semi-homomorphic encryption, *Control Engineering Practice,* 2017, **67**: 13–20.

[15] Lu Y and Zhu M H, Privacy preserving distributed optimization using homomorphic encryption, *Automatica*, 2018, **96**: 314–325.

[16] Mo Y L and Murray R M, Privacy preserving average consensus, *IEEE Transactions on Automatic Control*, 2017, **62**(2): 753–765.

[17] Liu X K, Zhang J F, and Wang J M, Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems, *Automatica*, 2020, **122**: Article 109283.

[18] Dwork C, Differential privacy, *Proceedings of the* 33*rd International Colloquium on Automata, Languages and Programming,* 2006, 1–12.

[19] Dwork C, Rothblum G N, and Vadhan S, Boosting and differential privacy, *Proceedings of the* 51*st Annual IEEE Symposium on Foundations of Computer Science*, 2010, 51–60.

[20] Liang W J, Chen H, Zhang J, et al., An effective scheme for top-k frequent itemset mining under differential privacy conditions, *Science China Information Sciences*, 2020, **63**: 159101:1–159101:3.

[21] Xue Q, Zhu Y W, Wang J, et al., Locally differentially private distributed algorithms for set intersection and union, *Science China Information Sciences*, 2021, **64**: 219101:1–219101:3

[22] Ding J, Gong Y, Zhang C, et al., Optimal differentially private ADMM for distributed machine learning, arXiv preprint arXiv: 1901.02094v2, 2019.

[23] Li C, Zhou P, Xiong L, et al., Differentially private distributed online learning, *IEEE Transactions on Knowledge and Data Engineering*, 2018, **30**(8): 1440–1453.

[24] Zhu J L, Xu C Q, Guan J F, et al., Differentially private distributed online algorithms over

time-varying directed networks, *IEEE Transactions on Signal and Information Processing over Networks*, 2018, **4**(1): 4–17.

[25] Huang Z, Mitra S, and Vaidya N, Differentially private distributed optimization, *Proceedings of the* 16*th International Conference on Distributed Computing and Networking*, 2015, 4:1–4:10.

[26] Han S, Topcu U, and Pappas G J, Differentially private distributed constrained optimization, *IEEE Transactions on Automatic Control*, 2017, **62**(1): 50–64.

[27] Showkatbakhsh M, Karakus C, and Diggavi S, Differentially private consensus-based distributed optimization, arXiv preprint arXiv: 1903.07792v1, 2019.

[28] Wang Y, Huang Z, Mitra S, et al., Differential privacy in linear distributed control systems: Entropy minimizing mechanisms and performance tradeoffs, *IEEE Transactions on Control of Network Systems*, 2017, **4**(1): 118–130.

[29] Huang Z, Mitra S, and Dullerud G E, Differentially private iterative synchronous consensus, *Proceeding of CCS Workshop on Privacy in the Electronic Society*, USA, 2012.

[30] Gao L, Deng S, and Ren W, Differentially private consensus with event-triggered mechanism, *IEEE Transactions on Control of Network Systems*, 2019, **6**(1): 60–71.

[31] Wang A, Liao X F, and He H B, Event-triggered differentially private average consensus for multi-agent network, *IEEE/CAA Journal of Automatica Sinica*, 2019, **6**(1): 75–83.

[32] Nozari E, Tallapragada P, and Cortes J, Differentially private average consensus: Obstructions, trade-offs, and optimal algorithm design, *Automatica*, 2017, **81**: 221–231.

[33] Fiore D and Russo G, Resilient consensus for multi-agent systems subject to differential privacy requirements, *Automatica*, 2019, **106**: 18–26.

[34] Ny J L and Pappas G J, Differentially private filtering, *IEEE Transactions on Automatic Control*, 2014, **59**(2): 341–354.

[35] Katewa V, Chakrabortty A, and Gupta V, Differential privacy for network identification, *IEEE Transactions on Control of Network Systems*, 2020, **7**(1): 266–277.

[36] Song S, Chaudhuri K, and Sarwate A D, Stochastic gradient descent with differentially private updates, *Proceedings of the Global Conference on Signal and Information Processing*, 2013, 245–248.

[37] Liu Y, Liu J, and Başar T, Differentially private gossip gradient descent, *IEEE Conference on Decision and Control*, 2018, 2777–2782.

[38] Liu Y, Zhang X, Qin S, et al., Differentially private linear regression over fully decentralized datasets, 33*rd Conference on Neural Information Processing Systems* (*NeurIPS*), 2019.

[39] Hale M T and Egerstedt M, Cloud-enabled differentially private multiagent optimization with constraints, *IEEE Transactions on Control of Network Systems*, 2018, **5**(4): 1693–1706.

[40] Goodwin G and Sin K, *Adaptive Filtering, Prediction and Control*, Englewood Cliffs, N.J.: Prentice-Hall, 1984.

[41] Mehran M and Magnus E, *Graph Theoretic Methods in Multiagent Network*, Princeton: Princeton University Press, 2010.