Automatica 142 (2022) 110440

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Differentially private distributed algorithms for stochastic aggregative games[☆]

Jimin Wang[a], Ji-Feng Zhang[b,c,*], Xingkang He[d]

[a] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, China
[b] Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[d] Department of Electrical Engineering, University of Notre Dame, IN 46556, USA

ARTICLE INFO

Article history:
Received 13 December 2021
Received in revised form 17 April 2022
Accepted 27 April 2022
Available online xxxx

Keywords:
Differential privacy
Stochastic aggregative games
Distributed algorithms
Stochastic approximation

ABSTRACT

Designing privacy-preserving distributed algorithms for stochastic aggregative games is urgent due to the privacy issues caused by information exchange between players. This paper proposes two differentially private distributed algorithms seeking the Nash equilibrium in stochastic aggregative games. By adding time-varying random noises, the input and output-perturbation methods are given to protect each player's sensitive information. For the case of output-perturbation, utilizing mini-batch methods, the algorithm's mean square error is inversely proportional to the privacy level ϵ and the number of samples. For the case of input-perturbation, a differentially private distributed stochastic approximation-type algorithm is developed to achieve almost sure convergence and (ϵ, δ)-differential privacy. Under suitable consensus time conditions, the algorithm's convergence rate is rigorously presented for the first time, where the optimal convergence rate $O(1/k)$ in a mean square sense is obtained. Then, utilizing mini-batch methods, the influence of added privacy noise on the algorithm's performance is reduced, and the convergence rate of the algorithm is improved. Specifically, when the batch sizes and the number of consensus times at each iteration grow at a suitable rate, an exponential rate of convergence can be achieved with the same privacy level. Finally, a simulation example demonstrates the algorithms' effectiveness.

1. Introduction

Game theory is a standard tool for studying the interaction behavior of self-interested players and has attracted considerable attention due to its broad applicability and technical challenge (Hao & Cheng, 2021; Li & Marden, 2013; Liu & Krstic, 2011; Pang & Hu, 2021; Salehisadaghiani & Pavel, 2016, 2018; Ye & Hu, 2017). In several practical situations, the player's objective function depends on its strategies and the sum-total of the other players' strategies, e.g., Cournot competition in economics, charging control of plug-in electric vehicles (Grammatico et al., 2016) and route choice on a road network (Paccagnan et al., 2019). This game type is known as an aggregative game. Due to its broad applicability, recently distributed algorithms for aggregative games have attracted increasing attention (Koshal et al., 2016; Paccagnan et al., 2019; Parise et al., 2015; Shokri & Kebriaei, 2020). When the players' objective functions are uncertain and stochastic, distributed algorithms for stochastic game problems have been provided (Franci & Grammatico, 2020; Lei & Shanbhag, 2020; Lei & Shanbhag, 2020; Yousefian et al., 2016). In particular, a distributed variable sample-size scheme for stochastic aggregative games has been developed in Lei and Shanbhag (2020). However, the communication among players in such a distributed manner raises privacy issues when players contain sensitive information.

Privacy issues in stochastic aggregative games can be encountered in various domains. For example, in a Cournot oligopoly, firms, i.e., players, compete to supply a product in a market with a price-responsive demand aiming to maximize profit. The firm's profit depends on its production cost and the market price, whereas the latter depends on all firms' aggregative quantity of the product offered in the market. Both production cost and market price are uncertain in practical applications (Yousefian et al., 2016). The production cost information is business sensitive, and

[☆] The work is supported by National Key R&D Program of China under Grant 2018YFA0703800, National Natural Science Foundation of China under Grant 61877057. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Xue-Jun Xie under the direction of Editor Miroslav Krstic.

[*] Corresponding author at: Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.

E-mail addresses: jimwang@ustb.edu.cn (J. Wang), jif@iss.ac.cn (J.-F. Zhang), xhe9@nd.edu (X. He).

https://doi.org/10.1016/j.automatica.2022.110440
0005-1098/© 2022 Elsevier Ltd. All rights reserved.

adversaries can exploit information sequence transmitted among firms to infer the production cost information. In the electric vehicles market, coordinating the charging schedules can provide services beneficial to the grid's operation. Since the electricity price depends on the aggregate consumption, the electric vehicle charging problem is formulated as an aggregative game (Grammatico et al., 2016). However, when an adversary has potential access to all communications in distributed schemes, the sensitive information will leak to the adversary (Han et al., 2017). In the traffic congestion case, every driver pursues its interest, e.g., minimizing traveling time, and is affected by the others' choices via congestion, and then the traffic congestion problem is formulated as an aggregative game (Paccagnan et al., 2019). Note that the origins and destinations of the drivers can easily be used as the basis for inferring their activities. Hence, it is sensitive information that urgently needs to be protected (Dong et al., 2015). Based on the above discussion, there is a great need to develop privacy-preserving algorithms for stochastic aggregative games.

A systematic and comprehensive view on privacy-preserving of control systems is presented in Zhang et al. (2021), which highlights that the system's privacy-preserving method should not lose its original control goal. Some privacy-preserving approaches for the systems have been recently proposed (Altafini, 2020; Lu & Zhu, 2018; Mo & Murray, 2017; Zhang et al., 2019), relying on time-varying transformation (Altafini, 2020), homomorphic encryption (Lu & Zhu, 2018; Zhang et al., 2019), and adding artificial noise (Mo & Murray, 2017). To achieve homomorphic encryption, the public and private keys should be generated and distributed in advance, allowing for computations performed on encrypted data without requiring access to a decryption key. The computation results are encrypted and can be revealed only by the private key owner, e.g., a player or a third party. The computation increases greatly as the number of iterations and players increases. Homomorphic encryption methods often require a large amount of computation. While time-varying transformation-based methods have small computation loads, they are only suitable for specific systems. Among others, differential privacy is a well-known privacy notion and has applications in many domains (Dwork, 2006; Dwork & Roth, 2014). So far, differential privacy has attracted substantial attention throughout computer, control and communication science, including areas like distributed learning (Abadi et al., 2016; Huang et al., 2019; Li et al., 2018; Zhou & Tang, 2020), data mining (Liang et al., 2020), and control and estimation (Han et al., 2017; Huang et al., 2012; Liu et al., 2020; Nozari et al., 2017; Ny & Pappas, 2014). In particular, the differentially private stochastic gradient descent is studied in Song et al. (2013). Generally, the added privacy noise causes a significant performance loss on the stochastic gradient descent algorithm. To improve the algorithm's performance, the differentially private variance with reduced stochastic gradient descent is presented in Bassily et al. (2019) and Lee (2017). However, the convergence analysis is not provided in Lee (2017) and Song et al. (2013).

In the study of networked games, several privacy-preserving methods have been proposed to seek the Nash equilibrium and protect the sensitive information (Alshehri et al., 2019; Dong et al., 2015; Gade et al., 2020; Hsu et al., 2013; Shakarami et al., 2019; Ye et al., 2021). For aggregative games, a privacy-preserving distributed algorithm is designed to seek the Nash equilibrium (Gade et al., 2020; Ye et al., 2021). However, the algorithm cannot protect the privacy of the players' sensitive information when all their neighbors are hostile or there exist eavesdroppers (Gade et al., 2020). A differentially private distributed algorithm for aggregative games is given in Ye et al. (2021). The above literature only focus on the privacy-preserving algorithm for deterministic aggregative games. Since stochastic aggregative games play an important role in both theory and applications, it is essential to develop privacy-preserving distributed algorithms against potential malicious attackers, which however has not been well studied in the literature.

This paper proposes two privacy-preserving distributed algorithms seeking the equilibrium solution in stochastic aggregative games and achieving (ϵ, δ)-differential privacy. By adding random noise, both the input and output-perturbation methods are given to protect each player's sensitive information. The main contributions of this paper are summarized as follows:

(i) By utilizing the technique of differential privacy, privacy-preserving distributed algorithms are proposed to seek the Nash equilibrium in stochastic aggregative games. To the best of our knowledge, this is the first attempt to use the input and output-perturbation methods to consider privacy issues in stochastic aggregative games.

(ii) For the case of output-perturbation, i.e., adding the privacy noise to each player's estimate, we prove that the mean square error is uniformly upper bounded by a finite scalar which is proportional to step size α and inversely proportional to the privacy level ϵ, the number of samples.

(iii) For the case of input-perturbation, with stochastic approximation-type step-size conditions, rigorous convergence and privacy analysis of the algorithm are provided, showing that the algorithm is noise-resilient and provably convergent. To the best of our knowledge, this is the first result of a stochastic approximation-type algorithm for stochastic aggregative games even without privacy-preserving. Moreover, when the number of consensus times at each iteration grows suitably, the convergence rate with stochastic approximation-type step sizes is also given for the first time, where the optimal convergence rate $O(1/k)$ in a mean square sense is displayed. Utilizing the adaptive batch sizes method to reduce the influence of added privacy noise on the algorithm's performance, the exponential convergence rate of the algorithm is given for the same privacy level, which is the same order as in Lei and Shanbhag (2020).

The results of this paper are significantly different from the literature. Compared with Gade et al. (2020) and Ye et al. (2021), stochastic objective functions are considered. Compared with Lei and Shanbhag (2020), two privacy-preserving distributed algorithms are proposed to seek the equilibrium solution in stochastic aggregative games.

Notations. Throughout this paper, the following standard notations are used. $\mathbf{1}$ stands for the appropriate-dimensional vector with all elements being one. $\lceil x \rceil$ denotes the smallest integer greater than x for $x \in \mathbb{R}$. $\mathbb{R}^n$ denotes the set of n-dimensional real-valued vectors. x^T the transpose of x, where x is either a matrix or a vector. $\|x\|$ refers to Euclidean norm of the vector x. I, 0 are identity matrix and zero matrix with appropriate dimensions, respectively. The expectation of a random variable X is denoted by $\mathbb{E}[X]$. For sequences $f(k)$ and $g(k)$ with $k = 1, 2, \ldots$, $O(\cdot)$ is defined as $f(k) = O(g(k))$ if there exists a positive number A and c such that $|\frac{f(k)}{g(k)}| \leqslant A$ for any $k > c$.

2. Preliminaries and problem formulation

2.1. Game theory

Definition 2.1 (*Ye et al., 2021, A Normal Form Game*). A game in a normal form is defined as a triple $\varGamma = \{\mathcal{V}, \mathcal{X}, \tilde{f}\}$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of players, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \cdots \times \mathcal{X}_N$, with $\mathcal{X}_i$ denoting the strategy set of player i, and $\tilde{f} = (\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_N)$, with $\tilde{f}_i$ referring to the loss function of player i.

Definition 2.2 (Franci & Grammatico, 2020, *Stochastic Nash Equilibrium*)**.** Stochastic Nash equilibrium is a set of strategies on which no player can reduce its loss function by unilaterally deviating from its strategy, assuming that the strategies of the other players are fixed, i.e., a set of strategies $x^* = (x_i^*, x_{-i}^*) \in \mathcal{X}$ is a stochastic Nash equilibrium if for all $i \in \mathcal{V}$, $\mathbb{E}[\tilde{f}_i(x_i^*, x_{-i}^*, \xi_i)] \leqslant \mathbb{E}[\tilde{f}_i(x_i, x_{-i}^*, \xi_i)]$, $\forall x_i \in \mathcal{X}_i$, where $x_{-i} = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N]^T$ and $\xi_i : \Omega \to \mathbb{R}^{m_i}$ is the random vector.

Throughout this paper, we consider there exist functions $f_i(x_i, \bar{x}, \xi_i)$ such that $\tilde{f}_i(x_i, x_{-i}, \xi_i) = f_i(x_i, \bar{x}, \xi_i)$, where $\bar{x} = \sum_{i=1}^N x_i$ denotes the aggregate of all players' strategies.

2.2. Problem formulation

Suppose there are N players in a stochastic aggregative game, where player i tries to solve

$$\begin{aligned} \min_{x_i \in \mathbb{R}^n} \quad & \mathbb{E}[f_i(x_i, \bar{x}, \xi_i)], \\ \text{subject to} \quad & x_i \in \mathcal{X}_i \end{aligned}$$

where $\mathcal{X}_i$ is a non-empty, compact and convex set for $i \in \mathcal{V}$. Instead of relying on a central authority, players exchange the information via a communication graph modeled as an undirected time-varying graph $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k)$ comprising a non-empty player set $\mathcal{V} = \{1, 2, \ldots, N\}$ and an edge set $\mathcal{E}_k \subseteq \mathcal{V} \times \mathcal{V}$. $\mathcal{N}_{i,k} = \{j \in \mathcal{V}, (j, i) \in \mathcal{E}_k\}$ denotes the neighborhood of player i at time k and player i is assumed to be a neighbor of itself. $\mathcal{G}_k$ is called connected if for any pair of players (i_1, i_m), a path exists from i_1 to i_m consisting of edges $(i_1, i_2), (i_2, i_3), \ldots, (i_{m-1}, i_m)$.

The following assumptions are needed in this paper.

Assumption 2.1 (*Joint Connectivity*)**.** There exists a positive integer $\bar{z}$ such that $(\mathcal{V}, \bigcup_{t=1}^{\bar{z}} \mathcal{E}_{t+k})$ is connected for all nonnegative integer k, where $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ is the undirected communication graph at time t and $\mathcal{E}_t$ is the corresponding edge set at time t.

Assumption 2.2. Let $\mathcal{A}_k = [a_{ij,k}]_{i,j \in \mathcal{V}}$ be the weight matrix associated with $\mathcal{G}_k$, which satisfies the following conditions: (i) There exists a positive constant η such that $a_{ij,k} > \eta$ for $j \in \mathcal{N}_{i,k}$, $a_{ij,k} = 0$ for $j \notin \mathcal{N}_{i,k}$; (ii) $\mathcal{A}_k$ is doubly stochastic, i.e., $\mathbf{1}^T \mathcal{A}_k = \mathbf{1}^T$, $\mathcal{A}_k \mathbf{1} = \mathbf{1}$.

Setting $F_i(x_i, z) = \mathbb{E}[g_i(x_i, z, \xi_i)]$, $g_i(x_i, z, \xi_i) = \nabla_{x_i} f_i(x_i, z, \xi_i)$, for all $z \in \mathbb{R}^n$, the following assumptions from Lei and Shanbhag (2020) are also needed.

Assumption 2.3. $F_i(x_i, z)$ is Lipschitz continuous w.r.t. z for each $i \in \mathcal{V}$ and any fixed $x_i \in \mathcal{X}_i$, i.e., there exists a positive constant L_i such that for all $z_1, z_2 \in \mathbb{R}^n$, $\|F_i(x_i, z_1) - F_i(x_i, z_2)\| \leqslant L_i \|z_1 - z_2\|$.

Assumption 2.4. The mapping $\phi(x)$ is L_ϕ-Lipschitz continuous, i.e., $\|\phi(x) - \phi(y)\| \leqslant L_\phi \|x - y\|$, where $\phi(x) = [F_1^T(x_1, \bar{x}), F_2^T(x_2, \bar{x}), \ldots, F_N^T(x_N, \bar{x})]^T$.

Assumption 2.5 (*Strong Monotonicity*)**.** There exists a positive constant m such that for $x, y \in \mathcal{X}$, $(x-y)^T(\phi(x) - \phi(y)) \geqslant m\|x-y\|^2$.

A common approach for seeking the Nash equilibrium in stochastic aggregative games is employing a synchronous iterative gradient-based algorithm. Specifically, every player exchanges information with its neighbors and subsequently updates its iterative state and the estimate of the aggregative decisions simultaneously. The time-varying jointly connected network graph models the player's communications in time. Due to incomplete information at each time-step, a player i only has an estimate of $\bar{x}$ in contrast to the actual value. At the beginning of the $(k+1)$th iteration, player i receives the estimates $\hat{v}_{j,k}$ from its neighbors $j \in \mathcal{N}_{i,k}$. Using this information, player i aligns its intermediate estimate according to $v_{i,k} = \sum_{j \in \mathcal{N}_{i,k}} a_{ij,k} \hat{v}_{j,k}$ and repeats this consensus step $\tau_k = k + 1$ times, where $a_{ij,k}$ is the nonnegative weight that player i assigns to player j's estimate. Then, using the average estimate $v_{i,k}$ and its own iterative state $x_{i,k}$, player i updates its iterative state and estimate as follows:

$$\begin{aligned} x_{i,k+1} &= \Pi_{\mathcal{X}_i}\Big(x_{i,k} - \frac{\alpha}{S_k} \sum_{\iota=1}^{S_k} g_i(x_{i,k}, N v_{i,k}, \xi_i^\iota)\Big), \\ \hat{v}_{i,k+1} &= v_{i,k} + x_{i,k+1} - x_{i,k}, \end{aligned}$$

where α is the constant step size corresponding to the influence of the sampling gradients on the state update rule at each time-step, S_k is the number of the sampling gradients used at time k and ξ_i^ι, $\iota = 1, \ldots, S_k$ denote the realizations of ξ_i.

The above result demonstrates that the update rule allows players in the network to solve the Nash equilibrium seeking problem distributively. Players are required to exchange and expose information with their neighbors, leading to undesirable privacy leakages of their sensitive information. For example, if an adversary can access all shared information by eavesdropping on the communications among players and has arbitrary auxiliary information, then the adversary can perform attacks to infer the sensitive information of each player.

2.3. Attack model

This paper considers two types of passive adversaries: semi-honest players and eavesdroppers defined as follows.

Semi-honest (i.e., honest-but-curious) players are assumed to follow the Nash equilibrium seeking algorithm and perform the correct computations. However, they may collect all intermediate and input/output information to learn sensitive information about other players.

Eavesdroppers are external adversaries who steal information through wiretapping all communication channels and intercepting exchanged information between players.

2.4. Privacy leakage in the above algorithm

In the above algorithm, the adversary can always collect $\hat{v}_{i,k}$ and $v_{i,k}$ at each time k. Recall that the goal of differential privacy is to provide a strong privacy guarantee in the presence of any auxiliary information that an adversary may have. In the worst case, the adversary has the knowledge of any auxiliary information, such as $\mathcal{A}_k$, α, S_k and the structure of cost function $f_i(\cdot)$. Then, with the help of all the information, if there is no privacy-preserving mechanism, the adversary can infer the players' sampled gradient. Next, we show that the leakage of sampled gradient information can lead to privacy issues. For example, in the Cournot competition model, each player seeks to maximize its profit, or equivalently, minimize its cost. Player i's cost function is $f_i(x_i, \bar{x}, \xi_i) = c_i(x_i) - x_i(a - b\bar{x} + \xi_i)$, where $c_i(x_i)$ is its production cost-information (business sensitive) which is worth protecting, x_i is the amount of goods that player i produces, and ξ_i is the uncertainty that is fixed for each sampling point ι. For each sampling point ι, the gradient of player i's cost function is $g_i(x_i, \bar{x}, \xi_i^\iota) = c_i'(x_i) - (a - b\bar{x} + \xi_i^\iota) - bx_i$. If the sampled gradient is leaked, then the observation of $(x_i, c_i'(x_i))$ allows adversary to infer the private production cost-information $c_i(x_i)$ since it has known the structure of $f_i(\cdot)$. Therefore, direct communication of the intermediate results can lead to severe privacy leakage in the above algorithm. It is imperative to provide a theoretical privacy guarantee on the sensitive information in distributed algorithms for stochastic aggregative games.

2.5. Differential privacy

Before introducing the privacy-preserving distributed algorithm for stochastic aggregative games, we first present some preliminaries of differential privacy. The basic idea of differential privacy is to "perturb" the exact result before delivering. In this case, an adversary cannot tell from the output information with a high probability of whether the player's sensitive information has changed. Differential privacy-preserving is equivalent to hiding changes in the datasets. Formally, changes in the datasets can be defined by a symmetric binary relation between two datasets called adjacency relation, which is denoted by $\mathrm{Adj}(\cdot,\cdot)$. Two datasets D_k and D_k' that satisfy $\mathrm{Adj}(D_k, D_k')$ are called adjacent datasets. Inspired by Bassily et al. (2019), we first define adjacent relation, which indicates the specific sensitive information that needs to be protected.

Definition 2.3 (*Adjacent Relation*)**.**: Two different samples of gradient information are recorded as $D_k = \{\xi_i^{\iota}, \iota = 1,\dots,S_k\}$, $D_k' = \{\xi_i^{\iota'}, \iota' = 1,\dots,S_k\}$, where ξ_i^{ι} and $\xi_i^{\iota'}$ denote two realizations of ξ_i, if only one of the sampling points is different, then D_k and D_k' are called adjacent datasets.

Definition 2.4 (*Differential Privacy, Ny & Pappas, 2014*)**.** Given $\epsilon, \delta \geqslant 0$, a randomized algorithm $\mathcal{R}$ is (ϵ,δ)-differentially private if for all adjacent datasets D_k and D_k', and for any subsets of outputs $\Upsilon \subseteq \mathrm{Range}(\mathcal{R})$, such that

$$\mathbb{P}\{\mathcal{R}(D_k) \in \Upsilon\} \leqslant e^{\epsilon}\mathbb{P}\{\mathcal{R}(D_k') \in \Upsilon\} + \delta.$$

Remark 2.1. The constant ϵ measures the privacy level of the randomized algorithm $\mathcal{R}$, i.e., a small ϵ implies a high privacy level. ϵ is taken to be a small constant, e.g., $\epsilon \approx 0.1$, or perhaps even $\ln 2$ or $\ln 3$.

Definition 2.5 (*Han et al., 2017, Sensitivity*)**.** The sensitivity of an output map q at the kth iteration is defined as $\Delta_k = \sup_{D_k, D_k':\mathrm{Adj}(D_k,D_k')} \|q(D_k) - q(D_k')\|$, where D_k and D_k' are input datasets at time k.

Remark 2.2. A randomized algorithm $\mathcal{R}$ is normally defined in conjunction with some query q of interest. The sensitivity of an output map q captures the magnitude by which a single individual's data can change the output map q in the worst case, and it is a key quantity that determines where and how much noise is added in each iteration for achieving (ϵ,δ)-differential privacy.

Problem. In this paper, we use two different perturbation methods to design privacy-preserving distributed algorithms for stochastic aggregative games, which protects each player's sensitive information in the sense of (ϵ,δ)-differential privacy and guarantees the convergence to a unique Nash equilibrium.

3. Differentially private distributed algorithms for stochastic aggregative games: output-perturbation

We propose a differentially private distributed algorithm for stochastic aggregative games via output-perturbation, i.e., Algorithm 1.

Remark 3.1. Different from the existing distributed algorithms for stochastic games (Lei & Shanbhag, 2020; Yousefian et al., 2016), to ensure (ϵ,δ)-differential privacy of Algorithm 1, we add the Gaussian noise to the estimate $v_{i,k}$ when broadcasting it.

Next, we analyze the (ϵ,δ)-differential privacy and convergence property of Algorithm 1.

Algorithm 1 Differentially private distributed algorithms for stochastic aggregative games via output-perturbation

Initialization: Let $k = 0$, and $v_{i,0} = x_{i,0} \in \mathcal{X}_i$ for any $i \in \mathcal{V}$. Let $\alpha > 0$ and $\{S_k\}$ be deterministic sequences.

Message passing. Player i sends the noisy estimate $p_{i,k} = v_{i,k}+n_{i,k}$ to its neighbors, where $v_{i,k}$ is the estimate of player i at time k, and each element of $n_{i,k} \in \mathbb{R}^n$ is the zero-mean independent and identically distributed (*i.i.d.*) Gaussian noise with variance $\sigma_{i,k}^2$, i.e. $n_{i,k}^l \sim \mathcal{N}(0, \sigma_{i,k}^2)$, $l = 1, 2, \dots, n$.

Consensus. Player i receives the noisy estimate $p_{j,k} = v_{j,k} + n_{j,k}$ from its neighbors and conducts the following step by one time: $\hat{v}_{i,k} = \sum_{j\in\mathcal{N}_{i,k}} a_{ij,k}p_{j,k}$, $\forall i \in \mathcal{V}$.

Strategy update. For every $i \in \mathcal{V}$:

$$x_{i,k+1} = \Pi_{\mathcal{X}_i}\Big(x_{i,k} - \frac{\alpha}{S_k}\sum_{\iota=1}^{S_k} g_i(x_{i,k}, Nv_{i,k}, \xi_i^{\iota})\Big), \tag{1}$$

$$v_{i,k+1} = \hat{v}_{i,k} + x_{i,k+1} - x_{i,k}. \tag{2}$$

where S_k is the number of the sampling gradients used at time k and ξ_i^{ι}, $\iota = 1,\dots,S_k$ denote the realizations of ξ_i.

3.1. Privacy analysis of Algorithm 1

In this subsection, we will prove the (ϵ,δ)-differential privacy of Algorithm 1. As explained above, to protect privacy each player i generates a noisy estimate by adding a noise vector to the local estimate $v_{i,k}$, i.e., $p_{i,k} = v_{i,k} + n_{i,k}$. This method guaranteeing differential privacy is known as output-perturbation (Ny & Pappas, 2014). Next, we derive conditions on the noise variances under which Algorithm 1 satisfies (ϵ,δ)-differential privacy.

Assumption 3.1. $g_i(x_i, Nv_i, \xi_i^{\iota})$ are uniformly bounded, i.e., there exists a positive constant C such that for each fixed sampled point ι, $\|g_i(x_i, Nv_i, \xi_i^{\iota})\| \leqslant C$.

Remark 3.2. Assumption 3.1 is a technical requirement for privacy analysis. A similar assumption is also used in the literature (Li et al., 2018). If privacy protection is not considered, then Assumption 3.1 can be removed. For given x_i and Nv_i, if ξ_i is uniformly bounded, then, by the characteristics of compact set, Assumption 3.1 holds. For instance, if we choose $f_i(\cdot)$ as the simulation example of this paper, then Assumption 3.1 holds. If $g_i(\cdot)$ itself is a bounded function, e.g., $g_i(\cdot) = \sin(x_i, Nv_i, \xi_i)$, then Assumption 3.1 holds.

Lemma 3.1. *The sensitivity of Algorithm 1 at the kth iteration satisfies*

$$\Delta_k \leqslant \frac{2\alpha C}{S_{k-1}}. \tag{3}$$

Proof. Recall in Definition 2.3, that D_k and D_k' are any two data vectors differing in one entry. $v_{i,k}$ is computed based on dataset D_k, while $v_{i,k}'$ is computed based on the dataset D_k'. For adjacent datasets D_k and D_k', we have

$$\begin{aligned}
&\|v_{i,k} - v_{i,k}'\| \\
&= \|\hat{v}_{i,k-1} + x_{i,k} - x_{i,k-1} - \hat{v}_{i,k-1}' - x_{i,k}' + x_{i,k-1}'\| \\
&= || \sum_{j\in\mathcal{N}_{i,k}} a_{ij,k}p_{i,k-1} + x_{i,k} - x_{i,k-1} \\
&\quad - \sum_{j\in\mathcal{N}_{i,k}} a_{ij,k}p_{i,k-1} - x_{i,k}' + x_{i,k-1}'||
\end{aligned}$$

$$
\begin{aligned}
&= \|x_{i,k} - x_{i,k-1} - x'_{i,k} + x'_{i,k-1}\| \\
&\leqslant \frac{\alpha}{S_{k-1}} || - g_i(x_{i,k-1}, Nv_{i,k-1}, \xi_i^{\iota}) \\
&\quad + g_i(x_{i,k-1}, Nv_{i,k-1}, \xi_i^{\iota'})|| \leqslant \frac{2\alpha C}{S_{k-1}}.
\end{aligned}
$$

The proof is completed. □

Remark 3.3. Motivated by Li et al. (2018), we use the mini-batch method to process multiple samples at the same iteration. Most importantly, the mini-batch method has a significant advantage in guaranteeing differential privacy for Algorithm 1. Observing the proof of Lemma 3.1, we find that the parameter $\frac{1}{S_k}$ has an effect on the sensitivity of Algorithm 1.

Theorem 3.1. *Let $\epsilon \in (0, 1]$ be arbitrary, $n_{i,k}$ be the noise sampled from Gaussian mechanism with variance $\sigma_{i,k}^2$ where*

$$
\sigma_{i,k} = \frac{2\alpha C\sqrt{2\ln(1.25/\delta)}}{\epsilon S_{k-1}}, \tag{4}
$$

Then, each iteration of Algorithm 1 is (ϵ, δ)-differentially private. Specially, for any adjacent datasets D_k, D'_k, and any output $p_{i,k}$, the following inequality holds:

$$
\mathbb{P}\{p_{i,k}|D_k\} \leqslant e^{\epsilon}\mathbb{P}\{p_{i,k}|D'_k\} + \delta.
$$

Proof. In the context of differential privacy, the corresponding mechanism for Algorithm 1 maps $\{v_{i,k}, i \in \mathcal{V}\}$ to $\{p_{i,k}, i \in \mathcal{V}\}$. Since the Jacobian matrix of the linear transformation from $p_{i,k}$ to $n_{i,k}$ is the identity matrix, the privacy loss from $p_{i,k}$ is given as follows

$$
|\ln\frac{\mathbb{P}\{p_{i,k}|D_k\}}{\mathbb{P}\{p_{i,k}|D'_k\}}| = |\ln\frac{\mathbb{P}\{n_{i,k}|D_k\}}{\mathbb{P}\{n_{i,k}|D'_k\}}|. \tag{5}
$$

Furthermore, the entries of $n_{i,k}$ denoted by $n_{i,k}^l$ are independent of each other and for any entry l, we have

$$
|\ln\frac{\mathbb{P}\{n_{i,k}|D_k\}}{\mathbb{P}\{n_{i,k}|D'_k\}}| = |\ln\frac{\mathbb{P}\{n_{i,k}^l|D_k\}}{\mathbb{P}\{n_{i,k}^l|D'_k\}}|. \tag{6}
$$

From (5) and (6) it follows that

$$
\begin{aligned}
|\ln\frac{\mathbb{P}\{p_{i,k}|D_k\}}{\mathbb{P}\{p_{i,k}|D'_k\}}| &= |\ln\frac{\mathbb{P}\{n_{i,k}^l|D_k\}}{\mathbb{P}\{n_{i,k}^l|D'_k\}}| \\
&= |\ln\frac{\exp\bigl(-\frac{1}{2\sigma_{i,k}^2}(n_{i,k}^l)^2\bigr)}{\exp\bigl(-\frac{1}{2\sigma_{i,k}^2}(n_{i,k}^l+\Delta_k)^2\bigr)}| \\
&\leqslant \frac{1}{2\sigma_{i,k}^2}|n_{i,k}^l\Delta_k| + \frac{\Delta_k^2}{2\sigma_{i,k}^2}.
\end{aligned} \tag{7}
$$

From Lemma 3.1, and substituting (3) and (4) into (7), it is obtained that $|\ln\frac{\mathbb{P}\{p_{i,k}|D_k\}}{\mathbb{P}\{p_{i,k}|D'_k\}}| \leqslant \frac{S_{k-1}\epsilon^2}{4\alpha C\ln(1.25/\delta)}|n_{i,k}^l + \frac{\alpha C}{S_{k-1}}|$. When $|n_{i,k}^l| \leqslant \frac{\alpha C}{S_{k-1}}(4\epsilon^{-1}\ln(1.25/\delta) - 1)$, $|\ln\frac{\mathbb{P}\{p_{i,k}|D_k\}}{\mathbb{P}\{p_{i,k}|D'_k\}}|$ is bounded by ϵ. Next, we prove that

$$
\mathbb{P}\{|n_{i,k}^l| > r\} \leqslant \delta, \tag{8}
$$

where $r = \frac{\alpha C}{S_{k-1}}(4\epsilon^{-1}\ln(1.25/\delta) - 1)$. Furthermore, (8) is equivalent to

$$
\mathbb{P}\{n_{i,k}^l > r\} \leqslant \frac{\delta}{2}. \tag{9}
$$

Using the tail bound of the normal distribution $\mathcal{N}(0, \sigma_{i,k}^2)$, we have $\mathbb{P}\{n_{i,k}^l > r\} \leqslant \frac{\sigma_{i,k}}{\sqrt{2\pi}r}\exp(-\frac{r^2}{2\sigma_{i,k}^2})$. When δ is small ($\leqslant$ 0.01) and $0 < \epsilon \leqslant 1$, it is obtained that $\frac{\sigma_{i,k}}{r} < 1$, $-\frac{r^2}{2\sigma_{i,k}^2} < \ln(\sqrt{2\pi}\frac{\delta}{2})$. Therefore, (9) holds, which further implies that (8) holds.

Setting

$$
\begin{aligned}
\mathbb{P}_1 &= \{n_{i,k}^l : |n_{i,k}^l| \leqslant \tfrac{\alpha C}{S_{k-1}}(4\epsilon^{-1}\ln(1.25/\delta) - 1)\}, \\
\mathbb{P}_2 &= \{n_{i,k}^l : |n_{i,k}^l| > \tfrac{\alpha C}{S_{k-1}}(4\epsilon^{-1}\ln(1.25/\delta) - 1)\},
\end{aligned}
$$

we have

$$
\begin{aligned}
\mathbb{P}\{p_{i,k}|D_k\} &= \mathbb{P}\{v_{i,k}^l + n_{i,k}^l : n_{i,k}^l \in \mathbb{P}_1\} \\
&\quad + \mathbb{P}\{v_{i,k}^l + n_{i,k}^l : n_{i,k}^l \in \mathbb{P}_2\} \\
&\leqslant e^{\epsilon}\mathbb{P}\{p_{i,k}|D'_k\} + \delta.
\end{aligned}
$$

Hence, the statement of this theorem is obtained. □

3.2. Converge analysis of Algorithm 1

To facilitate the convergence analysis of Algorithms 1–2, we define $e_{i,k} = \frac{1}{S_k}\sum_{\iota=1}^{S_k} g_i(x_{i,k}, Nv_{i,k}, \xi_i^{\iota}) - F_i(x_{i,k}, Nv_{i,k})$, and the following stacked vectors:

$$
\begin{aligned}
&V_k = [v_{1,k}^T, \ldots, v_{N,k}^T]^T, P_k = [p_{1,k}^T, \ldots, p_{N,k}^T]^T, \\
&n_k = [n_{1,k}^T, \ldots, n_{N,k}^T]^T, e_k = [e_{1,k}^T, \ldots, e_{N,k}^T]^T, \\
&F(x_k, Nv_k) = [F_1^T(x_{1,k}, Nv_{1,k}), \ldots, F_N^T(x_{N,k}, Nv_{N,k})]^T, \\
&X_k = [x_{1,k}^T, \ldots, x_{N,k}^T]^T, X^* = [(x_1^*)^T, \ldots, (x_N^*)^T]^T.
\end{aligned} \tag{10}
$$

Then we rewrite (1)–(2) in the following form:

$$
\begin{aligned}
x_{i,k+1} &= \Pi_{\mathcal{X}_i}\bigl(x_{i,k} - \alpha(F_i(x_{i,k}, Nv_{i,k}) + e_{i,k})\bigr), \\
V_{k+1} &= \mathcal{A}_k P_k + X_{k+1} - X_k.
\end{aligned} \tag{11}
$$

Before discussing the convergence property of Algorithm 1, we give the following lemmas:

Lemma 3.2 (Koshal et al., 2016)**.** *If Assumptions 2.1–2.2 hold, then there exists a constant $\theta > 0$ and $\rho \in (0, 1)$ such that $\|[\Psi_{k,s}]_{i,j} - \frac{1}{N}\| \leqslant \theta\rho^{k-s}$, where $\Psi_{k,s} = \mathcal{A}_k\mathcal{A}_{k-1}\cdots\mathcal{A}_{s+1}\mathcal{A}_s$, and $[\Psi_{k,s}]_{i,j}$ denotes the (i, j)th entry of the matrix $\Psi_{k,s}$, $\forall k \geqslant s \geqslant 0$.*

Lemma 3.3. *If Assumption 2.2 holds, and $S_k = \lceil q^{-k}\rceil$, $q \in (0, 1)$, then*

$$
\mathbb{E}[\|\mathbf{1}^T V_k - \mathbf{1}^T X_k\|] \leqslant \tfrac{2CN\alpha\sqrt{2\ln(1.25/\delta)}(1-q^k)}{\epsilon(1-q)}.
$$

Proof. From the strategy update of Algorithm 1, for $t = 0, \ldots, k$, we have

$$
\begin{aligned}
&\mathbf{1}^T V_{t+1} \\
&= \sum_{i=1}^N v_{i,t+1} \\
&= \sum_{i=1}^N(\hat{v}_{i,t} + x_{i,t+1} - x_{i,t}) \\
&= \sum_{i=1}^N\Bigl(\sum_{j\in\mathcal{N}_{i,t}} a_{ij,t}(v_{j,t} + n_{i,t}) + x_{i,t+1} - x_{i,t}\Bigr) \\
&= \sum_{i=1}^N\Bigl(\sum_{j\in\mathcal{N}_{i,t}} a_{ij,t}v_{j,t} + \sum_{j\in\mathcal{N}_{i,t}} a_{ij,t}n_{j,t}\Bigr) + \sum_{i=1}^N x_{i,t+1} - \sum_{i=1}^N x_{i,t} \\
&= \sum_{i=1}^N v_{i,t} + \sum_{i=1}^N n_{i,t} + \sum_{i=1}^N x_{i,t+1} - \sum_{i=1}^N x_{i,t} \\
&= \mathbf{1}^T V_t + \mathbf{1}^T n_t + \mathbf{1}^T X_{t+1} - \mathbf{1}^T X_t.
\end{aligned} \tag{12}
$$

Note that $x_{i,0} = v_{i,0}$, $i \in \mathcal{V}$. Then, from (12) and running iterations, we have

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{1}^T V_{k+1} - \mathbf{1}^T X_{k+1}\|] \\
\leqslant\ &\mathbb{E}[\|\mathbf{1}^T V_k - \mathbf{1}^T X_k\|] + \mathbb{E}[\|\mathbf{1}^T n_k\|] \\
\leqslant\ &\mathbb{E}[\|\mathbf{1}^T V_0 - \mathbf{1}^T X_0\|] + \sum_{t=0}^{k} \mathbb{E}[\|\mathbf{1}^T n_t\|] \\
\leqslant\ &\sum_{t=0}^{k} \mathbb{E}[\|\mathbf{1}^T n_t\|].
\end{aligned} \tag{13}$$

From (4) and (13) it follows that $\mathbb{E}[\|\mathbf{1}^T V_{k+1} - \mathbf{1}^T X_{k+1}\|] \leqslant \frac{2CN\alpha\sqrt{2\ln(1.25/\delta)}(1-q^{k+1})}{\epsilon(1-q)}$. The proof is completed. □

Lemma 3.4. *If Assumptions 2.1–2.2 hold, and* $S_k = \lceil q^{-k} \rceil$, $\rho < q \in (0, 1)$, *then for each positive integer k, we have*

$$\begin{aligned}
&\mathbb{E}[V_k - \tfrac{1}{N}\mathbf{11}^T X_k||] \\
\leqslant\quad &N\theta M + \tfrac{2\sqrt{2\ln(1.25/\delta)}CNn\alpha\theta q}{\epsilon(q-\rho)} + \tfrac{2CN\sqrt{N}\alpha\theta}{(1-\rho)} \\
&+ C\sqrt{N}\alpha + \tfrac{2\sqrt{2\ln(1.25/\delta)}CN\alpha}{\epsilon(1-q)}.
\end{aligned}$$

Proof. From (11) it follows that

$$V_{k+1} = \mathcal{A}_k P_k + X_{k+1} - X_k. \tag{14}$$

By iterating computation (14), we have

$$\begin{aligned}
V_{k+1} &= \mathcal{A}_k\big(\mathcal{A}_{k-1}P_{k-1} + X_k - X_{k-1} + n_k\big) \\
&\quad + X_{k+1} - X_k \\
&= \Psi_{k,k-1}P_{k-1} + \Psi_{k,k}(X_k - X_{k-1}) \\
&\quad + \Psi_{k,k}n_k + X_{k+1} - X_k \\
&= \cdots \\
&= \Psi_{k,0}V_0 + \sum_{s=1}^{k} \Psi_{k,s}(X_s - X_{s-1}) \\
&\quad + \sum_{s=0}^{k} \Psi_{k,s}n_s + X_{k+1} - X_k,
\end{aligned} \tag{15}$$

where $\Psi_{k,s}$ are defined in Lemma 3.2. Then, from (15) it follows that

$$\begin{aligned}
V_k &= \Psi_{k-1,0}V_0 + \sum_{s=1}^{k-1} \Psi_{k-1,s}(X_s - X_{s-1}) \\
&\quad + \sum_{s=0}^{k-1} \Psi_{k-1,s}n_s + X_k - X_{k-1}.
\end{aligned} \tag{16}$$

Furthermore, from (14) it follows that

$$\begin{aligned}
&\mathbf{1}^T V_k \\
= &\mathbf{1}^T(\mathcal{A}_{k-1}P_{k-1} + X_k - X_{k-1}) \\
= &\mathbf{1}^T(V_{k-1} + n_{k-1} + X_k - X_{k-1}) \\
= &\mathbf{1}^T V_0 + \sum_{s=1}^{k} \mathbf{1}^T(X_s - X_{s-1}) + \sum_{s=0}^{k-1} \mathbf{1}^T n_s.
\end{aligned} \tag{17}$$

Then, from (16) and (17) it follows that

$$\begin{aligned}
&\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{11}^T X_k\|] \\
\leqslant\ &\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{11}^T V_k\|] + \mathbb{E}[\|\frac{1}{N}(\mathbf{11}^T V_k - \mathbf{11}^T X_k)\|] \\
\leqslant\ &\mathbb{E}[\|(\Psi_{k-1,0} - \frac{1}{N}\mathbf{11}^T)V_0 + \sum_{s=0}^{k-1}(\Psi_{k-1,s} - \frac{1}{N}\mathbf{11}^T)n_s \\
&+ \sum_{s=1}^{k-1}(\Psi_{k-1,s} - \frac{1}{N}\mathbf{11}^T)(X_s - X_{s-1}) \\
&+ (I - \frac{1}{N}\mathbf{11}^T)(X_k - X_{k-1})\|] \\
&+ \mathbb{E}[\|\frac{1}{N}(\mathbf{11}^T V_k - \mathbf{11}^T X_k)\|] \\
\leqslant\ &N\theta\rho^{k-1}\|V_0\| + \sum_{s=1}^{k-1} N\theta\rho^{k-1-s}\|X_s - X_{s-1}\| \\
&+ \|X_k - X_{k-1}\| + \sum_{s=0}^{k-1} N\theta\rho^{k-1-s}\mathbb{E}[\|n_s\|] \\
&+ \frac{2CN\alpha\sqrt{2\ln(1.25/\delta)}(1-q^k)}{\epsilon(1-q)},
\end{aligned} \tag{18}$$

where the first inequality follows from the Cauchy–Schwarz inequality, while the last inequality holds since $\|\Psi_{k,s} - \frac{1}{N}\mathbf{11}^T\| \leqslant \theta\rho^{k-s}$, $\forall k \geqslant s \geqslant 0$. Then, we estimate $\|X_s - X_{s-1}\|$. From (1) it follows that

$$\begin{aligned}
&\|x_{i,s+1} - x_{i,s}\| \\
=&\|\Pi_{\mathcal{X}_i}\Big(x_{i,s} - \frac{\alpha}{S_s}\sum_{\iota=1}^{S_s} g_i(x_{i,s}, Nv_{i,s}, \xi_i^{\iota})\Big) - \Pi_{\mathcal{X}_i}(x_{i,s})\| \\
\leqslant&\|x_{i,s} - \frac{\alpha}{S_s}\sum_{\iota=1}^{S_s} g_i(x_{i,s}, Nv_{i,s}, \xi_i^{\iota}) - x_{i,s}\|
\end{aligned}$$

where the last inequality holds by using the standard non-expansiveness property, i.e. $\|\Pi_{\mathcal{X}}\big(x\big) - \Pi_{\mathcal{X}}\big(y\big)\| \leqslant \|x - y\|$ for any x and y. Then, from Assumption 3.1 it follows that for $\forall i \in \mathcal{V}$, $\|x_{i,s+1} - x_{i,s}\| \leqslant \| - \frac{\alpha}{S_s}\sum_{\iota=1}^{S_s} g_i(x_{i,s}, Nv_{i,s}, \xi_i^{\iota})\| \leqslant C\alpha$. Thus, we have

$$\|X_s - X_{s-1}\| \leqslant C\sqrt{N}\alpha. \tag{19}$$

From (4), (18) and (19) it follows that

$$\begin{aligned}
&\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{11}^T X_k\|] \\
\leqslant\ &N\theta\rho^{k-1}\|V_0\| + \frac{2\sqrt{2\ln(1.25/\delta)}CNn\alpha\theta(q^{k-1} - \rho^{k-1})}{\epsilon(q-\rho)} \\
&+ \frac{CN\sqrt{N}\alpha\theta(1-\rho^{k-1})}{1-\rho} + C\sqrt{N}\alpha \\
&+ \frac{2\sqrt{2\ln(1.25/\delta)}CN\alpha(1-q^k)}{\epsilon(1-q)} \\
\leqslant\ &N\theta M + \frac{2\sqrt{2\ln(1.25/\delta)}CNn\alpha\theta q}{\epsilon(q-\rho)} + \frac{CN\sqrt{N}\alpha\theta}{1-\rho} \\
&+ C\sqrt{N}\alpha + \frac{2\sqrt{2\ln(1.25/\delta)}CN\alpha}{\epsilon(1-q)}.
\end{aligned} \tag{20}$$

Note that $V_0 = X_0 \in \mathcal{X}$. Then, there exists a constant $M = \sum_{i=1}^{N} \max_{x_i \in \mathcal{X}_i} \|x_i\|$ such that the last inequality in (20) holds. □

Define the σ-algebra $\mathcal{F}_k = \sigma\{X_t, n_t, 0 \leqslant t \leqslant k\}$. We need the following assumption on e_k.

Assumption 3.2. There exist positive constants $c_{i,1}$ and $c_{i,2}$ such that for any $k \geqslant 0$, $i \in \mathcal{V}$, $\mathbb{E}[e_{i,k}|\mathcal{F}_k] = 0$, $\mathbb{E}[\|e_{i,k}\|^2|\mathcal{F}_k] \leqslant \frac{c_{i,1}^2\|x_{i,k}\|^2 + c_{i,2}^2}{S_k}$.

Remark 3.4. When $S_k = 1$, there exists some $\nu > 0$ such that $\mathbb{E}[\|e_k\|^2|\mathcal{F}_k] \leqslant \nu^2$, $\forall k \geqslant 0$, which is also used in Yousefian et al. (2016).

Next, we establish the mean square error of Algorithm 1, which is given as follows.

Theorem 3.2. *If Assumptions 2.1–2.5 and 3.1–3.2 hold, $S_k = \lceil q^{-k} \rceil$, $\rho < q \in (0,1)$, and the constant step size α satisfies $0 < 2\alpha m - \alpha^2(1+4L_\phi^2) < 1$, then for any $\epsilon, \delta > 0$ and $\sigma_{i,k}$ in (4), the sequence X_k generated by Algorithm 1 satisfies the following form:*

$$\lim_{k\to\infty} \mathbb{E}[\|X_k - X^*\|^2] \leqslant \frac{\Xi}{2\alpha m - \alpha^2(1+4L_\phi^2)}, \tag{21}$$

where $\Xi = 4\alpha^2 N^2 \max_{i\in\mathcal{V}}\{L_i^2\}\kappa + 4\alpha NM \max_{i\in\mathcal{V}}\{L_i\}\beta + (1+2\alpha^2)\bar{c}_1^2$.

Proof. From Lemma 1 in Lei and Shanbhag (2020), we know that $x^* \in \mathcal{X}$ is a Nash equilibrium if and only if x^* satisfies $x^* = \Pi_{\mathcal{X}}(x^* - \alpha\phi(x^*))$. From (11) and the non-expansiveness property of the projection operator, we have

$$\begin{aligned}
&\|x_{i,k+1} - x_i^*\|^2 \\
\leqslant\ & \|x_{i,k} - x_i^* - \alpha(F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*) + e_{i,k})\|^2 \\
=\ & \|x_{i,k} - x_i^*\|^2 + \alpha^2\|F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*)\|^2 \\
& - 2\alpha(x_{i,k} - x_i^*)^T(F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*)) \\
& + \alpha^2\|e_{i,k}\|^2 + 2\alpha^2(F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*))^T e_{i,k} \\
& - 2\alpha(x_{i,k} - x_i^*)^T e_{i,k}.
\end{aligned} \tag{22}$$

By using $\pm 2ab \leqslant a^2 + b^2$, it is obtained that

$$\begin{aligned}
&\|x_{i,k+1} - x_i^*\|^2 \\
\leqslant\ & (1+\alpha^2)\|x_{i,k} - x_i^*\|^2 + (1+2\alpha^2)\|e_{i,k}\|^2 \\
& + 2\alpha^2\|F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*)\|^2 \\
& - 2\alpha(x_{i,k} - x_i^*)^T(F_i(x_{i,k}, Nv_{i,k}) - \phi_i(x^*)).
\end{aligned}$$

Then, summing the above inequality over $i = 1, \ldots, N$ leads to

$$\begin{aligned}
&\|X_{k+1} - X^*\|^2 \\
\leqslant\ & (1+\alpha^2)\|X_k - X^*\|^2 + (1+2\alpha^2)\|e_k\|^2 \\
& + 2\alpha^2\|F(x_k, Nv_k) - \phi(x^*)\|^2 \\
& - 2\alpha(X_k - X^*)^T(F(x_k, Nv_k) - \phi(x^*)).
\end{aligned} \tag{23}$$

By $(a+b)^2 \leqslant 2(a^2+b^2)$, setting $y_k = \sum_{i=1}^N x_{i,k}$, from Assumptions 2.3–2.4, we have

$$\begin{aligned}
& 2\alpha^2\|F(x_k, Nv_k) - \phi(x^*)\|^2 \\
\leqslant\ & 4\alpha^2(\|F(x_k, Nv_k) - F(x_k, y_k)\|^2 \\
& + \|F(x_k, y_k) - \phi(x^*)\|^2) \\
\leqslant\ & 4\alpha^2 N^2 \max_{i\in\mathcal{V}}\{L_i^2\}\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|^2 \\
& + 4\alpha^2 L_\phi^2\|X_k - X^*\|^2.
\end{aligned} \tag{24}$$

Moreover, we have

$$\begin{aligned}
& -2\alpha(X_k - X^*)^T(F(x_k, Nv_k) - \phi(x^*)) \\
=\ & -2\alpha(X_k - X^*)^T(F(x_k, y_k) - \phi(x^*)) \\
& +2\alpha(X_k - X^*)^T(F(x_k, y_k) - F(x_k, Nv_k)).
\end{aligned}$$

From Assumption 2.5 it follows that

$$\begin{aligned}
& -2\alpha(X_k - X^*)^T(F(x_k, Nv_k) - \phi(x^*)) \\
\leqslant\ & -2\alpha m\|X_k - X^*\|^2 \\
& +2\alpha(X_k - X^*)^T(F(x_k, y_k) - F(x_k, Nv_k)).
\end{aligned}$$

Furthermore, from Assumption 2.3, and using the Cauchy–Schwarz inequality it follows that

$$\begin{aligned}
& -2\alpha(X_k - X^*)^T(F(x_k, Nv_k) - \phi(x^*)) \\
\leqslant\ & -2\alpha m\|X_k - X^*\|^2 \\
& +2\alpha N \max_{i\in\mathcal{V}}\{L_i\}\|X_k - X^*\|\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|.
\end{aligned} \tag{25}$$

From (23)–(25) it follows that

$$\begin{aligned}
& \mathbb{E}[\|X_{k+1} - X^*\|^2|\mathcal{F}_k] \\
\leqslant\ & (1+\alpha^2)\|X_k - X^*\|^2 + (1+2\alpha^2)\mathbb{E}[\|e_k\|^2|\mathcal{F}_k] \\
& + 2\alpha^2\mathbb{E}[\|F(x_k, Nv_k) - \phi(x^*)\|^2|\mathcal{F}_k] \\
& - 2\alpha\mathbb{E}[(X_k - X^*)^T(F(x_k, Nv_k) - \phi(x^*))|\mathcal{F}_k] \\
\leqslant\ & (1 - 2\alpha m + \alpha^2 + 4\alpha^2 L_\phi^2)\|X_k - X^*\|^2 \\
& + 4\alpha^2 N^2 \max_{i\in\mathcal{V}}\{L_i^2\}\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|^2|\mathcal{F}_k] \\
& + 4\alpha NM \max_{i\in\mathcal{V}}\{L_i\}\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\||\mathcal{F}_k] \\
& + (1+2\alpha^2)\mathbb{E}[\|e_k\|^2|\mathcal{F}_k].
\end{aligned} \tag{26}$$

Similar to the proof of Lemmas 3.3 and 3.4, we have

$$\begin{aligned}
& \mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|^2] \\
\leqslant\ & 4N^2\theta^2 M^2 + \frac{32\ln(1.25/\delta)C^2N^2n^2\alpha^2\theta^2q^2}{\epsilon^2(q^2-\rho^2)} + 16C^2N\alpha^2 \\
& + \frac{32C^2N^3\alpha^2\theta^2}{(1-\rho)^2} + \frac{32\ln(1.25/\delta)C^2N^2\alpha^2}{\epsilon^2(1-q^2)}.
\end{aligned} \tag{27}$$

From Assumption 3.2, $S_k \geqslant q^{-k} \geqslant 1$ and $\mathcal{X}_i$ is a compact set, it follows that

$$\mathbb{E}[\|e_k\|^2|\mathcal{F}_k] \leqslant \frac{\bar{c}_1^2}{S_k} \leqslant \bar{c}_1^2, \tag{28}$$

where $\bar{c}_1^2 = \max_{1\leqslant i\leqslant n}(c_{i,1}^2\|x_{i,k}\|^2 + c_{i,2}^2)$. From (26)–(28) and Lemma 3.4 it follows that

$$\begin{aligned}
& \mathbb{E}[\|X_{k+1} - X^*\|^2] \\
\leqslant\ & (1 - 2\alpha m + \alpha^2(1+4L_\phi^2))\mathbb{E}[\|X_k - X^*\|^2] \\
& + 4\alpha^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\kappa + 4\alpha NM\max_{i\in\mathcal{V}}\{L_i\}\beta + (1+2\alpha^2)\bar{c}_1^2.
\end{aligned}$$

where $\kappa = 4N^2\theta^2M^2 + \frac{32\ln(1.25/\delta)C^2N^2n^2\alpha^2\theta^2q^2}{\epsilon^2(q^2-\rho^2)} + 16C^2N\alpha^2 + \frac{32C^2N^3\alpha^2\theta^2}{(1-\rho)^2} + \frac{32\ln(1.25/\delta)C^2N^2\alpha^2}{\epsilon^2(1-q^2)}$, $\beta = N\theta M + \frac{2CN\sqrt{N}\alpha\theta}{1-\rho} + \frac{2\sqrt{2\ln(1.25/\delta)}CNn\alpha\theta q}{\epsilon(q-\rho)} + C\sqrt{N}\alpha + \frac{2\sqrt{2\ln(1.25/\delta)}CN\alpha}{\epsilon(1-q)}$. Based on the step size condition and iterating the above process, (21) is obtained. The proof is completed. □

Remark 3.5. In Theorem 3.2, the convergence to a neighborhood of the Nash equilibrium is achieved. From (21) it follows that the mean square error of Algorithm 1 is proportional to the step size α and inversely proportional to the privacy level ϵ (the number of samples), revealing a trade-off between accuracy and privacy. A tighter upper bound of the mean square error can be obtained by optimizing the right-hand side of (21) over constants α, q and ϵ.

4. Differentially private distributed algorithms for stochastic aggregative games: input-perturbation

Different from Algorithm 1, we present differentially private distributed algorithms for stochastic aggregative games via the input-perturbation method, i.e., gradient-perturbation. As shown below, the algorithm achieves privacy-preserving of the player's sensitive information with guaranteed convergence.

4.1. Privacy analysis of Algorithm 2

In Algorithm 2, the privacy noise $n_{i,k}$ is added directly to the gradient. Thus, we compute the sensitivity based on Adj (D_k, D_k'). From Assumption 3.1 it follows that the sensitivity of Algorithm 2 is $\Delta_k = \|g_i(x_{i,k}, Nv_{i,k}, \xi_i^\iota) - g_i(x_{i,k}, Nv_{i,k}, \xi_i^{\iota'})\| \leqslant 2C$. Then, we have the following theorem.

Algorithm 2 Differentially private distributed algorithms for stochastic aggregative games via input-perturbation

Initialization: Set $k = 0$, and $\hat{v}_{i,0} = x_{i,0} \in \mathcal{X}_i$ for any $i \in \mathcal{V}$. Let $\alpha > 0$, $\{\tau_k\}$ and $\{S_k\}$ be deterministic sequences.
Consensus. $v_{i,k} := \hat{v}_{i,k}$ for any $i \in \mathcal{V}$ and repeat the following update by τ_k times: $v_{i,k} = \sum\limits_{j\in\mathcal{N}_{i,k}} a_{ij,k} v_{j,k}, \forall i \in \mathcal{V}$.
Strategy update. For every $i \in \mathcal{V}$:

$$\begin{aligned} x_{i,k+1} &= \Pi_{\mathcal{X}_i}\big(x_{i,k} - \tfrac{\alpha_k}{S_k}(\sum_{\iota=1}^{S_k} g_i(x_{i,k}, N v_{i,k}, \xi_i^\iota) + n_{i,k})\big), \\ \hat{v}_{i,k+1} &= v_{i,k} + x_{i,k+1} - x_{i,k}, \end{aligned}$$

where S_k is the number of the sampling gradients used at time k and ξ_i^ι, $\iota = 1, \ldots, S_k$ denote the realizations of ξ_i.

Theorem 4.1. *Let $\epsilon \in (0, 1]$ be arbitrary, $n_{i,k}$ be the noise sampled from Gaussian mechanism with variance $\sigma_{i,k}^2$ where*

$$\sigma_{i,k} = \frac{2C\sqrt{2\ln(1.25/\delta)}}{\epsilon}. \tag{29}$$

Then, each iteration of Algorithm 2 is (ϵ, δ)-differentially private.

Proof. Similar to the proof process of Theorem 3.1, we have the randomized mechanism $\mathcal{R}(D_k) = \sum_{\iota=1}^{S_k} g_i(x_{i,k}, N v_{i,k}, \xi_i^\iota) + n_{i,k}$ is (ϵ, δ)-differential privacy. Then, from the strategy update of Algorithm 2 it follows that $\hat{v}_{i,k+1} = v_{i,k} - x_{i,k} + \Pi_{\mathcal{X}_i}\big(x_{i,k} - \frac{\alpha_k}{S_k}(\sum_{\iota=1}^{S_k} g_i(x_{i,k}, N v_{i,k}, \xi_i^\iota) + n_{i,k})\big)$. Therefore, the estimate $\hat{v}_{i,k+1}$ is a function of $\mathcal{R}(D_k)$ and accesses the private dataset only indirectly via the output of a differentially private mechanism. As shown in Theorem 1 of Ny and Pappas (2014), differential privacy is robust to post-processing. Therefore, the privacy guarantee cannot be weakened, and further, (ϵ, δ)-differential privacy is preserved.

4.2. Convergence analysis of algorithm 2

Next, we establish the almost sure convergence of Algorithm 2 by using the following stochastic approximation-type conditions.

Assumption 4.1. The step size $\{\alpha_k\}_{k\geqslant 0}$ satisfies the following conditions: (i) (non-increasing) $0 \leqslant \alpha_{k+1} \leqslant \alpha_k \leqslant 1$, $\forall k \geqslant 0$; (ii) (non-summable) $\sum_{k\geqslant 0} \alpha_k = \infty$; (iii) (square-summable) $\sum_{k\geqslant 0} \alpha_k^2 < \infty$.

For example, Assumption 4.1 is satisfied for the step size of the form $\alpha_k = (k+1)^{-\gamma}$, where $\frac{1}{2} < \gamma \leqslant 1$.

For stochastic approximation-type sizes, instead of Assumption 2.5, we consider the following assumption.

Assumption 4.2 (*Strictly Monotonicity*). For $x, y \in \mathcal{X}, x \neq y$,

$$(x-y)^T(\phi(x) - \phi(y)) > 0.$$

Lemma 4.1 (Koshal et al., 2016). *Let ς_k be a non-negative scalar sequence. If $\sum_{k=0}^{\infty} \varsigma_k < \infty$ and $0 < \rho < 1$, then $\sum_{k=0}^{\infty}(\sum_{s=0}^{k} \rho^{k-s}\varsigma_s) < \infty$.*

Lemma 4.2 (Koshal et al., 2016). *If Assumptions 2.1–2.2 hold, $\mathbf{1}^T\widehat{V}_k = \mathbf{1}^T X_k$ for all $k \geqslant 0$.*

Lemma 4.3. *If Assumptions 2.1–2.2 and consensus times $\tau_k = 1$ hold, then $\sum_{k=0}^{\infty} \alpha_k \mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| | \mathcal{F}_k] < \infty$ for all $k \geqslant 0$.*

Proof. Similar to (16) and (17), from Lemmas 3.2 and 4.2 it follows that

$$\begin{aligned} &\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| \\ &\leqslant \|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T\widehat{V}_k\| + \|\frac{1}{N}(\mathbf{1}\mathbf{1}^T\widehat{V}_k - \mathbf{1}\mathbf{1}^T X_k)\| \\ &\leqslant \|(\Psi_{k,0} - \frac{1}{N}\mathbf{1}\mathbf{1}^T)\widehat{V}_0 + \sum_{s=1}^{k}(\Psi_{k,s} - \frac{1}{N}\mathbf{1}\mathbf{1}^T)(X_s - X_{s-1})\| \\ &\leqslant \|\Psi_{k,0} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\|\|\widehat{V}_0\| + \sum_{s=1}^{k}\|\Psi_{k,s} - \frac{1}{N}\mathbf{1}\mathbf{1}^T\|\|X_s - X_{s-1}\| \\ &\leqslant \theta\rho^k\|\widehat{V}_0\| + \sum_{s=1}^{k}\theta\rho^{k-s}\|X_s - X_{s-1}\|. \end{aligned} \tag{30}$$

Utilizing the strategy update of Algorithm 2, we obtain

$$\begin{aligned} &\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| | \mathcal{F}_k] \\ &\leqslant \theta\rho^k\|\widehat{V}_0\| + \sum_{s=1}^{k}\theta\rho^{k-s}\mathbb{E}[\|X_s - X_{s-1}\| | \mathcal{F}_k] \\ &\leqslant \theta\rho^k\|\widehat{V}_0\| + (\frac{2CNn\theta\sqrt{2\ln(1.25/\delta)}}{\epsilon} + CN\theta)\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1} \\ &\leqslant \theta\rho^k M + (\frac{2CNn\theta\sqrt{2\ln(1.25/\delta)}}{\epsilon} + CN\theta)\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1}. \end{aligned} \tag{31}$$

Note that $\widehat{V}_0 = X_0 \in \mathcal{X}$. Then, there exists a constant $M = \sum_{i=1}^{N}\max_{x_i\in\mathcal{X}_i}\|x_i\|$ such that the last inequality in (31) holds.

Next, we establish the convergence of $\sum_{k=0}^{\infty}\alpha_k\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| | \mathcal{F}_k]$. From (31) it follows that

$$\begin{aligned} &\sum_{k=0}^{\infty}\alpha_k\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| | \mathcal{F}_k] \\ &\leqslant (\frac{2\sqrt{2\ln(1.25/\delta)}CNn\theta}{\epsilon} + CN\theta)\sum_{k=0}^{\infty}\alpha_k\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1} \\ &\quad + \theta M\sum_{k=0}^{\infty}\alpha_k\rho^k. \end{aligned} \tag{32}$$

Now, we show that each term on the right side of (32) is summable, hence $\sum_{k=0}^{\infty}\alpha_k\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\| | \mathcal{F}_k] < \infty$.

Noting that $\alpha_k \leqslant \alpha_0$, $k \in \mathbb{N}$, $0 < \rho < 1$, we have $\sum_{k=0}^{\infty}\alpha_k\rho^k \leqslant \alpha_0\sum_{k=0}^{\infty}\rho^k < \infty$. Moreover, since $\alpha_k \leqslant \alpha_s$, $\forall k \geqslant s$, for the series $\sum_{k=0}^{\infty}\alpha_k\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1}$, we have

$$\begin{aligned} \sum_{k=0}^{\infty}\alpha_k(\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1}) &= \sum_{k=0}^{\infty}\sum_{s=1}^{k}\rho^{k-s}\alpha_k\alpha_{s-1} \\ &\leqslant \sum_{k=0}^{\infty}\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1}^2. \end{aligned}$$

From Lemma 4.1 and $\sum_{k=0}^{\infty}\alpha_k^2 < \infty$, it is obtained that $\sum_{k=0}^{\infty}\alpha_k\big(\sum_{s=1}^{k}\rho^{k-s}\alpha_{s-1}\big) < \infty$. Hence, the proof is completed. □

Lemma 4.4 (Polyak, 1987). *Let $Z_k, u_k, \beta_k, \zeta_k$ be non-negative random variables adapted to σ-algebra $\mathcal{F}_k$. If $\sum_{k=0}^{\infty} u_k < \infty$, $\sum_{k=0}^{\infty}\beta_k < \infty$, and $\mathbb{E}[Z_{k+1}|\mathcal{F}_k] \leqslant (1+u_k)Z_k - \zeta_k + \beta_k$ for all $k \geqslant 0$, then Z_k converges almost surely and $\sum_{k=0}^{\infty}\zeta_k < \infty$ almost surely.*

Theorem 4.2. *If Assumptions 2.1–2.4, 3.1-3.2 and 4.1–4.2 hold, the consensus times $\tau_k = 1$ and $S_k = 1$, then for any $\epsilon, \delta > 0$ and $\sigma_{i,k}$*

in (29), the sequence X_k generated by Algorithm 2 with stochastic approximation-type step size converges almost surely to X^*.

Proof. Similar to (22), considering the strategy update of Algorithm 2, we have

$$\begin{aligned}
&\|x_{i,k+1}-x_i^*\|^2\\
&\leqslant \|x_{i,k}-x_i^*\|^2+\alpha_k^2\|F_i(x_{i,k},Nv_{i,k})-\phi_i(x^*)\|^2\\
&\quad+\alpha_k^2\|e_{i,k}\|^2+\alpha_k^2\|n_{i,k}\|^2-2\alpha_k(x_{i,k}-x_i^*)^Te_{i,k}\\
&\quad-2\alpha_k(x_{i,k}-x_i^*)^T(F_i(x_{i,k},Nv_{i,k})-\phi_i(x^*))\\
&\quad+2\alpha_k^2(F_i(x_{i,k},Nv_{i,k})-\phi_i(x^*))^Te_{i,k}\\
&\quad+2\alpha_k^2(F_i(x_{i,k},Nv_{i,k})-\phi_i(x^*))^Tn_{i,k}\\
&\quad-2\alpha_k(x_{i,k}-x_i^*)^Tn_{i,k}+2\alpha_k e_{i,k}^Tn_{i,k}.
\end{aligned}\tag{33}$$

Then, summing the above inequality over $i=1,\dots,N$, and noting that $\mathbb{E}[n_k|\mathcal{F}_k]=\mathbb{E}[e_k|\mathcal{F}_k]=\mathbb{E}[e_k^Tn_k|\mathcal{F}_k]=0$, we obtain

$$\begin{aligned}
&\mathbb{E}[\|X_{k+1}-X^*\|^2|\mathcal{F}_k]\\
&\leqslant \|X_k-X^*\|^2+\alpha_k^2\mathbb{E}[\|e_k\|^2|\mathcal{F}_k]+\alpha_k^2\mathbb{E}[\|n_k\|^2|\mathcal{F}_k]\\
&\quad+2\alpha_k^2\|F(x_k,Nv_k)-\phi(x^*)\|^2\\
&\quad-2\alpha_k\mathbb{E}[(X_k-X^*)^T(F(x_k,Nv_k)-\phi(x^*))|\mathcal{F}_k].
\end{aligned}\tag{34}$$

Setting $y_k=\sum_{i=1}^N x_{i,k}$, we have

$$\begin{aligned}
&-2\alpha_k(X_k-X^*)^T(F(x_k,Nv_k)-\phi(x^*))\\
&=-2\alpha_k(X_k-X^*)^T(F(x_k,y_k)-\phi(x^*))\\
&\quad+2\alpha_k(X_k-X^*)^T(F(x_k,y_k)-F(x_k,Nv_k)).
\end{aligned}$$

Moreover, from Assumption 2.3 and by using the Cauchy–Schwarz inequality, it is obtained that

$$\begin{aligned}
&-2\alpha_k(X_k-X^*)^T(F(x_k,Nv_k)-\phi(x^*))\\
&\leqslant -2\alpha_k(X_k-X^*)^T(F(x_k,y_k)-\phi(x^*))\\
&\quad+2\alpha_kN\max_{i\in\mathcal{V}}\{L_i\}\|X_k-X^*\|\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|.
\end{aligned}\tag{35}$$

From (24), (34) and (35) it follows that

$$\begin{aligned}
&\mathbb{E}[\|X_{k+1}-X^*\|^2|\mathcal{F}_k]\\
&\leqslant (1+4\alpha_k^2L_\phi^2)\|X_k-X^*\|^2+\alpha_k^2\mathbb{E}[\|e_k\|^2|\mathcal{F}_k]\\
&\quad+\alpha_k^2\mathbb{E}[\|n_k\|^2|\mathcal{F}_k]+4\alpha_k^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|^2\\
&\quad+2\alpha_kN\max_{i\in\mathcal{V}}\{L_i\}\|X_k-X^*\|\mathbb{E}[\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\||\mathcal{F}_k]\\
&\quad-2\alpha_k(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big)\\
&\leqslant (1+4\alpha_k^2L_\phi^2)\|X_k-X^*\|^2\\
&\quad+\alpha_k^2\nu^2+\alpha_k^2\frac{8N^2n^2C^2\ln(1.25/\delta)}{\epsilon^2}\\
&\quad+4\alpha_k^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|^2\\
&\quad+2\alpha_kN\max_{i\in\mathcal{V}}\{L_i\}\|X_k-X^*\|\mathbb{E}[\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\||\mathcal{F}_k]\\
&\quad-2\alpha_k(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big).
\end{aligned}\tag{36}$$

From (30) it follows that

$$\begin{aligned}
\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|^2&\leqslant N(\theta M\rho^k+2\theta M\sum_{s=1}^k\rho^{k-s})^2\\
&\leqslant N(\theta M+\frac{2\theta M}{1-\rho})^2.
\end{aligned}$$

Hence, by $\sum_{k=0}^\infty\alpha_k^2<\infty$, we have $\sum_{k=0}^\infty 4\alpha_k^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|^2<\infty$. Furthermore, we have

$$\begin{cases}
\displaystyle\sum_{k=0}^\infty\alpha_k^2\nu^2<\infty,\\
\displaystyle\sum_{k=0}^\infty\alpha_k^2\frac{8N^2n^2C^2\ln(1.25/\delta)}{\epsilon^2}<\infty.
\end{cases}\tag{37}$$

Now, we apply Lemma 4.4 for

$$\begin{aligned}
Z_k:&=\|X_k-X^*\|^2,\\
u_k:&=4\alpha_k^2L_\phi^2,\\
\beta_k:&=\alpha_k^2\nu^2+\alpha_k^2\frac{8N^2n^2C^2\ln(1.25/\delta)}{\epsilon^2}\\
&\quad+4\alpha_k^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|^2\\
&\quad+2\alpha_kN\max_{i\in\mathcal{V}}\{L_i\}\|X_k-X^*\|\mathbb{E}[\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\||\mathcal{F}_k],\\
\zeta_k:&=2\alpha_k(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big).
\end{aligned}$$

From Lemma 4.3 and (37) it follows that $\sum_{k=0}^\infty\beta_k<\infty$. By using Lemma 4.4, it is obtained that $\|X_k-X^*\|^2$ converges almost surely, and $\sum_{k=0}^\infty\zeta_k=\sum_{k=0}^\infty 2\alpha_k(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big)<\infty$. Since the α_k-sequence is nonsummable and $F(\cdot)$ is strict monotonic, we have

$$\liminf_{k\to\infty}(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big)=0.$$

Notice that the sequence of X_k is bounded. Then, we consider its bounded subsequence X_{k_l}, which satisfies $\lim_{k\to\infty}(X_{k_l}-X^*)^T\big(F(x_{k_l},y_{k_l})-\phi(x^*)\big)=\liminf_{k\to\infty}(X_k-X^*)^T\big(F(x_k,y_k)-\phi(x^*)\big)=0$. Strict monotonicity of $F(\cdot)$ (Assumption 4.2) implies that this subsequence converges to X^*. Furthermore, the convergence of $\|X_k-X^*\|^2$ shows that $\lim_{k\to\infty}X_k=X^*$. The proof of the theorem is completed. □

Remark 4.1. The almost sure convergence and (ϵ,δ)-differential privacy of Algorithm 2 are established simultaneously. While we only study the privacy issue for stochastic aggregative games in Lei and Shanbhag (2020), Algorithm 2 can be extended to those presented in Franci and Grammatico (2020) and Yousefian et al. (2016).

From Theorem 4.2 it follows that the strategy update states X_k converge to Nash equilibrium X^*. The following theorem provides an estimate of the convergence rate.

Lemma 4.5 (*Lei & Shanbhag, 2020*). *If Assumptions 2.1–2.2 hold and consensus times $\tau_k=k+1$, $k\geqslant 1$, then there exists a constant $\theta>0$ and $\rho\in(0,1)$, such that*

$$\begin{aligned}
&\|V_k-\frac{1}{N}\mathbf{1}\mathbf{1}^TX_k\|\\
&\leqslant\sqrt{N}M\theta\rho^{\frac{(k+1)(k+2)}{2}}+2\sqrt{N}M\theta\rho^{\frac{(2k+1)}{2}}(1+\frac{2}{k\ln(1/\rho)})\\
&\quad+2\sqrt{N}M\theta e(\ln(\rho^{-1/2}))^{-\frac{1}{2}}\rho^{\frac{(k+1)^2}{2}}.
\end{aligned}$$

Theorem 4.3. *If Assumptions 2.1–2.5 and 3.1–3.2 hold, the step size satisfies $\alpha_k=\frac{1}{k^\gamma}$, $\frac{1}{2}<\gamma\leqslant 1$, consensus times $\tau_k=k+1$ and $S_k=1$, then for any $\epsilon,\delta>0$ and $\sigma_{i,k}$ in (29), the convergence rate of Algorithm 2 is given as follows. When $\frac{1}{2}<\gamma<1$, it holds that*

$$\mathbb{E}[\|X_{k+1}-X^*\|^2]=O(\frac{1}{k^\gamma}).$$

When $\gamma = 1$, it holds that

$$\mathbb{E}[\|X_{k+1} - X^*\|^2] = \begin{cases} O(\frac{1}{k^m}), & m < 1 \\ O(\frac{\ln k}{k}), & m = 1 \\ O(\frac{1}{k}), & m > 1 \end{cases}$$

where m is a positive constant in Assumption 2.5.

Proof. From (36) and by Assumption 2.5, we have

$$\begin{aligned} &\mathbb{E}[\|X_{k+1} - X^*\|^2] \\ &\leqslant (1 - 2\alpha_k m + 4\alpha_k^2 L_\phi^2)\mathbb{E}[\|X_k - X^*\|^2] + \alpha_k^2\mathbb{E}[\|e_k\|^2] \\ &\quad + \alpha_k^2\mathbb{E}[\|n_k\|^2] + 4\alpha_k^2 N^2 \max_{i\in\mathcal{V}}\{L_i^2\}\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|^2 \\ &\quad + 2\alpha_k N \max_{i\in\mathcal{V}}\{L_i\}\|X_k - X^*\|\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|]. \end{aligned}$$

By using Lemma 4.5, when $\alpha_k = \frac{1}{k^\gamma}$, $\frac{1}{2} < \gamma \leqslant 1$, there exists $k > k_0$ and $\beta > 0$ such that

$$\begin{aligned} &-2\alpha_k m + 4\alpha_k^2 L_\phi^2 \leqslant -\frac{m}{k^\gamma}, \\ &\alpha_k^2\mathbb{E}[\|e_k\|^2] + 4\alpha_k^2 N^2 \max_{i\in\mathcal{V}}\{L_i^2\}\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|^2 \\ &+ 2\alpha_k N \max_{i\in\mathcal{V}}\{L_i\}\|X_k - X^*\|\mathbb{E}[\|V_k - \frac{1}{N}\mathbf{1}\mathbf{1}^T X_k\|] \\ &+ \alpha_k^2\mathbb{E}[\|n_k\|^2] \leqslant \frac{\beta}{k^{2\gamma}}. \end{aligned} \tag{38}$$

From (38) it follows that

$$\begin{aligned} &\mathbb{E}\big[\|X_{k+1} - X^*\|^2\big] \\ &\leqslant \big[1 - \frac{m}{k^\gamma}\big]\mathbb{E}\big[\|X_k - X^*\|^2\big] + \frac{\beta}{k^{2\gamma}}, \quad \text{as} \quad k > k_0. \end{aligned}$$

Thus, iterating the above process, we have

$$\begin{aligned} \mathbb{E}\big[\|X_{k+1} - X^*\|^2\big] \leqslant &\prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big]\mathbb{E}\big[\|X_{k_0} - X^*\|^2\big] \\ &+ \sum_{l=k_0}^{k-1}\prod_{t=l+1}^{k}(1 - \frac{m}{t^\gamma})\frac{\beta}{l^{2\gamma}} + \frac{\beta}{k^{2\gamma}}. \end{aligned} \tag{39}$$

Note that $\prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big] = \exp\big(\sum_{t=k_0}^{k}\log(1 - \frac{m}{t^\gamma})\big) = O\big(\exp\big(-\sum_{t=k_0}^{k}\frac{m}{t^\gamma}\big)\big)$. Then, when $\gamma = 1$, it is obtained that

$$\begin{aligned} \prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big] &= O\big(\exp\big(-\sum_{t=k_0}^{k}\frac{m}{t}\big)\big) \\ &= O\big(\exp\big(-m\log\frac{k}{k_0}\big)\big) = O\big(\frac{1}{k^m}\big). \end{aligned} \tag{40}$$

From (39) and (40) it follows that

$$\begin{aligned} &\mathbb{E}\big[\|X_{k+1} - X^*\|^2\big] \\ &= O\big(\exp\big(-m\log\frac{k}{k_0}\big)\big) + O\big(\sum_{l=k_0}^{k-1}(\frac{l}{k})^m\frac{\beta}{l^2}\big) + O(\frac{\beta}{k^2}) \\ &= O\big(\exp\big(-m\log\frac{k}{k_0}\big)\big) + O\big(\frac{1}{k^m}\sum_{l=k_0}^{k-1}\frac{\beta}{l^{2-m}}\big) + O(\frac{\beta}{k^2}) \\ &= O\big(\frac{1}{k^m}\big) + O\big(\frac{1}{k^m}\sum_{l=k_0}^{k-1}\frac{\beta}{l^{2-m}}\big) + O(\frac{\beta}{k^2}). \end{aligned}$$

By using $\sum_{l=k_0}^{k-1}\frac{\beta}{l^{2-m}} \leqslant \int_{k_0-1}^{k}\frac{\beta}{x^{2-m}}dx$, we have

$$\sum_{l=k_0}^{k-1}\frac{\beta}{l^{2-m}} = \begin{cases} O(1), & m < 1 \\ O(\ln k), & m = 1 \\ O(\frac{1}{k^{1-m}}), & m > 1 \end{cases}$$

Thus, the results hold for $\gamma = 1$. When $\frac{1}{2} < \gamma < 1$, it is obtained that

$$\begin{aligned} \prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big] &= O\big(\exp\big(-\sum_{t=k_0}^{k}\frac{m}{t^\gamma}\big)\big) \\ &= O\big(\exp\big(-\frac{m}{1-\gamma}[(k+1)^{1-\gamma} - k_0^{1-\gamma}]\big)\big). \end{aligned}$$

Noting that for large enough k_0 and $l \geqslant k_0$, we have $(1 - \frac{m}{l^\gamma})^{-1} \leqslant 2$. Therefore, from (39) it follows that

$$\begin{aligned} &\mathbb{E}\big[\|X_{k+1} - X^*\|^2\big] \\ &\leqslant \prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big]\mathbb{E}\big[\|X_{k_0} - X^*\|^2\big] \\ &\quad + \sum_{l=k_0}^{k-1}\prod_{t=l+1}^{k}(1 - \frac{m}{t^\gamma})\frac{\beta}{l^{2\gamma}} + \frac{\beta}{k^{2\gamma}} \\ &\leqslant \prod_{t=k_0}^{k}\big[1 - \frac{m}{t^\gamma}\big]\mathbb{E}\big[\|X_{k_0} - X^*\|^2\big] \\ &\quad + 2\sum_{l=k_0}^{k-1}\prod_{t=l}^{k}(1 - \frac{m}{t^\gamma})\frac{\beta}{l^{2\gamma}} + \frac{\beta}{k^{2\gamma}} \\ &= O\big(\exp\big(-\frac{m}{1-\gamma}(k+1)^{1-\gamma}\big)\big) + O\big(\frac{1}{k^{2\gamma}}\big) \\ &\quad + O\big(\sum_{l=k_0}^{k-1}\exp\big(-\frac{m}{1-\gamma}(k+1)^{1-\gamma}\big) \\ &\quad \times \exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}}\big). \end{aligned} \tag{41}$$

Note that for large k_0, $\frac{\gamma}{mk_0^{1-\gamma}} < \frac{1}{2}$. Then, we have

$$\begin{aligned} &\sum_{l=k_0}^{k-1}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}} \\ &\leqslant \int_{k_0}^{k}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}}dl \\ &= \frac{1}{m}\int_{k_0}^{k}\frac{\beta}{l^\gamma}d(\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)) \\ &= \frac{1}{m}\frac{\beta}{l^\gamma}(\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big))|_{k_0}^{k} - \frac{\beta}{m}\int_{k_0}^{k}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)d(\frac{1}{l^\gamma}) \\ &= \frac{1}{m}\frac{\beta}{l^\gamma}(\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big))|_{k_0}^{k} \\ &\quad + \frac{\gamma}{m}\int_{k_0}^{k}\frac{1}{l^{1-\gamma}}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}}dl \\ &\leqslant \frac{1}{m}\frac{\beta}{k^\gamma}(\exp\big(\frac{m}{1-\gamma}k^{1-\gamma}\big)) \\ &\quad + \frac{\gamma}{mk_0^{1-\gamma}}\int_{k_0}^{k}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}}dl \\ &\leqslant \frac{1}{m}\frac{\beta}{k^\gamma}(\exp\big(\frac{m}{1-\gamma}k^{1-\gamma}\big)) + \frac{1}{2}\int_{k_0}^{k}\exp\big(\frac{m}{1-\gamma}l^{1-\gamma}\big)\frac{\beta}{l^{2\gamma}}dl. \end{aligned}$$

Furthermore, we have

$$\sum_{l=k_0}^{k-1}\exp(\frac{m}{1-\gamma}l^{1-\gamma})\frac{\beta}{l^{2\gamma}}=O(\frac{1}{k^{\gamma}}(\exp(\frac{m}{1-\gamma}k^{1-\gamma}))).$$

From (41) it follows that

$$\begin{aligned}&\mathbb{E}\big[\|X_{k+1}-X^*\|^2\big]\\&=O\big(\exp(-\frac{m}{1-\gamma}(k+1)^{1-\gamma})\big)+O\big(\frac{1}{k^{2\gamma}}\big)\\&\quad+O\big(\exp(-\frac{m}{1-\gamma}(k+1)^{1-\gamma})\frac{1}{k^{\gamma}}\exp\big(\frac{m}{1-\gamma}k^{1-\gamma}\big)\big)\\&=O\big(\frac{1}{k^{\gamma}}\big).\end{aligned}$$

The proof is completed. □

Remark 4.2. The convergence rate of the algorithm with stochastic approximation-type step sizes is given for the first time in Theorem 4.3, and the related results are not given even if privacy-preserving is not considered. From Theorem 4.3, we see that the privacy level ϵ does not affect the convergence rate of the algorithm when privacy noise is added directly to the gradient.

Next, by using the mini-batch method, observing the strategy update of Algorithm 2, we find that the parameter $\frac{1}{S_k}$ has reduced the influence of added privacy noise on the performance of the algorithm. Since the influence of added privacy noise on the algorithm's performance is reduced, the faster convergence rate will be achieved at the same privacy level, which will be shown in the following theorem.

Theorem 4.4. *If Assumptions 2.1–2.5 and 3.1–3.2 hold, consensus times $\tau_k=k+1$, $\alpha_k=\alpha$, $S_k=\lceil\alpha^{-2}q^{-(k+1)}\rceil$ for some $\rho\in(0,1)$, and $\mathbb{E}[\|X_0-X^*\|^2]\leqslant C_1$, then for any $\epsilon,\delta>0$ and $\sigma_{i,k}$ in (29), the following results hold:*

$$\begin{aligned}&\mathbb{E}[\|X_k-X^*\|^2]\\&\leqslant\begin{cases}(C_1+\frac{C_4}{\max\{\gamma/C_0,C_0/\gamma\}-1})\max\{C_0,\gamma\}^k, & C_0\neq\gamma\\(C_1+\frac{C_4}{\ln((\vartheta/C_0)^e)})\vartheta^k, & C_0=\gamma\end{cases}\end{aligned}$$

where

$$\begin{aligned}C_0&=1-2\alpha m+\alpha^2+4\alpha^2L_\phi^2+2(1+2\alpha^2)\alpha^2\bar{c}_1^2,\\C_2&=(1+2\alpha^2)(2\bar{c}_1^2\|X^*\|^2+N\bar{c}_2^2)\\&\quad+\frac{8N^2n^2\alpha^4C^2\ln(1.25/\delta)}{\epsilon^2},\\C_3&=2M\theta\rho^{-\frac{1}{2}}(1+\frac{2}{k\ln(1/\rho)})+2M\theta e(\rho\ln(\rho^{-1/2}))^{-\frac{1}{2}},\\C_4&=4\alpha^2N^3\max_{i\in\mathcal{V}}\{L_i^2\}(M^2\theta^2+C_3^2)\\&\quad+4\alpha N^{\frac{3}{2}}M\max_{i\in\mathcal{V}}\{L_i\}(M\theta+C_3)+C_2\alpha^2.\end{aligned}$$

Proof. Similar to the proof of (33)–(35), by considering the strategy update of Algorithm 2, we have

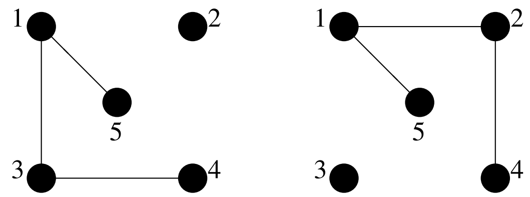


Fig. 1. Communication topology.

$$\begin{aligned}&\mathbb{E}[\|X_{k+1}-X^*\|^2|\mathcal{F}_k]\\&\leqslant(1-2\alpha m+\alpha^2+4\alpha^2L_\phi^2)\|X_k-X^*\|^2\\&\quad+4\alpha^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{11}^TX_k\|^2\\&\quad+4\alpha NM\max_{i\in\mathcal{V}}\{L_i\}\|V_k-\frac{1}{N}\mathbf{11}^TX_k\|\\&\quad+(1+2\alpha^2)\mathbb{E}[\|e_k\|^2|\mathcal{F}_k]+\frac{\alpha^2}{S_k^2}\mathbb{E}[\|n_k\|^2|\mathcal{F}_k]\\&\leqslant(1-2\alpha m+\alpha^2+4\alpha^2L_\phi^2)\|X_k-X^*\|^2\\&\quad+4\alpha^2N^2\max_{i\in\mathcal{V}}\{L_i^2\}\|V_k-\frac{1}{N}\mathbf{11}^TX_k\|^2\\&\quad+4\alpha NM\max_{i\in\mathcal{V}}\{L_i\}\|V_k-\frac{1}{N}\mathbf{11}^TX_k\|\\&\quad+(1+2\alpha^2)\mathbb{E}[\|e_k\|^2|\mathcal{F}_k]+\frac{8N^2n^2\alpha^2C^2\ln(1.25/\delta)}{S_k^2\epsilon^2}.\qquad(42)\end{aligned}$$

By Assumption 3.2, we have

$$\mathbb{E}[\|e_k\|^2|\mathcal{F}_k]\leqslant2\alpha^2\bar{c}_1^2\|X_k-X^*\|^2+\frac{2\bar{c}_1^2\|X^*\|^2+N\bar{c}_2^2}{S_k},$$

where $\bar{c}_1=\max_{1\leqslant i\leqslant n}c_{i,1}$, $\bar{c}_2=\max_{1\leqslant i\leqslant n}c_{i,2}$. Further, based on Lemma 4.5 and (42), $S_k\geqslant\alpha^{-2}$, we have

$$\begin{aligned}&\mathbb{E}[\|X_{k+1}-X^*\|^2]\leqslant C_0\mathbb{E}[\|X_k-X^*\|^2]\\&+4\alpha^2N^3\max_{i\in\mathcal{V}}\{L_i^2\}(M^2\theta^2\rho^{(k+1)(k+2)}+C_3^2\rho^{2(k+1)})\\&+4\alpha N^{\frac{3}{2}}M\max_{i\in\mathcal{V}}\{L_i\}(M\theta\rho^{\frac{(k+1)(k+2)}{2}}+C_3\rho^{k+1})+\frac{C_2}{S_k}\\&\leqslant C_0\mathbb{E}[\|X_k-X^*\|^2]+C_4\gamma^{k+1},\end{aligned}$$

where $\gamma=\max\{\rho,q\}$.

$$\mathbb{E}[\|X_{k+1}-X^*\|^2]\leqslant C_0^{k+1}\mathbb{E}[\|X_0-X^*\|^2]+C_4\sum_{m=1}^{k+1}C_0^{k+1-m}\gamma^m.$$

When $C_0\neq\gamma$, for $\gamma<C_0$, we have

$$\sum_{m=1}^{k+1}C_0^{k+1-m}\gamma^m=C_0^{k+1}\sum_{m=1}^{k+1}C_0^{-m}\gamma^m\leqslant\frac{1}{C_0/\gamma-1}C_0^{k+1}.$$

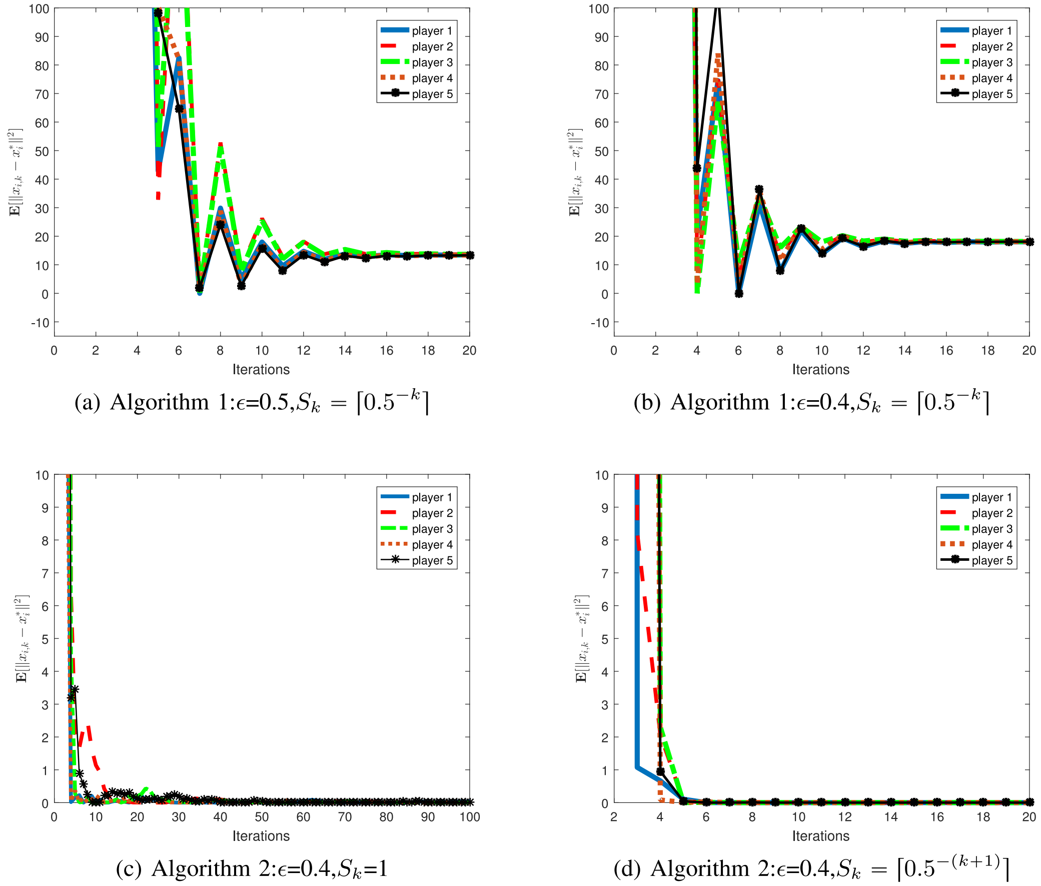
Similarly, when $\gamma>C_0$, it is obtained that $\sum_{m=1}^{k+1}C_0^{k+1-m}\gamma^m\leqslant\frac{1}{\gamma/C_0-1}\gamma^{k+1}$. Therefore, $\sum_{m=1}^{k+1}C_0^{k+1-m}\gamma^m\leqslant\frac{\max\{C_0,\gamma\}^{k+1}}{\max\{\gamma/C_0,C_0/\gamma\}-1}$. When $C_0=\gamma$, by using $kC_0^k\leqslant\vartheta^k/\ln((\vartheta/C_0)^e)$, $\vartheta\in(C_0,1)$, it is obtained that $\mathbb{E}[\|X_{k+1}-X^*\|^2]\leqslant C_0^{k+1}\mathbb{E}[\|X_0-X^*\|^2]+C_4(k+1)C_0^{k+1}\leqslant(C_1+\frac{C_4}{\ln((\vartheta/C_0)^e)})\vartheta^{k+1}$. □

Remark 4.3. In Theorems 4.4, we show that the convergence rate of the algorithm has the same order as that of the non-private algorithm in Lei and Shanbhag (2020). Different from Lei and Shanbhag (2020), the privacy level ϵ affects the convergence rate of the algorithm in the form of $O(\frac{1}{\epsilon^2})$.

Remark 4.4. We present two different differentially private distributed algorithms seeking the equilibrium solution in stochastic aggregative games. If a better convergent accuracy is required, then it is better to use the input-perturbation method; but if a simple operation or low computational complexity is preferred, then it is better to use the output-perturbation method.

5. Simulation example

This section provides a numerical simulation to testify the effectiveness of Algorithms 1–2. We consider a 5-player Cournot



(a) Algorithm 1:ϵ=0.5,$S_k=\lceil 0.5^{-k}\rceil$

(b) Algorithm 1:ϵ=0.4,$S_k=\lceil 0.5^{-k}\rceil$

(c) Algorithm 2:ϵ=0.4,S_k=1

(d) Algorithm 2:ϵ=0.4,$S_k=\lceil 0.5^{-(k+1)}\rceil$

Fig. 2. The expectations of the players' Nash equilibrium seeking mean square errors by using Algorithms 1–2.

competition model, where the player i's loss function is given as follows:

$$\mathbb{E}[f_i(x_i,\bar{x},\xi_i)] = \mathbb{E}[(x_i-\widehat{x}_i)^2+(0.04\sum_{i=1}^{5}x_i+\xi_i+5)x_i],$$

where $\widehat{x}_1=50$, $\widehat{x}_2=55$, $\widehat{x}_3=60$, $\widehat{x}_4=65$, $\widehat{x}_5=70$, $\xi_i\sim U(-\frac{c_i}{5},\frac{c_i}{5})$, $c_i\sim U(3,5)$. Since $\mathbb{E}[(x_i-\widehat{x}_i)^2+(0.04\sum_{i=1}^{5}x_i+\xi_i+5)x_i]=(x_i-\widehat{x}_i)^2+(0.04\sum_{i=1}^{5}x_i+5)x_i$, as shown in Ye et al. (2021), the game has a unique Nash equilibrium $X^*=(41.5,46.4,51.3,56.2,61.6)$. The communication topology among players switches between two graphs is given in Fig. 1. Setting $q=0.5$, from Fig. 2 it follows that $\mathbb{E}[\|X_k-X^*\|^2]$ converges to a small neighborhood of zero by using Algorithm 1. This indicates that the mean square convergence of the algorithm cannot be guaranteed when the privacy noise is added to the estimate of each player. In addition, comparing Fig. 2(a) and (b), it follows that the mean square error of the algorithm is inversely proportional to privacy level ϵ. From Fig. 2(c) and (d) it follows that $\mathbb{E}[\|X_k-X^*\|^2]$ exactly converges to zero by using Algorithm 2, while comparing Fig. 2(c) and (d), it follows that Algorithm 2 converges faster for larger batch sizes at the same level of privacy, which is consistent with theoretical analysis.

6. Conclusion

This paper designs privacy-preserving distributed algorithms seeking the Nash equilibrium in stochastic aggregative games and protecting each player's sensitive information. The (ϵ,δ)-differentially private method is used, where additional noise is introduced. The input and output-perturbation methods are both given. In particular, the algorithm converges almost surely to the equilibrium using stochastic approximation-type conditions and is (ϵ,δ)-differentially private. Under suitable conditions of consensus times, the algorithm's convergence rate is also given. Then, by using mini-batch methods, the effect of privacy noise on the algorithm's performance is reduced, and better performance than stochastic approximation-type distributed algorithms is obtained. Under appropriate assumptions, the algorithm can achieve exponential convergence and (ϵ,δ)-differential privacy. Finally, a simulation example is provided to verify the effectiveness of the algorithms.

Note that the attackers of this paper are passive. Ensuring that the algorithm performs well when active attackers exist is a challenging topic. Future work shall introduce a novel strategy that combines differential privacy methods and homomorphic encryption techniques. In addition, coupling constraints are of fundamental importance in technical applications, such as electricity markets or road networks. Hence, privacy-preserving distributed algorithms for stochastic aggregative games with coupling constraints are also an interesting topic, which deserves to be studied.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308–318).

Alshehri, K., Liu, J., Chen, X. D., & Basar, T. (2019). Privacy-preserving multi-period demand response: A game theoretic approach. arXiv:1710.00145v4.

Altafini, C. (2020). A system-theoretic framework for privacy preservation in continuous-time multiagent dynamics. *Automatica*, *122*, Article 109253.

Bassily, R., Feldman, V., & Talwar, K. (2019). Private stochastic convex optimization with optimal rates. In *Advances in neural information processing systems*. Vancouver, Canada.

Dong, R., Krichene, W., Bayen, A. M., & Sastry, S. S. (2015). Differential privacy of populations in routing games. In *IEEE 54th annual conference on decision and control* (pp. 2798–2803).

Dwork, C. (2006). Differential privacy. In *Proceedings of the 33rd international colloquium on automata, languages and programming* (pp. 1–12).

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Databases*.

Franci, B., & Grammatico, S. (2020). Stochastic generalized Nash equilibrium seeking in merely monotone games. *IEEE Transactions on Automatic Control*, http://dx.doi.org/10.1109/TAC.2021.3108496.

Gade, S., Winnicki, A., & Bose, S. (2020). On privatizing equilibrium computation in aggregate games over networks. *IFAC-PapersOnLine*, *53*(2), 3272–3277.

Grammatico, S., Parise, F., Colombino, M., & Lygeros, J. (2016). Decentralized convergence to Nash equilibria in constrained deterministic mean field control. *IEEE Transactions on Automatic Control*, *61*(11), 3315–3329.

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

Hao, Y., & Cheng, D. Z. (2021). Finite element approach to continuous potential games. *Science China. Information Sciences*, *64*, 149202:1–149202:3.

Hsu, J., Roth, A., & Ullman, J. (2013). Differential privacy for the analyst via private equilibrium computation. In *Symposium on theory of computing conference* (pp. 341–350).

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

Huang, Z., Mitra, S., & Dullerud, G. E. (2012). Differentially private iterative synchronous consensus. In *Proceedings of the 2012 ACM workshop on privacy in the electronic society* (pp. 81–90).

Koshal, J., Nedić, A., & Shanbhag, U. V. (2016). Distributed algorithms for aggregative games on graphs. *Operations Research*, *64*(3), 680–704.

Lee, J. (2017). Differentially private variance reduced stochastic gradient descent. In *International conference on new trends in computing sciences* (pp. 161-166).

Lei, J. L., & Shanbhag, U. V. (2020). Asynchronous schemes for stochastic and misspecified potential games and nonconvex optimization. *Operations Research*, *68*(6), 1742–1766.

Lei, J. L., & Shanbhag, U. V. (2020). Distributed variable sample-size gradient-response and best-response schemes for stochastic Nash equilibrium problems over graphs. arXiv:1811.11246v2.

Li, N., & Marden, J. (2013). Designing games for distributed optimization. *IEEE Journal of Selected Topics in Signal Processing*, *7*, 230–242.

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

Liang, W. J., Chen, H., Zhang, J., Zhao, D., & Li, C. P. (2020). An effective scheme for top-k frequent itemset mining under differential privacy conditions. *Science China. Information Sciences*, *63*(5), Article 159101:1-159101:3.

Liu, S. J., & Krstic, M. (2011). Stochastic Nash equilibrium seeking for games with general nonlinear payoffs. *SIAM Journal on Control and Optimization*, *49*(4), 1659–1679.

Liu, X. K., Zhang, J. F., & Wang, J. M. (2020). Differentially private consensus algorithm for continuous-time heterogeneous multi-agent systems. *Automatica*, *12*, Article 109283.

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

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

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

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

Paccagnan, D., Gentile, B., Parise, F., Kamgarpour, M., & Lygeros, J. (2019). Nash and Wardrop equilibria in aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, *64*(4), 1373–1388.

Pang, Y., & Hu, G. Q. (2021). Distributed Nash equilibrium seeking with limited cost function knowledge via a consensus-based gradient-free method. *IEEE Transactions on Automatic Control*, *66*(4), 1832–1839.

Parise, F., Gentile, B., Grammatico, S., & Lygeros, J. (2015). Network aggregative games: Distributed convergence to Nash equilibria. In *IEEE 54th annual conference on decision and control* (pp. 2295–2300).

Polyak, B. (1987). *Introduction to optimization*. New York, NY, USA: Optimization Software, Inc..

Salehisadaghiani, F., & Pavel, L. (2016). Distributed Nash equilibrium seeking: A gossip-based algorithm. *Automatica*, *72*, 209–216.

Salehisadaghiani, F., & Pavel, L. (2018). Distributed Nash equilibrium seeking in networked graphical games. *Automatica*, *87*, 17–24.

Shakarami, M., Persis, C. D., & Monshizadeh, N. (2019). Privacy and robustness guarantees in distributed dynamics for aggregative games. arXiv:1910.13928v2.

Shokri, M., & Kebriaei, H. (2020). Leader-follower network aggregative game with stochastic agents' communication and activeness. *IEEE Transactions on Automatic Control*, *65*(12), 5496–5502.

Song, S., Chaudhuri, K., & Sarwate, A. D. (2013). Stochastic gradient descent with differentially private updates. In *IEEE global conference on signal and information processing* (pp. 245–248).

Ye, M. J., & Hu, G. Q. (2017). Distributed Nash equilibrium seeking by a consensus based approach. *IEEE Transactions on Automatic Control*, *62*(9), 4811–4818.

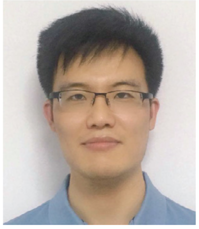
Ye, M. J., Hu, G. Q., Xie, L. H., & Xu, S. Y. (2021). Differentially private distributed Nash equilibrium seeking for aggregative games. *IEEE Transactions on Automatic Control*, http://dx.doi.org/10.1109/TAC.2021.3075183.

Yousefian, F., Nedić, A., & Shanbhag, U. V. (2016). Self-tuned stochastic approximation schemes for non-Lipschitzian stochastic multi-user optimization and Nash games. *IEEE Transactions on Automatic Control*, *61*(7), 1753–1766.

Zhang, C. L., Ahmad, M., & Wang, Y. Q. (2019). ADMM based privacy-preserving decentralized optimization. *IEEE Transactions on Information Forensics and Security*, *14*(3), 565–580.

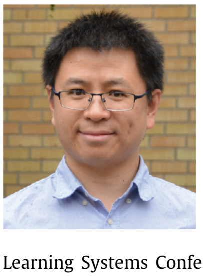
Zhang, J. F., Tan, J. W., & Wang, J. M. (2021). Privacy security in control systems. *Science China. Information Sciences*, *64*, 176201:1–176201:3.

Zhou, Y., & Tang, S. (2020). Differentially private distributed learning. *INFORMS Journal on Computing*, *32*(3), 779–789.



Jimin Wang 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 was a recipient of Shandong University's excellent doctoral dissertation.

Ji-Feng Zhang 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. He is now with the ISS, Academy of Mathematics and Systems Science, CAS. His current research interests include system modeling, adaptive control, stochastic systems, and multi-agent systems. He is an IEEE Fellow, IFAC Fellow, CAA Fellow, CSIAM 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 is a Vice-President of the Chinese Mathematical Society and the Chinese Association of Automation. 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; Vice-President of the Systems Engineering Society of China. He has served as Editor-in-Chief, Deputy Editor-in-Chief, Senior Editor or Associate Editor for more than 10 journals, including *Science China Information Sciences*, *National Science Review*, *IEEE Transactions on Automatic Control*, and *SIAM Journal on Control and Optimization* etc.



Xingkang He is a postdoctoral research associate in the Department of Electrical Engineering, University of Notre Dame, USA. From Oct. 2018 to Oct. 2021, he was a postdoctoral researcher in the Division of Decision and Control Systems, KTH Royal Institute of Technology, Sweden. He received his Ph.D. degree at the University of Chinese Academy of Sciences in 2018. His research interests include security and privacy of cyber–physical systems, networked control systems, robotics, and machine learning. He received the Best Paper Award in 2018 IEEE Data Driven Control and Learning Systems Conference.