Differentially-private Distributed Algorithms for Aggregative Games with Guaranteed Convergence

Yongqiang Wang, Angelia Nedić

Abstract—**The distributed computation of a Nash equilibrium in aggregative games is gaining increased traction in recent years. Of particular interest is the coordinator-free scenario where individual players only access or observe the decisions of their neighbors due to practical constraints. Given the non-cooperative relationship among participating players, protecting the privacy of individual players becomes imperative when sensitive information is involved. We propose a fully distributed equilibrium-seeking approach for aggregative games that can achieve both rigorous differential privacy and guaranteed computation accuracy of the Nash equilibrium. This is in sharp contrast to existing differential-privacy solutions for aggregative games that have to either sacrifice the accuracy of equilibrium computation to gain rigorous privacy guarantees, or allow the cumulative privacy budget to grow unbounded, hence losing privacy guarantees, as iteration proceeds. Our approach uses independent noises across players, thus making it effective even when adversaries have access to all shared messages as well as the underlying algorithm structure. The encryption-free nature of the proposed approach, also ensures efficiency in computation and communication. The approach is also applicable in stochastic aggregative games, able to ensure both rigorous differential privacy and guaranteed computation accuracy of the Nash equilibrium when individual players only have stochastic estimates of their pseudo-gradient mappings. Numerical comparisons with existing counterparts confirm the effectiveness of the proposed approach.**

Index Terms –**Aggregative games, distributed Nash equilibrium seeking, differential privacy**

I. Introduction

The distributed seeking of a Nash equilibrium over networks has gained increased attention in recent years. It has found applications in various domains where multiple players (agents) compete to maximize their individual payoff functions, with typical examples including energy management in smart grids [1], congestion control in communication networks [2], market analysis in economics [3], and route coordination in road networks [4]. In many of these application scenarios, a participating player's payoff function depends on the aggregate (e.g., total sum) of all players' decisions, but such an aggregate is inaccessible to individual players. Namely, no central coordinator exits to collect and distribute the aggregate information, and a player can only access the decisions of its immediate neighbors. Consequently, individual players cannot compute their accurate payoff functions, but instead, they share information among neighboring players to estimate the aggregate decision [5], [6], [7], [8].

Despite recent progress in such aggregative games where distributed computation can be conducted under partial-decision information obtained through local information sharing [5], [6], [7], [8], [9], all these distributed algorithms explicitly share estimate/decision variables in every iteration, which can lead to a disclosure of players' sensitive information. This problem is significant in that the players are non-cooperative and have every reason to protect their individual private information in competition. Take the Nash-Cournot game as an example, players' cost functions could be market sensitive and every player is well motivated to protect its cost function to gain an edge over its competitors [10]. Moreover, sometimes privacy legislations require the privacy of players' information to be protected during equilibrium-behavior implementation in a game. For example, in routing games [11], California Privacy Rights Act forbids disclosing the spatiotemporal information of drivers because these information can be used as the basis for inferences of a person's activities [12].

To address the urgent need of privacy protection in aggregative games, some efforts have been reported in recent years (see, e.g., [13], [14], [15]). However, these efforts mainly address Nash-equilibrium seeking in the presence of a coordinator which greatly simplifies the privacy design problem. The work [16] proposes a privacy approach for fully distributed Nash-equilibrium seeking, but the use of correlated noise restricts its applicability when players can have arbitrary communication patterns. The paper [17] proposes to use an uncertain parameter to obscure the pseudo-gradient mapping in continuous-time Nash-equilibrium seeking algorithms. However, the fact that the uncertain parameter is a constant scalar restricts its privacy-protection strength. In fact, the approach can only avoid the payoff function from being uniquely identifiable, while the relations among private parameters are still revealed. Given that differential privacy can provide strong protection against arbitrary post-processing and auxiliary information [18], and is becoming the de facto standard for privacy protection, the recent works [19] and [20] propose differential-privacy mechanisms for equilibrium seeking in fully distributed aggregative games. However, to ensure rigorous ϵ-differential privacy (with finite cumulative privacy budget), these approaches have to sacrifice provable convergence to the exact Nash equilibrium.

To avoid the problem of trading convergence accuracy for differential privacy that is plaguing existing differential-privacy approaches for aggregative games, this paper presents the first distributed Nash-equilibrium seeking approach that can simultaneously achieve both rigorous ϵ-differential privacy (with finite cumulative privacy budget) and guaranteed conver-

The work was supported in part by the National Science Foundation under Grants ECCS-1912702, CCF-2106293, CCF-2106336, CCF-2215088, and CNS-2219487.

Yongqiang Wang is with the Department of Electrical and Computer Engineering, Clemson University, Clemson, SC 29634, USA yongqiw@clemson.edu

Angelia Nedić is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85281, USA angelia.nedich@asu.edu

gence to the Nash equilibrium. Motivated by the observation that persistent differential-privacy noise has to be repeatedly injected in every iteration of information sharing, which results in significant reduction in algorithmic accuracy, our key idea is to gradually weaken the coupling strength to attenuate the effect of differential-privacy noise added on shared messages. We judiciously design the weakening factor sequence to ensure that convergence to the Nash equilibrium is guaranteed even in the presence of persistent differential-privacy noise. It is worth noting that compared with our recent result for differentially-private distributed optimization [21], [22], the results here are significantly different: 1) In distributed optimization, agents cooperate to minimize a common objective function, whereas in aggregative games players are non-cooperative and only mind their own payoff functions; 2) Adding differential-privacy noise can easily alter the equilibrium of a game (just as evidenced by the loss of accurate convergence in existing differential-privacy approaches for aggregative games [19]), and hence we have to judiciously design our noise-adding mechanism to avoid perturbing the equilibrium.

Contributions: The main contributions are summarized as follows:

1) By judiciously designing the aggregate estimation mechanism, we propose a fully distributed equilibrium-seeking approach for aggregative games that can ensure rigorous ϵ-differential privacy without losing guaranteed convergence to the Nash equilibrium. The algorithm can ensure both a finite cumulative privacy budget and accurate convergence, which is in sharp contrast to existing differential-privacy approaches for aggregative games (see, e.g., [19] and [20]) that have to trade accurate convergence for differential privacy. To the best of our knowledge, this is the first such algorithm in the literature.

2) We propose a new proof technique for the convergence analysis of the fully distributed equilibrium-seeking approach for aggregative games in the presence of information-sharing noise (caused by, e.g., differential-privacy design). The new convergence derivation does not impose the restriction that the pseudo-gradient mapping is uniformly bounded, an assumption that is used in existing distributed algorithms (e.g., [19] and [20]) for aggregative games subject to noises. Note that avoiding the uniformly bounded pseudo-gradient assumption is significant since in the presence of differential-privacy noise (e.g., Laplacian or Gaussian noise) which are not uniformly bounded, the aggregative estimation may become unbounded, which makes the pseudo-gradient mapping unbounded in many common games such as the Nash-Cournot game under a price governed by the linear inverse-demand function.

3) Even without taking privacy into consideration, the proposed algorithms and theoretical derivations are of interest themselves. The convergence analysis for the proposed algorithms has fundamental differences from existing proof techniques. More specifically, existing convergence analysis of distributed (generalized) Nash-equilibrium seeking algorithms for aggregative games (e.g., [5], [23], [24], [25], [26]) and their stochastic variants (e.g., [20] and [27]) rely on the geometric (exponential) decreasing of the aggregate-estimation error (consensus error) among the players, which is possible only when all nonzero coupling weights are lower bounded by a positive constant. Such geometric (exponential) decreasing of aggregate-estimation error is key to proving exact convergence of all players' iterates to the Nash equilibrium. In our case, since the coupling strength decays to zero, such geometric (exponential) decreasing of players' aggregate-estimation error does not exist any more, which makes it impossible to use the proof techniques in existing results.

4) We extend the approach to the case where the pseudo-gradient mapping is stochastic, and prove that rigorous ϵ-differential privacy and guaranteed convergence can still be achieved simultaneously in this case. Note that different from [20], [27] which consider stochastic pseudo-gradients with decreasing variances (via increasing sample sizes), we allow the variance of the stochastic pseudo-gradient to be constant, or even increasing with time, as specified in Remark 12.

The organization of the paper is as follows. Sec. II gives the problem formulation and some results for a later use. Sec. III presents a differentially-private distributed equilibrium-seeking algorithm for aggregative games. This section also proves that the algorithm can ensure all players' convergence to the exact Nash equilibrium while ensuring rigorous ϵ-differential privacy with a finite cumulative privacy budget, even when the number of iterations goes to infinity. Sec. IV extends the approach to the case of stochastic aggregative games and prove that it can ensure both guaranteed computation accuracy of the Nash equilibrium and differential privacy with guaranteed finite cumulative privacy budget when individual players only have stochastic estimates of their pseudo-gradient mappings. Sec. V presents numerical comparisons with existing distributed computation approaches for aggregative games to confirm the obtained results. Finally, Sec. VI concludes the paper.

Notations: We use $\mathbb{R}^d$ to denote the Euclidean space of dimension d. We write I_d for the identity matrix of dimension d, and $\mathbf{1}_d$ for the d-dimensional column vector will all entries equal to 1; in both cases we suppress the dimension when it is clear from the context. For a vector x, $[x]_i$ denotes its ith element. We write $x > 0$ (resp. $x \geq 0$) if all elements of x are positive (resp. non-negative). We use $\langle \cdot, \cdot \rangle$ to denote the inner product and $\|x\|$ for the standard Euclidean norm of a vector x. We use $\|x\|_1$ to represent the L_1 norm of a vector x. We write $\|A\|$ for the matrix norm induced by the vector norm $\|\cdot\|$. We let A^T denote the transpose of a matrix A. For two vectors u and v with the same dimension, we use $u \leq v$ to represent the relationship that every entry of the vector u is no greater than the corresponding entry of v. Often, we abbreviate *almost surely* by *a.s.*

II. Problem Formulation and Preliminaries

A. On Aggregative Games

We consider a set of m players (or agents), i.e., $[m] = \{1, 2, \ldots, m\}$, which are indexed by $1, 2, \cdots, m$. Player i is characterized by a strategy set $K_i \subseteq \mathbb{R}^d$ and a payoff function $f_i(x_i, \bar{x})$ where x_i denotes the decision of player i and $\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ denotes the average of all players' decisions. Note that sometimes the payoff function depends on the aggregate decision $m\bar{x} = \sum_{i=1}^{m} x_i$ rather than the average decision $\bar{x}$. In these cases, all algorithms and analysis in this paper are still valid by replacing $\bar{x}$ with $m\bar{x}$.

Since each decision variable x_i is restricted in K_i, the average $\bar{x}$ is restricted to the set that is the $1/m$-scaling of the Minkowski sum[1] of the sets K_i, denoted by $\bar{K}$, i.e., $\bar{K} = \frac{1}{m}(K_1 + K_2 + \cdots + K_m)$. With this notation, we can formalize $f_i(x_i, \bar{x})$ as a mapping from $K_i \times \bar{K}$ to $\mathbb{R}$, and further formulate the game that player i faces as the following parameterized optimization problem:

$$\min f_i(x_i, \bar{x}) \quad \text{s.t.} \quad x_i \in K_i \text{ and } \bar{x} \in \bar{K}. \tag{1}$$

The constraint set K_i and the function $f_i(\cdot)$ are assumed to be known to player i only.

To characterize a Nash equilibrium of the aggregative game (1), following [5], we introduce the following notations

$$F_i(x_i, \bar{x}) \triangleq \nabla_{x_i} f_i(x_i, \bar{x}), \quad K \triangleq \Pi_{i=1}^{m} K_i, \tag{2}$$

and $x \triangleq \left[x_1^T, \cdots, x_m^T\right]^T$.

These notions allow us to define two mappings

$$F(x,u) \triangleq \begin{pmatrix} F_1(x_1,u) \\ \vdots \\ F_m(x_m,u) \end{pmatrix}, \tag{3}$$

$$\phi(x) = F(x, \bar{x}), \quad \forall x \in K. \tag{4}$$

Similar to [5], we make the following assumptions on the constraint sets K_i and the functions f_i:

Assumption 1. *Each $K_i \in \mathbb{R}^d$ is compact and convex. Each function $f_i(x_i, y)$ is continuously differentiable in (x_i, y) over some open set containing the set $K_i \times \bar{K}$, while each function $f_i(x_i, \bar{x})$ is convex in x_i over the set K_i. The mapping $\phi(x)$ is strictly monotone over K, i.e., for all $x \neq x'$ in K, we always have*

$$\left(\phi(x) - \phi(x')\right)^T (x - x') > 0.$$

Remark 1. *It is worth noting that the strictly monotone assumption on $\phi(x)$ is weaker than the commonly used strongly monotone assumption in [7], [19], [20], [27], [28], [29].*

According to [5], Assumption 1 ensures that the aggregative game (1) has a unique Nash equilibrium $x^* = [(x_1^*)^T, (x_2^*)^T, \ldots, (x_m^*)^T]^T \in \mathbb{R}^{md}$. Moreover, following [5], we also make the following assumption:

Assumption 2. *Each mapping $F_i(x_i, u)$ satisfies the following Lipschitz continuous condition with respect to u: for all $x_i \in K_i$, all $u_1, u_2 \in \bar{K}$, and all $i \in [m]$, we always have $\|F_i(x_i, u_1) - F_i(x_i, u_2)\| \leq \tilde{L}\|u_1 - u_2\|$ for some $\tilde{L} > 0$.*

We consider distributed algorithms for equilibrium seeking of the game in (1). Namely, no player has a direct access to the average decision $\bar{x}$ or the aggregate decision $m\bar{x}$. Instead, each player has to construct a local estimate of the average/aggregate through local interactions with its neighbors. We describe the local interaction using a weight matrix $L = \{L_{ij}\}$, where $L_{ij} > 0$ if players j and i can directly communicate with each other, and $L_{ij} = 0$ otherwise. For a player $i \in [m]$, its neighbor set $\mathbb{N}_i$ is the collection of players j such that $L_{ij} > 0$ holds. We define $L_{ii} \triangleq -\sum_{j \in \mathbb{N}_i} L_{ij}$ for all $i \in [m]$, where $\mathbb{N}_i$ is the neighbor set of agent i. Furthermore, we make the following assumption on L:

Assumption 3. *The matrix $L = \{w_{ij}\} \in \mathbb{R}^{m \times m}$ is symmetric and satisfies $\mathbf{1}^T L = \mathbf{0}^T$, $L\mathbf{1} = \mathbf{0}$, and $\|I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| < 1$.*

Remark 2. *$I + L$ here corresponds to the commonly-used weight matrix W in the literature (see, e.g., [30], [31], where in Sec. 2.4 of [30] several examples of local interaction patterns satisfying Assumption 3 are given). We decompose the commonly-used weight matrix W into I and L because this will facilitate convergence analysis under our differential-privacy oriented design, which gradually attenuates inter-player interaction (represented by L) while keeping intact self-interaction (i.e., the influence of a player's state at k to its own state at $k+1$, which corresponds to the "I" matrix in $I+W$). Please see Algorithms 1 and 2 for details.*

One can verify $\|I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| = \max\{|1 + \rho_2|, |1 + \rho_m|\}$, where $\{\rho_i, i \in [m]\}$ are the eigenvalues of L, with $\rho_m \leq \ldots \leq \rho_2 \leq \rho_1 = 0$. Hence, the inequality in Assumption 3 ensures $|1 + \rho_2| < 1$ and further $\rho_2 \neq 0$. Given that $-L$ corresponds to the conventional weighted Laplacian matrix, the inequality $\|I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| < 1$ in Assumption 3 ensures that the second smallest eigenvalue of $-L$ is greater than zero, i.e., the interaction graph induced by L is connected (there is a path from each player to every other player) [32]. Moreover, this inequality also ensures $\rho_m > -2$, which can be enforced by requiring $\sum_{j \in \mathbb{N}_i} L_{ij} < 1$ for all $i \in [m]$. It is worth noting that in algorithms with a constant stepsize (see, e.g., [7]), the condition of $\rho_m > -2$ (and hence $\sum_{j \in \mathbb{N}_i} L_{ij} < 1$ for all $i \in [m]$) can be relaxed.

In the analysis of our methods, we use the following results:

Lemma 1. *[21] Let $\{v^k\}$,$\{\alpha^k\}$, and $\{p^k\}$ be random non-negative scalar sequences, and $\{q^k\}$ be a deterministic non-negative scalar sequence satisfying $\sum_{k=0}^{\infty} \alpha^k < \infty$ almost surely, $\sum_{k=0}^{\infty} q^k = \infty$, $\sum_{k=0}^{\infty} p^k < \infty$ almost surely, and the following inequality:*

$$\mathbb{E}\left[v^{k+1} | \mathcal{F}^k\right] \leq (1 + \alpha^k - q^k) v^k + p^k, \quad \forall k \geq 0 \quad a.s.$$

where $\mathcal{F}^k = \{v^\ell, \alpha^\ell, p^\ell; 0 \leq \ell \leq k\}$. Then, $\sum_{k=0}^{\infty} q^k v^k < \infty$ and $\lim_{k\to\infty} v^k = 0$ hold almost surely.

Lemma 2. *[21] Let $\{\mathbf{v}^k\} \subset \mathbb{R}^d$ and $\{\mathbf{u}^k\} \subset \mathbb{R}^p$ be random nonnegative vector sequences, and $\{a^k\}$ and $\{b^k\}$ be random nonnegative scalar sequences such that*

$$\mathbb{E}\left[\mathbf{v}^{k+1} | \mathcal{F}^k\right] \leq (V^k + a^k \mathbf{1}\mathbf{1}^T)\mathbf{v}^k + b^k \mathbf{1} - H^k \mathbf{u}^k, \quad \forall k \geq 0$$

holds almost surely, where $\{V^k\}$ and $\{H^k\}$ are random sequences of nonnegative matrices and $\mathbb{E}\left[\mathbf{v}^{k+1} | \mathcal{F}^k\right]$ denotes the conditional expectation given $\mathbf{v}^\ell, \mathbf{u}^\ell, a^\ell, b^\ell, V^\ell, H^\ell$ for $\ell = 0, 1, \ldots, k$. Assume that $\{a^k\}$ and $\{b^k\}$ satisfy $\sum_{k=0}^{\infty} a^k < \infty$ and $\sum_{k=0}^{\infty} b^k < \infty$ almost surely, and that there exists a (deterministic) vector $\pi > 0$ such that $\pi^T V^k \leq \pi^T$ and $\pi^T H^k \geq 0$ hold almost surely for all $k \geq 0$. Then, we have

[1]A scaling tX of a set X with a scalar t is the set given by $tX = \{tx \mid x \in X\}$. A Minkowski sum of two sets X and Y is the set $X + Y = \{x + y \mid x \in X, y \in Y\}$.

1) $\{\pi^T \mathbf{v}^k\}$ converges to some random variable $\pi^T \mathbf{v} \geq 0$ almost surely; 2) $\{\mathbf{v}^k\}$ is bounded almost surely; and 3) $\sum_{k=0}^{\infty} \pi^T H^k \mathbf{u}^k < \infty$ holds almost surely.

B. On Differential Privacy

We adopt the notion of ϵ-differential privacy for continuous bit streams [33], which has recently been applied to distributed optimization algorithms (see [34] as well as our work [21]). A commonly used approach to enabling differential privacy is injecting Laplace noise to shared messages. For a constant $\nu > 0$, we use $\mathrm{Lap}(\nu)$ to denote a Laplace distribution of a scalar random variable with the probability density function $x \mapsto \frac{1}{2\nu} e^{-\frac{|x|}{\nu}}$. It can be verified that $\mathrm{Lap}(\nu)$ has zero mean and variance $2\nu^2$. Following the formulation of distributed optimization in [34], for the convenience of differential-privacy analysis, we represent the distributed game $\mathcal{P}$ in (1) by three parameters $(K, \mathbb{F}, L)$, where K defined in (2) is the domain of decision variables, $\mathbb{F} \triangleq \{f_1, \cdots, f_m\}$, and L is the inter-player interaction weight matrix L. Then we define adjacency between two games as follows:

Definition 1. *Two distributed Nash-equilibrium seeking problems $\mathcal{P} = (K, \mathbb{F}, L)$ and $\mathcal{P}' = (K', \mathbb{F}', L')$ are adjacent if the following conditions hold:*

- *$K = K'$ and $L = L'$, i.e., the domains of decision variables and the interaction weight matrices are identical;*
- *there exists an $i \in [m]$ such that $f_i \neq f_i'$ but $f_j = f_j'$ for all $j \in [m]$, $j \neq i$;*
- *the different payoff functions f_i and f_i' have similar behaviors around x^*, the Nash equilibrium of $\mathcal{P}$. More specifically, there exits some $\delta > 0$ such that for all ι and ι' in $B_\delta(x_i^*) \triangleq \{u : u \in \mathbb{R}^d, \|u - x_i^*\| < \delta\}$, we have $\Pi_{K_i}[\iota - \lambda \nabla_\iota f_i(\iota, \cdot)] - \iota = \Pi_{K_i'}[\iota' - \lambda \nabla_{\iota'} f_i'(\iota', \cdot)] - \iota'$ for all $\lambda > 0$, where $\Pi_{K_i}[\cdot]$ denotes the Euclidean projection of a vector onto the set K_i.*

In Definition 1, since the change of a payoff function from f_i to f_i' in the second condition can be arbitrary, additional restrictions have to be imposed to ensure rigorous DP in distributed Nash equilibrium seeking. Different from [19] which restricts all pseudo-gradients to be uniformly bounded, we use the third condition, which, as shown later, allows us to ensure rigorous DP while maintaining provable convergence to the exact Nash equilibrium. It is worth noting that in the constraint-free case, the condition reduces to requiring $\nabla_\iota f_i(\iota, \cdot) = \nabla_{\iota'} f_i'(\iota', \cdot)$ for ι and ι' in the neighborhood of the Nash equilibrium of $\mathcal{P}$.

Given a distributed Nash-equilibrium seeking problem $\mathcal{P}$, we represent an iterative distributed Nash-equilibrium seeking algorithm as a mapping $\mathcal{R}_{\mathcal{P}}(\vartheta_0) : \vartheta^0 \mapsto \mathcal{O}$, where ϑ^0 is the initial state and $\mathcal{O}$ is the observation sequence (the sequence of shared messages). Under a fixed distributed Nash-equilibrium seeking algorithm, for a given distributed Nash-equilibrium seeking problem $\mathcal{P}$, observation sequence $\mathcal{O}$, and initial state ϑ^0, we denote the corresponding internal state at iteration k as $\mathcal{A}_{\mathcal{P},\mathcal{O},\vartheta^0}[k]$.

Definition 2. *(ϵ-differential privacy, adapted from [34]). For a given $\epsilon > 0$, an iterative distributed algorithm solving problem (1) is ϵ-differentially private if for any two adjacent $\mathcal{P}$ and $\mathcal{P}'$, any set of observation sequences $\mathcal{O}_s \subseteq \mathbb{O}$ (with $\mathbb{O}$ denoting the set of all possible observation sequences), and any initial state ϑ^0, we always have*

$$\mathbb{P}[\mathcal{R}_{\mathcal{P}}(\vartheta^0) \in \mathcal{O}_s] \leq e^{\epsilon} [\mathcal{R}_{\mathcal{P}'}(\vartheta^0) \in \mathcal{O}_s], \tag{5}$$

where the probability $\mathbb{P}$ is taken over the randomness over iteration processes.

The above definition of ϵ-differential privacy ensures that an adversary having access to all shared messages in the network cannot gain information with a significant probability of any participating player's payoff function. It can also be seen that a smaller ϵ means a higher level of privacy protection. It is also worth noting that the considered notion of ϵ-differential privacy is more stringent than other relaxed (approximate) differential privacy notions such as (ϵ, δ)-differential privacy [35], zero-concentrated differential privacy [36], or Rényi differential privacy [37].

III. A Differentially-Private Distributed Equilibrium-Seeking Algorithm for Aggregative Games

To achieve strong differential privacy, independent noise should be injected repeatedly in every round of message sharing and, hence, constantly affecting the algorithm through inter-player interactions and leading to significant reduction in algorithmic accuracy. Motivated by this observation, we propose to gradually weaken inter-player interactions to reduce the influence of differential-privacy noise on computation accuracy. Interestingly, we prove that by judiciously designing the interaction weakening mechanism, we can ensure convergence of all players to the Nash equilibrium even in the presence of persistent differential-privacy noise.

Algorithm 1: Differentially-private distributed algorithm for aggregative games with guaranteed convergence

---

Parameters: Stepsize $\lambda^k > 0$ and weakening factor $\gamma^k > 0$.

Every player i maintains one decision variable x_i^k, which is initialized with a random vector in $K_i \subseteq \mathbb{R}^d$, and an estimate of the aggregate decision v_i^k, which is initialized as $v_i^0 = x_i^0$.

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

a) Every player j adds persistent differential-privacy noise ζ_j^k to its estimate v_j^k, and then sends the obscured estimate $v_j^k + \zeta_j^k$ to player $i \in \mathbb{N}_j$.

b) After receiving $v_j^k + \zeta_j^k$ from all $j \in \mathbb{N}_i$, player i updates its decision variable and estimate as follows:

$$\begin{aligned} x_i^{k+1} =& \Pi_{K_i}\left[x_i^k - \lambda^k F_i(x_i^k, v_i^k)\right], \\ v_i^{k+1} =& v_i^k + \gamma^k \sum_{j \in \mathbb{N}_i} L_{ij}(v_j^k + \zeta_j^k - v_i^k - \zeta_i^k) \\ &+ x_i^{k+1} - x_i^k, \end{aligned} \tag{6}$$

where $\Pi_{K_i}[\cdot]$ denotes the Euclidean projection of a vector onto the set K_i.

c) **end**

Remark 3. *In the iterates in (6), we judiciously let player i use $v_i^k+\zeta_i^k$ that it shares with its neighbors in its interaction terms ($L_{ij}(v_j^k+\zeta_j^k-v_i^k-\zeta_i^k)$ for $j\in\mathbb{N}_i$) to cancel out the influence of noises on the aggregate estimation (average estimation, more precisely). As shown latter in Lemma 3, player i using $v_i^k+\zeta_i^k$ in its interaction terms rather than v_i^k is key to ensure that the average v_i^k among all players can accurately track the average decision x_i^k among all players. Note that different from [16] where players use correlated noise which restricts the strength of privacy protection, here the noises ζ_i^k ($i=1,2,\cdots,m$) of all players are completely independent of each other, and hence can enable strong differential privacy. However, it is worth noting that when a player i has only one neighboring player (say, player j), then according to the update rule in (6), the dynamics of the two players could be correlated, which may allow the neighboring player j to infer certain information of player i. This is a limitation of the conventional differential privacy framework which implicitly relies on a data curator/aggregator to collect data and inject noises, and hence, in the decentralized setting, implicitly assumes that players trust each other to cooperatively decide (like a data curator/aggregator) how to mask shared information. To completely avoid correlated dynamics among interacting players, the local differential privacy framework (see, e.g., [38]) can be exploited.*

The sequence $\{\gamma^k\}$, which diminishes with time, is used to suppress the influence of persistent differential-privacy noise ζ_j^k on the convergence point of the iterates. The stepsize sequence $\{\lambda^k\}$ and attenuation sequence $\{\gamma^k\}$ have to be designed appropriately to guarantee the accurate convergence of the iterate vector $x^k\triangleq[(x_1^k)^T,\cdots,(x_m^k)^T]^T$ to the Nash equilibrium point $x^*\triangleq[(x_1^*)^T,\cdots,(x_m^*)^T]^T$. It is worth noting that they are hard-coded into all players' programs and need no adjustment/coordination in implementation. The persistent differential-privacy noise sequences $\{\zeta_i^k\}, i\in[m]$ have zero-mean and γ^k-bounded (conditional) variances, which will be specified later in Assumption 4.

A. Convergence Analysis

To prove the convergence of the decision vector x^k to the Nash equilibrium x^*, we have to present some properties of the iterates. The first property pertains to the average of the estimates v_i^k, which is defined as $\bar{v}^k\triangleq\frac{1}{m}\sum_{i=1}^m v_i^k$. More specifically, we will prove that $\bar{v}^k$ is equal to the average of decisions $\bar{x}^k\triangleq\frac{1}{m}\sum_{i=1}^m x_i^k$. Namely, $\bar{v}^k$ captures the exact average decision. Such a property has been proven and employed in [5] in the absence of noise. Now we prove that this relationship still holds under our proposed Algorithm 1 even all agents add independent noises to their shared messages.

Lemma 3. *Under Assumption 3, we have $\bar{v}^k=\bar{x}^k$ for all $k\geq 0$.*

Proof. According to the definitions of $\bar{v}^k$ and $\bar{x}^k$, we only have to prove

$$\sum_{i=1}^m v_i^k=\sum_{i=1}^m x_i^k. \tag{7}$$

We prove the relationship in (7) using induction.

For $k=0$, the relationship holds trivially since we have initialized all v_i^k as $v_i^0=x_i^0$.

Next we proceed to prove that if (7) holds for some iteration $k>0$, i.e.,

$$\sum_{i=1}^m v_i^k=\sum_{i=1}^m x_i^k, \tag{8}$$

then it also holds for iteration $k+1$.

According to (6), we have

$$\begin{aligned}\sum_{i=1}^m v_i^{k+1}=&\sum_{i=1}^m v_i^k+\gamma^k\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}(v_j^k+\zeta_j^k-v_i^k-\zeta_i^k)\\&+\sum_{i=1}^m x_i^{k+1}-\sum_{i=1}^m x_i^k.\end{aligned} \tag{9}$$

Plugging (8) into (9) leads to

$$\sum_{i=1}^m v_i^{k+1}=\gamma^k\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}(v_j^k+\zeta_j^k-v_i^k-\zeta_i^k)+\sum_{i=1}^m x_i^{k+1}. \tag{10}$$

We decompose the first term (excluding γ^k) on the right hand side of (10) as

$$\begin{aligned}&\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}(v_j^k+\zeta_j^k-v_i^k-\zeta_i^k)\\&=\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}v_j^k-\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}v_i^k+\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}\zeta_j^k\\&\quad-\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}\zeta_i^k.\end{aligned} \tag{11}$$

Using the symmetric property of L_{ij} in Assumption 3, the preceding relationship can be rewritten as

$$\begin{aligned}&\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}(v_j^k+\zeta_j^k-v_i^k-\zeta_i^k)\\&=\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}v_j^k-\sum_{i=1}^m\sum_{i\in\mathbb{N}_j}L_{ji}v_i^k+\sum_{i=1}^m\sum_{j\in\mathbb{N}_i}L_{ij}\zeta_j^k\\&\quad-\sum_{i=1}^m\sum_{i\in\mathbb{N}_j}L_{ji}\zeta_i^k\\&=0.\end{aligned} \tag{12}$$

Plugging (12) into (10) leads to $\sum_{i=1}^m v_i^{k+1}=\sum_{i=1}^m x_i^{k+1}$, which completes the proof. ■

Using Lemma 3, we have the following results under Assumption 1:

Lemma 4. *Under Assumption 1, and v_i^k governed by Algorithm 1, the following inequalities hold for some $C>0$ and all $k\geq 0$:*

$$\|F_i(x_i^k,\bar{v}^k)\|\leq C,\quad \|F_i(x_i^k,v_i^k)\|\leq C+\tilde{L}\|v_i^k-\bar{v}^k\|. \tag{13}$$

Proof. According to Lemma 3, we have $\bar{v}^k=\bar{x}^k$ for all $k\geq 0$, where $\bar{x}^k\triangleq\frac{\sum_{i=1}^m x_i^k}{m}$. Hence, $\bar{v}^k\in\bar{K}$, where $\bar{K}$ is compact since each K_i is compact according to Assumption 1. From

Assumption 1, $F_i(x_i^k, \bar{x}^k)$ is continuous over $K_i \times \bar{K}$, so we have the first inequality.

To show the second inequality, we use the following relationship

$$\begin{aligned}\|F_i(x_i^k, v_i^k)\| &= \|F_i(x_i^k, v_i^k) - F_i(x_i^k, \bar{x}^k) + F_i(x_i^k, \bar{x}^k)\| \\ &\leq \|F_i(x_i^k, v_i^k) - F_i(x_i^k, \bar{v}^k)\| + \|F_i(x_i^k, \bar{v}^k)\|.\end{aligned}$$

Then using the Lipschitz continuous condition in Assumption 2 and the proven fact that $\|F_i(x_i^k, \bar{v}^k)\|$ is bounded, we can arrive at the second inequality in (13). ■

Remark 4. *Note that different from [19], [20] whose convergence analysis requires $F_i(x_i, v_i^k)$ to be uniformly bounded in the presence of noise, we will provide a new proof technique that removes the uniformly bounded constraint in convergence analysis. This relaxation is significant in that under differential-privacy design, v_i^k will be subject to unbounded noise, such as Laplace noise or Gaussian noise, and becomes unbounded. Therefore, restricting $F_i(x_i, v_i^k)$ to be uniformly bounded with respect to v_i^k will significantly limit the applicability of the algorithm. For example, in the Nash-Cournot market game considered in the numerical simulations in Sec. V, the sale price function (the inverse demand function) is usually modeled as a function decreasing linearly with the aggregative production, which will result in a mapping $F_i(x_i, v_i^k)$ that is not uniformly bounded.*

We now apply Lemma 2 to arrive at a general convergence theory for distributed algorithms for the problem in (1):

Proposition 1. *Assume that problem (1) has a Nash equilibrium $x^* = [(x_1^*)^T, (x_2^*)^T, \ldots, (x_m^*)^T]^T \in \mathbb{R}^{md}$. Suppose that a distributed algorithm generates sequences $\{x_i^k\} \subseteq \mathbb{R}^d$ and $\{v_i^k\} \subseteq \mathbb{R}^d$ such that almost surely we have*

$$\begin{aligned}&\begin{bmatrix} \mathbb{E}\left[\sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2 | \mathcal{F}^k\right] \\ \mathbb{E}\left[\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2 | \mathcal{F}^k\right] \end{bmatrix} \\ &\leq \left(\begin{bmatrix} 1 & \kappa_1\gamma^k \\ 0 & 1-\kappa_2\gamma^k \end{bmatrix} + a^k \mathbf{1}\mathbf{1}^T \right) \begin{bmatrix} \sum_{i=1}^m \|x_i^k - x_i^*\|^2 \\ \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 \end{bmatrix} \\ &\quad + b^k\mathbf{1} - c^k \begin{bmatrix} \left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*) \\ 0 \end{bmatrix}, \quad \forall k \geq 0\end{aligned} \tag{14}$$

where $\bar{v}^k = \frac{1}{m}\sum_{i=1}^m v_i^k$, $\mathcal{F}^k = \{x_i^\ell, v_i^\ell, i \in [m], 0 \leq \ell \leq k\}$, the random nonnegative scalar sequences $\{a^k\}$, $\{b^k\}$ satisfy $\sum_{k=0}^\infty a^k < \infty$ and $\sum_{k=0}^\infty b^k < \infty$ almost surely, the deterministic nonnegative sequences $\{c^k\}$ and $\{\gamma^k\}$ satisfy $\sum_{k=0}^\infty c^k = \infty$ and $\sum_{k=0}^\infty \gamma^k = \infty$, and the scalars κ_1 and κ_2 satisfy $\kappa_1 > 0$ and $0 < \kappa_2\gamma^k < 1$, respectively, for all $k \geq 0$. Then, we have $\lim_{k\to\infty} \|v_i^k - \bar{v}^k\| = 0$ almost surely for all i, and $\lim_{k\to\infty} \|x_i^k - x_i^*\| = 0$ almost surely.

Proof. According to Assumption 1, we always have $\left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*) > 0$ for all k. Hence, by letting $\mathbf{v}^k = \left[\sum_{i=1}^m \|x_i^k - x_i^*\|^2, \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2\right]^T$, from relation (14) it follows that almost surely for all $k \geq 0$,

$$\mathbb{E}\left[\mathbf{v}^{k+1} | \mathcal{F}^k\right] \leq \left(\begin{bmatrix} 1 & \kappa_1\gamma^k \\ 0 & 1-\kappa_2\gamma^k \end{bmatrix} + a^k\mathbf{1}\mathbf{1}^T \right) \mathbf{v}^k + b^k\mathbf{1}. \tag{15}$$

Consider the vector $\pi = [1, \frac{\kappa_1}{\kappa_2}]^T$ and note

$$\pi^T \begin{bmatrix} 1 & \kappa_1\gamma^k \\ 0 & 1-\kappa_2\gamma^k \end{bmatrix} = \pi^T.$$

Thus, relation (15) satisfies all conditions of Lemma 2. By Lemma 2, it follows that $\lim_{k\to\infty} \pi^T\mathbf{v}^k$ exists almost surely, and that the sequences $\{\sum_{i=1}^m \|x_i^k - x_i^*\|^2\}$ and $\{\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2\}$ are bounded almost surely. From (15), we have the following relationship almost surely for the second element of $\mathbf{v}^k$:

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2 | \mathcal{F}^k\right] \\ &\leq (1 + a^k - \kappa_2\gamma^k) \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 + \beta^k \quad \forall k \geq 0,\end{aligned}$$

where $\beta^k = a^k \left(\sum_{i=1}^m \left(\|x_i^k - x_i^*\|^2 + \|v_i^k - \bar{v}^k\|^2\right)\right)$. Since $\sum_{k=0}^\infty a^k < \infty$ holds almost surely by our assumption, and the sequences $\{\sum_{i=1}^m \|x_i^k - x_i^*\|^2\}$ and $\{\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2\}$ are bounded almost surely, it follows that $\sum_{k=0}^\infty \beta^k < \infty$ holds almost surely. Thus, the preceding relation satisfies the conditions of Lemma 1 with $v^k = \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$, $q^k = \kappa_2\gamma^k$, and $p^k = \beta^k$ due to our assumptions $\sum_{k=0}^\infty b^k < \infty$ almost surely and $\sum_{k=0}^\infty \gamma^k = \infty$. By Lemma 1, it follows that almost surely

$$\sum_{k=0}^\infty \kappa_2\gamma^k \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 < \infty, \ \lim_{k\to\infty} \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 = 0. \tag{16}$$

It remains to show that $\sum_{i=1}^m \|x_i^k - x_i^*\|^2 \to 0$ almost surely. For this, we use Lemma 2. Under the assumption that $\{a^k\}$ and $\{b^k\}$ are summable, we have that the inequality in (14) satisfies the relationship in Lemma 2 with $\mathbf{v}^k = \left[\sum_{i=1}^m \|x_i^k - x_i^*\|^2, \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2\right]^T$, $V^k = \begin{bmatrix} 1 & \kappa_1\gamma^k \\ 0 & 1-\kappa_2\gamma^k \end{bmatrix}$, $H^k = \begin{bmatrix} c^k & 0 \\ 0 & 0 \end{bmatrix}$, and $\pi^T = [1, \frac{\kappa_1}{\kappa_2}]^T$. Therefore, according to Lemma 2, we know that $\pi^T\mathbf{v}^k$ converges almost surely, i.e., $\sum_{i=1}^m \|x_i^k - x_i^*\|^2 + \frac{\kappa_1}{\kappa_2}\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ converges almost surely. Given that we have proven that $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ converges almost surely (see (16)), we have that $\sum_{i=1}^m \|x_i^k - x_i^*\|^2$ (or $\|x^k - x^*\|^2$) converges almost surely. According to Lemma 2, we also have $\sum_{k=0}^\infty \pi^T H^k \mathbf{u}^k < \infty$ almost surely, i.e.,

$$\sum_{k=0}^\infty \left[1, \frac{\kappa_1}{\kappa_2}\right]^T \begin{bmatrix} c^k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*) \\ 0 \end{bmatrix} < \infty,$$

or

$$\sum_{k=0}^\infty c^k \left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*) < \infty. \tag{17}$$

Now using (17) and the proven fact that $\|x^k - x^*\|^2$ converges almost surely, we proceed to prove that x^k converges to x^* almost surely. Because the augmented state decision vector x^k belongs to the compact set K defined in (2), we know that the sequence $\{x^k\}$ must have accumulation points in K. So the condition $\sum_{k=0}^\infty c^k = \infty$ and (17) mean that there exists a subsequence of $\{x^k\}$, say $\{x^{k_\ell}\}$, along which $\left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*)$ converges to zero almost surely.

Recalling that $\phi(\cdot)$ is strictly monotone (see Assumption 1), one has that the subsequence $\{x^{k_\ell}\}$ must converge to x^* almost surely. This and the fact that $\|x^k - x^*\|^2$ converges almost surely imply that x^k converges to x^* almost surely. ■

We also need the following Lemma about matrix L:

Lemma 5. *Under Assumption 3 and a positive sequence $\{\gamma^k\}$ satisfying $\sum_{k=0}^{\infty}\gamma^k = \infty$ and $\sum_{k=0}^{\infty}(\gamma^k)^2 < \infty$, there always exists a $T > 0$ such that when $k \geq T$, we always have*

$$\|I + \gamma^k L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| \leq 1 - \gamma^k|\rho_2|,$$

where ρ_2 is the second largest eigenvalue of L.

Proof. Under Assumption 3, the matrix L is symmetric, so we have that all eigenvalues of L are real numbers. Since L has non-negative off-diagonal entries and the diagonal entries L_{ii} are given by $L_{ii} = -\sum_{j\in\mathbb{N}_i} L_{ij}$, we know that all eigenvalues of L are non-positive (according to the Gershgorin circle theorem), and there is always an eigenvalue equal to 0. Arrange the eigenvalues of L as $\rho_m \leq \rho_{m-1} \leq \cdots \leq \rho_2 \leq \rho_1 = 0$. It can be verified that the eigenvalues of $I + L$ are given by $1+\rho_m \leq 1+\rho_{m-1} \leq \cdots \leq 1+\rho_2 \leq 1+\rho_1 = 1$, and the eigenvalues of $I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}$ are given by $\{1+\rho_m,\, 1+\rho_{m-1},\, \cdots,\, 1+\rho_2,\, 0\}$. Furthermore, the condition $\|I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| < 1$ in Assumption 3 implies that only one eigenvalue of L is zero, and its all other eigenvalues are strictly less than 0. Hence, we have $\rho_m \leq \rho_{m-1} \leq \cdots \leq \rho_2 < 0$, i.e., $|\rho_m| \geq |\rho_{m-1}| \geq \cdots \geq |\rho_2| > 0$. Since the eigenvalues of $I+\gamma^k L-\frac{\mathbf{1}\mathbf{1}^T}{m}$ are $\{1+\gamma^k\rho_m,\, 1+\gamma^k\rho_{m-1},\, \cdots,\, 1+\gamma^k\rho_2,\, 0\}$, we have the norm $\|I + \gamma^k L - \frac{\mathbf{1}\mathbf{1}^T}{m}\|$ being no larger than $|1 + \gamma^k\rho_m|$ or $|1 + \gamma^k\rho_2|$. Further taking into account the fact that $\{\gamma^k\}$ is square summable and hence γ^k decays to zero, we have that there always exists a $T > 0$ such that $|1 + \gamma^k\rho_m| = 1 - \gamma^k|\rho_m|$ and $|1 + \gamma^k\rho_2| = 1 - \gamma^k|\rho_2|$ hold for $k \geq T$. Given $|\rho_m| \geq |\rho_2|$, we have the stated result of the Lemma. ■

Using Proposition 1, we are in position to establish convergence of Algorithm 1 assuming that persistent differential-privacy noise satisfies the following assumption:

Assumption 4. *For every $i \in [m]$ and every k, conditional on $\mathcal{F}^k = \{v^0, \ldots, v^k\}$, the random noise ζ_i^k satisfies $\mathbb{E}\left[\zeta_i^k \mid \mathcal{F}^k\right] = 0$ and $\mathbb{E}\left[\|\zeta_i^k\|^2 \mid \mathcal{F}^k\right] = (\sigma_i^k)^2$ for all $k \geq 0$, and*

$$\sum_{k=0}^{\infty}(\gamma^k)^2 \max_{i\in[m]}(\sigma_i^k)^2 < \infty, \tag{18}$$

where $\{\gamma^k\}$ is the attenuation sequence from Algorithm 1. The initial random vectors satisfy

$$\mathbb{E}\left[\|v_i^0\|^2\right] < \infty, \quad \forall i \in [m].$$

Remark 5. *Given that γ^k decreases with time, (18) can be satisfied even when $\{\sigma_i^k\}$ increases with time. For example, under $\gamma^k = \mathcal{O}(\frac{1}{k^{0.9}})$, an increasing $\{\sigma_i^k\}$ with increasing rate no faster than $\mathcal{O}(k^{0.3})$ still satisfies the summable condition in (18). Allowing $\{\sigma_i^k\}$ to be increasing with time is key to enabling the strong ϵ-differential privacy in Theorem 2. In addition, this assumption allows σ_i^k to be different for different i, i.e., it allows different players to inject noises with different variances.*

Theorem 1. *Under Assumptions 1, 2, 3, and 4, if there exists some $T \geq 0$ such that γ^k and λ^k satisfy the following conditions:*

$$\sum_{k=T}^{\infty}\gamma^k = \infty,\ \sum_{k=T}^{\infty}\lambda^k = \infty,\ \sum_{k=T}^{\infty}(\gamma^k)^2 < \infty,\ \sum_{k=T}^{\infty}\frac{(\lambda^k)^2}{\gamma^k} < \infty,$$

then Algorithm 1 converges to the Nash equilibrium of the game in (1) almost surely.

Proof. The basic idea is to apply Proposition 1 to the quantities $\sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2$ and $\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2$. Since the results of Proposition 1 are asymptotic, they remain valid when the starting index is shifted from $k = 0$ to $k = T$, for an arbitrary $T \geq 0$. We divide the proof into two parts to analyze $\sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2$ and $\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2$, respectively.

Part I: We first analyze $\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2$. For the convenience of analysis, we write the iterates of v_i^k on per-coordinate expressions. Define for all $\ell = 1, \ldots, d$, and $k \geq 0$, $v^k(\ell) = \left[[v_1^k]_\ell, \ldots, [v_m^k]_\ell\right]^T$ where $[v_i^k]_\ell$ represents the ℓth element of the vector v_i^k. Similarly, we define $x^k(\ell) = \left[[x_1^k]_\ell, \ldots, [x_m^k]_\ell\right]^T$ and $\zeta^k(\ell) = \left[[\zeta_1^k]_\ell, \ldots, [\zeta_m^k]_\ell\right]^T$. In this per-coordinate view, (6) has the following form for all $\ell = 1, \ldots, d$, and $k \geq 0$:

$$v^{k+1}(\ell) = v^k(\ell) + \gamma^k L v^k(\ell) + \gamma^k L\zeta^k(\ell) + x^{k+1}(\ell) - x^k(\ell). \tag{19}$$

Note that the diagonal entries of L are defined as $L_{ii} \triangleq -\sum_{j\in\mathbb{N}_i} L_{ij}$.

The dynamics of the average v_i^k, i.e., $\bar{v}^k$, is given by

$$\begin{aligned}[\bar{v}^{k+1}]_\ell &= \frac{\mathbf{1}^T}{m} v^{k+1}(\ell) \\ &= \frac{\mathbf{1}^T}{m}\left(v^k(\ell) + \gamma^k L v^k(\ell) + \gamma^k L\zeta^k(\ell) + x^{k+1}(\ell) - x^k(\ell)\right),\end{aligned} \tag{20}$$

where $[\bar{v}^{k+1}]_\ell$ represents the ℓ-th element of $\bar{v}^{k+1}$.

Under Assumption 3, we have $\mathbf{1}^T L = 0$, which simplifies the preceding equation (20) to:

$$[\bar{v}^{k+1}]_\ell = \frac{\mathbf{1}^T}{m}\left(v^k(\ell) + x^{k+1}(\ell) - x^k(\ell)\right), \tag{21}$$

where $[\bar{x}^{k+1}]_\ell$ represents the ℓ-th element of the vector $\bar{x}^{k+1}$.

Combining (19), (20), and (21) yields

$$\begin{aligned}v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell &= \left(I + \gamma^k L - \frac{\mathbf{1}\mathbf{1}^T}{m}\right)v^k(\ell) + \gamma^k L\zeta^k(\ell) \\ &\quad + \left(I - \frac{\mathbf{1}\mathbf{1}^T}{m}\right)\left(x^{k+1}(\ell) - x^k(\ell)\right).\end{aligned} \tag{22}$$

For the sake of notational simplicity, we define

$$W^k \triangleq I + \gamma^k L - \frac{\mathbf{1}\mathbf{1}^T}{m}, \quad \Pi^k \triangleq I - \frac{\mathbf{1}\mathbf{1}^T}{m}. \tag{23}$$

It can be verified that $W^k\mathbf{1}=0$ holds, and hence $W^k\mathbf{1}[\bar{v}^k]_\ell = 0$ always holds for any $1 \le \ell \le m$ under Assumption 3. Therefore, (22) can be rewritten as

$$v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell = W^k\left(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell\right) + \gamma^k L\zeta^k(\ell) + \Pi^k\left(x^{k+1}(\ell) - x^k(\ell)\right), \tag{24}$$

which further implies

$$\begin{aligned}
&\|v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell\|^2 \\
&= \left\|W^k(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell) + \Pi^k(x^{k+1}(\ell) - x^k(\ell))\right\|^2 \\
&+ 2\left\langle W^k(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell) + \Pi^k(x^{k+1}(\ell) - x^k(\ell)),\, \gamma^k L\zeta^k(\ell)\right\rangle \\
&+ \|\gamma^k L\zeta^k(\ell)\|^2 \\
&\le \left\|W^k(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell) + \Pi^k(x^{k+1}(\ell) - x^k(\ell))\right\|^2 \\
&+ 2\left\langle W^k(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell) + \Pi^k(x^{k+1}(\ell) - x^k(\ell)),\, \gamma^k L\zeta^k(\ell)\right\rangle \\
&+ (\gamma^k)^2\|L\|^2\zeta^k(\ell)\|^2.
\end{aligned} \tag{25}$$

Taking the conditional expectation, given $\mathcal{F}^k = \{v^0, \ldots, v^k\}$, and using Assumption 4, from the preceding relation we obtain for all $k \ge 0$:

$$\begin{aligned}
&\mathbb{E}\left[\|v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell\|^2\mathcal{F}^k\right] \\
&\le \left\|W^k(v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell) + \Pi^k(x^{k+1}(\ell) - x^k(\ell))\right\|^2 \\
&\quad + (\gamma^k)^2\|L\|^2\mathbb{E}\left[\|\zeta^k(\ell)\|^2\right] \\
&\le \left(\|W^k\|\|v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell\| + \|\Pi^k\|\|x^{k+1}(\ell) - x^k(\ell)\|\right)^2 \\
&\quad + (\gamma^k)^2\|L\|^2\mathbb{E}\left[\|\zeta^k(\ell)\|^2\right].
\end{aligned} \tag{26}$$

Now we analyze the first term on the right hand side of the preceding inequality. Combined with the facts that $\|\Pi^k\| = 1$ and there exists a $T \ge 0$ such that $0 < \|W^k\| \le 1 - \gamma^k|\rho_2|$ holds for $k \ge T$ (see Lemma 5), equation (26) implies that there always exists a $T \ge 0$ such that the following inequality always holds for $k \ge T$:

$$\begin{aligned}
&\mathbb{E}\left[\|v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell\|^2\mathcal{F}^k\right] \\
&\le \left((1-\gamma^k|\rho_2|)\|v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell\| + \|x^{k+1}(\ell) - x^k(\ell)\|\right)^2 \\
&\quad + (\gamma^k)^2\|L\|^2\mathbb{E}\left[\|\zeta^k(\ell)\|^2\right].
\end{aligned} \tag{27}$$

Using the inequality $(a+b)^2 \le (1+\epsilon)a^2 + (1+\epsilon^{-1})b^2$ valid for any scalars a, b, and $\epsilon > 0$, we further have

$$\begin{aligned}
&\mathbb{E}\left[\|v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell\|^2\mathcal{F}^k\right] \\
&\le (1+\epsilon)(1-\gamma^k|\rho_2|)^2\|v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell\|^2 \\
&+ (1+\epsilon^{-1})\|\, x^{k+1}(\ell) - x^k(\ell)\|^2 + (\gamma^k)^2\|L\|^2\mathbb{E}\left[\|\zeta^k(\ell)\|^2\right].
\end{aligned} \tag{28}$$

Setting $\epsilon = \frac{\gamma^k|\rho_2|}{1-\gamma^k|\rho_2|}$ (which leads to $(1+\epsilon) = \frac{1}{1-\gamma^k|\rho_2|}$ and $1+\epsilon^{-1} = \frac{1}{\gamma^k|\rho_2|}$) yields

$$\begin{aligned}
&\mathbb{E}\left[\|v^{k+1}(\ell) - \mathbf{1}[\bar{v}^{k+1}]_\ell\|^2|\mathcal{F}^k\right] \\
&\le (1-\gamma^k|\rho_2|)\|v^k(\ell) - \mathbf{1}[\bar{v}^k]_\ell\|^2 \\
&+ \frac{1}{\gamma^k|\rho_2|}\|x^{k+1}(\ell) - x^k(\ell)\|^2 + (\gamma^k)^2\|L\|^2\mathbb{E}\left[\|\zeta^k(\ell)\|^2\right].
\end{aligned} \tag{29}$$

Summing these relations over $\ell = 1, \ldots, d$, and noting $\sum_{\ell=1}^d \|v^k(\ell) - [\bar{v}^k]_\ell\mathbf{1}\|^2 = \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$, $\sum_{\ell=1}^d \|x^{k+1}(\ell) - x^k(\ell)\|^2 = \sum_{i=1}^m \|x_i^{k+1} - x_i^k\|^2$, and $\sum_{\ell=1}^d \|\zeta^k(\ell)\|^2 = \sum_{i=1}^m \|\zeta_i^k\|^2$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2\mathcal{F}^k\right] \\
&\le (1-\gamma^k|\rho_2|)\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 \\
&+ \frac{1}{\gamma^k|\rho_2|}\sum_{i=1}^m \|x_i^{k+1} - x_i^k\|^2 + (\gamma^k)^2\|L\|^2\sum_{i=1}^m (\sigma_i^k)^2.
\end{aligned} \tag{30}$$

Next, we characterize $\sum_{i=1}^m \|x_i^{k+1} - x_i^k\|^2$. According to (6), we have

$$\begin{aligned}
\|x_i^{k+1} - x_i^k\| &= \left\|\Pi_{K_i}\left[x_i^k - \lambda^k F_i(x_i^k, v_i^k)\right] - x_i^k\right\| \\
&\le \left\|x_i^k - \lambda^k F_i(x_i^k, v_i^k) - x_i^k\right\| \\
&= \lambda^k\|F_i(x_i^k, v_i^k)\| \le \lambda^k C + \lambda^k\tilde{L}\|v_i^k - \bar{v}^k\|,
\end{aligned} \tag{31}$$

where in the last inequality we used Lemma 4. The preceding inequality further implies

$$\|x_i^{k+1} - x_i^k\|^2 \le 2(\lambda^k)^2C^2 + 2(\lambda^k)^2\tilde{L}^2\|v_i^k - \bar{v}^k\|^2. \tag{32}$$

Plugging (32) into (30) yields

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^m \|v_i^{k+1} - \bar{v}^{k+1}\|^2\mathcal{F}^k\right] \\
&\le (1-\gamma^k|\rho_2|)\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 + \frac{2(\lambda^k)^2\tilde{L}^2}{\gamma^k|\rho_2|}\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 \\
&\quad + \frac{2m(\lambda^k)^2C^2}{\gamma^k|\rho_2|} + (\gamma^k)^2\|L\|^2\sum_{i=1}^m (\sigma_i^k)^2.
\end{aligned} \tag{33}$$

Part II: Next, we analyze $\sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2$.

At the Nash equilibrium $x^* = [(x_1^*)^T, (x_2^*)^T, \ldots, (x_m^*)^T]^T$, we always have $x_i^* = \Pi_{K_i}\left[x_i^* - \lambda^k F_i(x_i^*, \bar{x}^*)\right]$, where $\bar{x}^* = \frac{1}{m}\sum_{i=1}^m x_i^*$. Therefore, using (6), we have

$$\begin{aligned}
&\left\|x_i^{k+1} - x_i^*\right\|^2 \\
&= \left\|\Pi_{K_i}\left[x_i^k - \lambda^k F_i(x_i^k, v_i^k)\right] - x_i^*\right\|^2 \\
&= \left\|\Pi_{K_i}\left[x_i^k - \lambda^k F_i(x_i^k, v_i^k)\right] - \Pi_{K_i}\left[x_i^* - \lambda^k F_i(x_i^*, \bar{x}^*)\right]\right\|^2 \\
&\le \left\|x_i^k - x_i^* - \lambda^k(F_i(x_i^k, v_i^k) - F_i(x_i^*, \bar{x}^*))\right\|^2 \\
&\le \left\|x_i^k - x_i^*\right\|^2 + (\lambda^k)^2\left\|F_i(x_i^k, v_i^k) - F_i(x_i^*, \bar{x}^*)\right\|^2 \\
&\quad - 2\left\langle x_i^k - x_i^*,\, \lambda^k(F_i(x_i^k, v_i^k) - F_i(x_i^*, \bar{x}^*))\right\rangle.
\end{aligned} \tag{34}$$

By adding and subtracting $F_i(x_i^k, \bar{v}^k)$ to the inner-product term, we arrive at

$$\begin{aligned}
&\left\|x_i^{k+1} - x_i^*\right\|^2 \\
&\le \left\|x_i^k - x_i^*\right\|^2 + \underbrace{(\lambda^k)^2\left\|F_i(x_i^k, v_i^k) - F_i(x_i^*, \bar{x}^*)\right\|^2}_{\text{Term 1}} \\
&\quad - \underbrace{2\left\langle x_i^k - x_i^*,\, \lambda^k(F_i(x_i^k, v_i^k) - F_i(x_i^k, \bar{v}^k))\right\rangle}_{\text{Term 2}} \\
&\quad - \underbrace{2\left\langle x_i^k - x_i^*,\, \lambda^k(F_i(x_i^k, \bar{v}^k) - F_i(x_i^*, \bar{x}^*))\right\rangle}_{\text{Term 3}}.
\end{aligned} \tag{35}$$

Next, we characterize the three terms on the right hand side of (35), i.e., Term 1, Term 2, and Term 3, respectively.

Using Lemma 4, we can bound Term 1 as follows:

$$\begin{aligned}\text{Term 1} &\leq 2(\lambda^k)^2 \left\|F_i(x_i^k, v_i^k)\right\|^2 + 2(\lambda^k)^2 \left\|F_i(x_i^*, \bar{x}^*)\right\|^2 \\ &\leq 2(\lambda^k)^2 \left(C + \tilde{L}\|v_i^k - \bar{v}^k\|\right)^2 + 2(\lambda^k)^2 C^2 \\ &\leq 4(\lambda^k)^2 C^2 + 4(\lambda^k)^2 \tilde{L}^2 \|v_i^k - \bar{v}^k\|^2 + 2(\lambda^k)^2 C^2 \\ &= 6(\lambda^k)^2 C^2 + 4(\lambda^k)^2 \tilde{L}^2 \|v_i^k - \bar{v}^k\|^2.\end{aligned} \tag{36}$$

Applying Cauchy–Schwarz inequality to Term 2 yields

$$\begin{aligned}&\text{Term 2} \\ &\geq -2\lambda^k \|x_i^k - x_i^*\| \left\|F_i(x_i^k, v_i^k) - F_i(x_i^k, \bar{v}^k)\right\| \\ &\geq -\frac{(\lambda^k)^2 \|x_i^k - x_i^*\|^2}{\gamma^k} - \gamma^k \left\|F_i(x_i^k, v_i^k) - F_i(x_i^k, \bar{v}^k)\right\|^2 \\ &\geq -\frac{(\lambda^k)^2 \|x_i^k - x_i^*\|^2}{\gamma^k} - \gamma^k \tilde{L}^2 \left\|v_i^k - \bar{v}^k\right\|^2,\end{aligned} \tag{37}$$

where in the last inequality we used Assumption 2, and in the second inequality we used the inequality $2ab \leq \frac{a^2}{\epsilon} + \epsilon b^2$ valid for any $a \in \mathbb{R}$, $b \in \mathbb{R}$, and $\epsilon > 0$.

We use Lemma 3 to treat Term 3. According to Lemma 3, we always have $\bar{v}^k = \bar{x}^k$, which further leads to

$$\begin{aligned}\text{Term 3} &= 2\left\langle x_i^k - x_i^*,\, \lambda^k (F_i(x_i^k, \bar{x}^k) - F_i(x_i^*, \bar{x}^*))\right\rangle \\ &= 2\lambda^k \left(F_i(x_i^k, \bar{x}^k) - F_i(x_i^*, \bar{x}^*)\right)^T (x_i^k - x_i^*).\end{aligned} \tag{38}$$

Plugging (36), (37), and (38) into (35) yields

$$\begin{aligned}\left\|x_i^{k+1} - x_i^*\right\|^2 &\leq \left\|x_i^k - x_i^*\right\|^2 + 4(\lambda^k)^2 \tilde{L}^2 \|v_i^k - \bar{v}^k\|^2 \\ &+ 6(\lambda^k)^2 C^2 + \frac{(\lambda^k)^2 \|x_i^k - x_i^*\|^2}{\gamma^k} + \gamma^k \tilde{L}^2 \left\|v_i^k - \bar{v}^k\right\|^2 \\ &- 2\lambda^k \left(F_i(x_i^k, \bar{x}^k) - F_i(x_i^*, \bar{x}^*)\right)^T (x_i^k - x_i^*).\end{aligned} \tag{39}$$

Summing (39) from $i = 1$ to $i = m$ yields

$$\begin{aligned}\sum_{i=1}^m \left\|x_i^{k+1} - x_i^*\right\|^2 &\leq \sum_{i=1}^m \left\|x_i^k - x_i^*\right\|^2 + 4(\lambda^k)^2 \tilde{L}^2 \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2 \\ &+ \frac{(\lambda^k)^2 \sum_{i=1}^m \|x_i^k - x_i^*\|^2}{\gamma^k} + \gamma^k \tilde{L}^2 \sum_{i=1}^m \left\|v_i^k - \bar{v}^k\right\|^2 \\ &+ 6m(\lambda^k)^2 C^2 - 2\lambda^k \left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*).\end{aligned} \tag{40}$$

From Assumption 1, we know that $\left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*)$ in the last term on the right hand side of the proceeding inequality is positive for all $x^k \neq x^*$.

Next, we combine Step I and Step II to prove the theorem.

Defining $\mathbf{v}^k = \left[\sum_{i=1}^m \|x_i^k - x_i^*\|^2,\ \sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2\right]^T$, we have the following relations from (33) and (40):

$$\mathbb{E}\left[\mathbf{v}^{k+1} | \mathcal{F}^k\right] \leq (V^k + A^k)\mathbf{v}^k - 2\lambda^k \Phi^k + B^k, \tag{41}$$

where

$$V^k = \begin{bmatrix} 1 & \tilde{L}^2 \gamma^k \\ 0 & 1 - \gamma^k |\rho_2| \end{bmatrix},$$

$$A^k = \begin{bmatrix} \frac{(\lambda^k)^2}{\gamma^k} & 4(\lambda^k)^2 \tilde{L}^2 \\ 0 & \frac{2(\lambda^k)^2 \tilde{L}^2}{\gamma^k |\rho_2|} \end{bmatrix},$$

$$\Phi^k = \begin{bmatrix} \left(\phi(x^k) - \phi(x^*)\right)^T (x^k - x^*) \\ 0 \end{bmatrix},$$

$$B^k = \begin{bmatrix} 6m(\lambda^k)^2 C^2 \\ \frac{2m(\lambda^k)^2 C^2}{\gamma^k |\rho_2|} + (\gamma^k)^2 \|L\|^2 \sum_{i=1}^m (\sigma_i^k)^2 \end{bmatrix}.$$

Using Assumption 4 and the conditions of the theorem $\sum_{k=T}^{\infty} (\gamma^k)^2 < \infty$ and $\sum_{k=T}^{\infty} \frac{(\lambda^k)^2}{\gamma^k} < \infty$, we have that all elements of the matrices of A^k and B^k are summable. Therefore, we have $\sum_{i=1}^m \|x_i^k - x_i^*\|^2$ and $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ satisfying the conditions of Proposition 1 with $\kappa_1 = \tilde{L}^2$, $\kappa_2 = |\rho_2|$, $c^k = 2\lambda^k$, $a^k = \max\{\frac{(\lambda^k)^2}{\gamma^k},\ 4(\lambda^k)^2 \tilde{L}^2,\ \frac{2(\lambda^k)^2 \tilde{L}^2}{\gamma^k |\rho_2|}\}$, and $b^k = \max\{6m(\lambda^k)^2 C^2,\ \frac{2m(\lambda^k)^2 C^2}{\gamma^k |\rho_2|} + (\gamma^k)^2 \|L\|^2 \sum_{i=1}^m (\sigma_i^k)^2\}$. ■

Remark 6. *The conditions for $\{\gamma^k\}$ and $\{\lambda^k\}$ can be satisfied, e.g., by setting $\lambda^k = \frac{c_1}{1 + c_2 k}$ and $\gamma^k = \frac{c_3}{1 + c_4 k^{\varrho}}$ with any $0.5 < \varrho < 1$, $c_1 > 0$, $c_2 > 0$, $c_3 > 0$, and $c_4 > 0$.*

Remark 7. *In the derivation, it can be seen that the aggregate-estimation error $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|$ does not decrease geometrically with k, which makes it impossible to use existing proof techniques for distributed Nash-equilibrium computation algorithms. In fact, in existing distributed Nash-equilibrium computation algorithms (e.g., [5], [23], [24], [25], [26]) and their stochastic variants (e.g., [27] and [20]), because the inter-player interaction is persistent, the aggregate-estimation error $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|$ always decreases geometrically, which makes it possible to separate the evolution analysis of the aggregate-estimation error and the decision distance from the Nash equilibrium. However, in the proposed algorithm, the diminishing γ^k leads to a non-geometric decreasing of the aggregate-estimation error, which makes it impossible to analyze the evolution of the aggregate estimate v_i^k and the decision x_i^k separately, and hence makes the proposed proof technique fundamentally different from existing analysis.*

Remark 8. *Communication imperfections can be modeled as channel noises, which can be regarded as the differential-privacy noise here. Therefore, Algorithm 1 can also counteract such communication imperfections in distributed equilibrium seeking of aggregative games.*

Remark 9. *Using Lemma 4 in [21], we can obtain from (33) that $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ decreases to zero with a rate of $\mathcal{O}((\frac{\lambda^k}{\gamma^k})^2)$. Similarly, when $\phi(x)$ is strongly monotone, i.e., $\left(\phi(x) - \phi(x')\right)^T (x - x') > c\|x - x'\|^2$ for some $c > 0$ and all $x, x' \in K$, we can obtain that $\sum_{i=1}^m \|x_i^{k+1} - x_i^*\|^2$ decreases to zero with a rate of $\mathcal{O}(\frac{\lambda^k}{\gamma^k})$. Moreover, from (33), it can be seen that the decreasing speed of $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ increases with an increase in $|\rho_2|$, which corresponds to the second largest eigenvalue of L. Therefore, the decreasing speed of $\sum_{i=1}^m \|v_i^k - \bar{v}^k\|^2$ to zero increases with an increase* *in the absolute value of the second largest eigenvalue of L in Assumption 3. Given that $-L$ corresponds to the conventional Laplacian weight matrix and an increase in $|\rho_2|$ implies a stronger network connectivity according to the algebraic graph theory [32], we have that the converging speed of $\sum_{i=1}^{m} \|v_i^k - \bar{v}^k\|^2$ to zero increases with an increase in network connectivity. Of course, since the algorithm requires each agent to share its noisy estimate with all its neighbors, a stronger connectivity could also mean more information exchange.*

B. Privacy Analysis for Algorithm 1

Note that in the proposed algorithm, the output is the estimation of the aggregate decision, i.e., $v^k \triangleq [(v_1^k)^T, \cdots, (v_m^k)^T]^T$. Then, following the idea of differential-privacy design for distributed optimization in [34], we define the sensitivity of a distributed Nash-equilibrium seeking algorithm to problem (1) as follows (see Sec. II.B. for the definition of the mapping $\mathcal{A}_{\mathcal{P},\mathcal{O},\vartheta^0}[k]$):

Definition 3. *At each iteration k, for any initial state ϑ^0 and any adjacent distributed games $\mathcal{P}$ and $\mathcal{P}'$, the sensitivity of a Nash-equilibrium seeking algorithm is*

$$\Delta^k \triangleq \sup_{\mathcal{O}\in\mathbb{O}} \left\{ \sup_{v^k\in\mathcal{A}_{\mathcal{P},\mathcal{O},\vartheta^0}[k],\, v'^k\in\mathcal{A}_{\mathcal{P}',\mathcal{O},\vartheta^0}[k]} \|v^k - v'^k\|_1 \right\}. \tag{42}$$

Then, we have the following lemma:

Lemma 6. *In Algorithm 1, at each iteration k, if each player adds a noise vector $\zeta_i^k \in \mathbb{R}^d$ consisting of d independent Laplace noises with parameter ν^k to shared messages v_i^k such that $\sum_{k=1}^{T_0} \frac{\Delta^k}{\nu^k} \le \bar{\epsilon}$, then the iterative distributed Algorithm 1 is ϵ-differentially private with the cumulative privacy budget for iterations from $k = 0$ to $k = T_0$ less than $\bar{\epsilon}$.*

Proof. The lemma can be obtained following the same line of reasoning of Lemma 2 in [34] (also see Theorem 3 in [19]). ■

Remark 10. *By replacing the ℓ_1 norm in the sensitivity Definition 3 with the ℓ_2 norm, we can also employ Gaussian noise to achieve a given differential-privacy protection (budget) ϵ^k for iteration k (see Appendix A of [18]). Then, as long as the added noises are independent across different iterations, we can leverage the sequential composition property of differential privacy (see Theorem 3.20 of [18]) to compute the cumulative privacy budget and obtain a result similar to Lemma 6 under Gaussian-noise based differential-privacy design.*

Theorem 2. *Under Assumptions 1, 2 3, if $\{\lambda^k\}$ and $\{\gamma^k\}$ satisfy the conditions in Theorem 1, and all elements of ζ_i^k are drawn independently from Laplace distribution $\mathrm{Lap}(\nu^k)$ with $(\sigma_i^k)^2 = 2(\nu^k)^2$ satisfying Assumption 4, then all players will converge almost surely to the Nash equilibrium. Moreover,*

1) *For any finite number of iterations T_0, Algorithm 1 is ϵ-differentially private with the cumulative privacy budget bounded by $\epsilon \le \sum_{k=1}^{T_0} \frac{C\varsigma^k}{\nu^k}$ where $\varsigma^k \triangleq \sum_{p=1}^{k-1}(\Pi_{q=p}^{k-1}(1-\bar{L}\gamma^q)) + 1$, $\bar{L} \triangleq \min_i\{|L_{ii}|\}$, and $C \triangleq \max_{i\in[m],0\le k\le T_0-1}\{\|x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k)\|_1\}$ (note that C is always finite since the algorithm ensures convergence in both $\mathcal{P}$ and $\mathcal{P}'$);*
2) *The cumulative privacy budget is always finite for $T_0 \to \infty$ when the sequence $\{\frac{\lambda^k}{\nu^k}\}$ is summable.*

Proof. Because the Laplace noise satisfies Assumption 4, it follows from Theorem 1 that the iterate x_i^k of every player i will converge to the Nash equilibrum x_i^* almost surely.

To prove the statements on differential privacy, we first analyze the sensitivity of Algorithm 1. Given two adjacent distributed games $\mathcal{P}$ and $\mathcal{P}'$, for any given fixed observation $\mathcal{O}$ and initial state $\left[(x^0)^T,\, (v^0)^T\right]^T$, the sensitivity is determined by $\|v^k - v'^k\|_1$ according to Definition 3. Since in $\mathcal{P}$ and $\mathcal{P}'$, there is only one payoff function that is different, we represent this different payoff function as the ith one, i.e., $f_i(\cdot)$, without loss of generality.

Because the initial conditions, payoff functions, and observations of $\mathcal{P}$ and $\mathcal{P}'$ are identical for $j \neq i$, we have $v_j^k = v_j'^k$ for all $j \neq i$ and k. Therefore, $\|v^k - v'^k\|_1$ is always equal to $\|v_i^k - v_i'^k\|_1$.

Since the sensitivity is independent of the magnitude of Laplace noises added for differential privacy, we calculate sensitivity in the noise-free case. Then, according to the update rule in (6), we have

$$\begin{aligned} v_i^{k+1} - v_i'^{k+1} =& (1 + L_{ii}\gamma^k)(v_i^k - v_i'^k) \\ &+ x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k). \end{aligned}$$

Note that we have used the definition $L_{ii} \triangleq -\sum_{j\in\mathbb{N}_i} L_{ij}$ and the fact that the observations v_j^k and $v_j'^k$ are the same.

Hence, the sensitivity Δ^k satisfies

$$\Delta^{k+1} \le (1 - |L_{ii}|\gamma^k)\Delta^k + \|x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k)\|_1.$$

which, implies the first statement by iteration using Lemma 6.

For the infinite horizon result in the second statement, we exploit the fact that our algorithm ensures convergence in both $\mathcal{P}$ and $\mathcal{P}'$. This means that $\|x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k)\|_1 = 0$ will be satisfied when k is large enough using the third condition in Definition 1. Furthermore, the ensured convergence also means that $\|x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k)\|_1$ is always bounded. Hence, there always exists some constant C such that the sequence $\{\|x_i^{k+1} - x_i^k - (x_i'^{k+1} - x_i'^k)\|_1\}$ is upper bounded by the sequence $\{C\gamma^k\lambda^k\}$.

Therefore, according to Lemma 4 in [21], there always exists a constant $\bar{C}$ such that $\Delta^k \le \bar{C}\lambda^k$ holds. Using Lemma 6, we can easily obtain $\epsilon \le \sum_{k=1}^{T_0} \frac{\bar{C}\lambda^k}{\nu^k}$. Hence, ϵ will be finite even when T_0 tends to infinity if the sequence $\{\frac{\lambda^k}{\nu^k}\}$ is summable, i.e., $\sum_{k=0}^{\infty} \frac{\lambda^k}{\nu^k} < \infty$. ■

Note that to ensure that the cumulative differential-privacy budget is finite (an unbounded privacy budget means complete loss of privacy protection), [19] and [34] have to use a summable stepsize (geometrically-decreasing stepsize, more specifically), which, however, also makes it impossible to ensure convergence to the exact desired equilibrium. In our approach, by allowing the stepsize sequence to be non-summable, we achieve both accurate convergence and finite cumulative privacy budget, even when the number of iterations goes to infinity. In fact, to our knowledge, this is the first time that almost-sure convergence to a Nash equilibrium is achieved under rigorous ϵ-differential privacy even with the number of iterations going to infinity.

Remark 11. *It is worth noting that to ensure the boundedness of the cumulative privacy budget $\epsilon = \sum_{k=1}^{\infty} \frac{\bar{C}\lambda^k}{\nu^k}$ when $k \to \infty$, our algorithm uses Laplace noise with parameter ν^k increasing with time (since we require $\frac{\lambda^k}{\nu^k}$ to be summable while $\{\lambda^k\}$ is non-summable). Because the strength of shared signal is always v_i^k, an increasing ν^k makes the relative level between noise ζ_i^k and signal v_i^k increase with time. However, since what actually feeds into the algorithm is $\gamma^k \text{Lap}(\nu^k)$, and the increase in the noise level ν^k is outweighed by the decrease of γ^k (see Assumption 4), the actual noise fed into the algorithm still decays with time, which makes it possible for Algorithm 1 to ensure every player's almost sure convergence to the Nash equilibrium. Moreover, according to Theorem 1, such almost sure convergence is not affected by scaling ν^k by any constant coefficient $\frac{1}{\epsilon} > 0$ so as to achieve any desired level of ϵ-differential privacy, as long as the Laplace noise parameter ν^k (with associated variance $(\sigma_i^k)^2 = 2(\nu^k)^2$) satisfies Assumption 4.*

It is worth noting that our algorithms' simultaneous achievement of both provable accuracy and ϵ-differential privacy does not contradict the fundamental trade-off between utility and privacy in the differential-privacy theory [18]. In fact, despite avoiding trading off convergence accuracy, our approach does pay a utility price in convergence speed. More specifically, in order to reduce ϵ to enhance privacy, we can use a faster-increasing $\{\nu^k\}$ according to Theorem 2, which requires $\{\gamma^k\}$ to decrease faster from Assumption 4. Given that according to Remark 9, the convergence speed is determined by $\mathcal{O}(\frac{\lambda^k}{\gamma^k})$, whose decreasing speed to zero reduces with an increase in the decreasing speed of γ^k, we arrive at the conclusion that a stronger privacy protection comes with a price of a slower convergence speed.

IV. Extension to Stochastic Aggregative Games

In this section, we prove that the proposed distributed algorithm can ensure the almost sure convergence of all agents to the Nash equilibrium even when individual agents only have access to a stochastic estimate of their payoff functions. Such stochastic Nash-equilibrium seeking problems arise frequently in practical applications like electricity markets [8], [28] and transportation systems [39] where the payoff functions are subject to stochastic uncertainties.

Representing the stochastic version of the payoff functions as $f_i(x_i, \bar{x}, \xi_i)$ for player i, where $\bar{x} \triangleq \frac{\sum_{i=1}^m x_i}{m}$, and $\xi_i \in \mathbb{R}^d$ is a random vector, we can formulate the stochastic game that player i faces as the following parameterized optimization problem:

$$\min \mathbb{E}\left[f_i(x_i, \bar{x}, \xi_i)\right] \quad \text{s.t.} \quad x_i \in K_i \text{ and } \bar{x} \in \bar{K}, \tag{43}$$

where the expected value is taken with respect to ξ_i. The constraint set K_i and the function $f_i(\cdot)$ are assumed to be known to player i only.

When the payoff functions are given through the expectation, the pseudo-gradients that individual players can access become stochastic, i.e., the gradient mapping $F(x, \bar{x})$ has components

$$F_i(x, \bar{x}) = \mathbb{E}\left[\nabla_{x_i} f_i(x_i, \bar{x}, \xi_i)\right], \quad \forall i \in [m].$$

In this case, in Algorithm 1, the mapping $F_i(x_i^k, v_i^k)$ is replaced with a sampled mapping

$$\tilde{F}_i(x_i^k, v_i^k, \xi_i^k) = \nabla_{x_i} f_i(x_i^k, v_i^k, \xi_i^k), \quad \forall i \in [m].$$

Accordingly, our privacy-preserving distributed algorithm reduces to:

Algorithm 2: Differentially-private distributed algorithm for stochastic aggregative games with guaranteed convergence

---

Parameters: Stepsize $\lambda^k > 0$ and weakening factor $\gamma^k > 0$.

Every player i maintains one decision variable x_i^k, which is initialized with a random vector in $K_i \subseteq \mathbb{R}^d$, and an estimate of the aggregate decision v_i^k, which is initialized as $v_i^0 = x_i^0$.

for $k = 1, 2, \dots$ **do**

a) Every player j adds persistent differential-privacy noise ζ_j^k to its estimate v_j^k, and then sends the obscured estimate $v_j^k + \zeta_j^k$ to agent $i \in \mathbb{N}_j$.

b) After receiving $v_j^k + \zeta_j^k$ from all $j \in \mathbb{N}_i$, player i updates its decision variable and estimate as follows:

$$x_i^{k+1} = \Pi_{K_i}\left[x_i^k - \lambda^k \nabla \tilde{F}_i(x_i^k, v_i^k, \xi_i^k)\right],$$
$$v_i^{k+1} = v_i^k + \gamma^k \sum_{j \in \mathbb{N}_i} L_{ij}(v_j^k + \zeta_j^k - v_i^k - \zeta_i^k) + x_i^{k+1} - x_i^k. \tag{44}$$

c) **end**

A. Convergence Analysis

Next, we prove that Algorithm 2 can ensure the convergence of the decision vector $x^k \triangleq [(x_1^k)^T, \cdots, (x_m^k)^T]^T$ to the exact Nash equilibrium point $x^* \triangleq [(x_1^*)^T, \cdots, (x_m^*)^T]^T$, even in the presence of differential-privacy noise ζ_i^k and stochastic pseudo-gradient $\tilde{F}_i(x_i, v_i^k, \xi_i^k)$. To this end, similar to [28], we first formalize the noise in pseudo-gradients:

Assumption 5. *Let $\mathcal{F}^k \triangleq \{\xi^0, \cdots, \xi^k\}$ be the family of sigma algebra with $\xi^k = [(\xi_1^k)^T, \cdots, (\xi_m^k)^T]^T$, we have the following relationship almost surely:*

$$\mathbb{E}\left[\tilde{F}_i(x_i^k, v_i^k, \xi_i^k) | \mathcal{F}^k\right] = F_i(x_i^k, v_i^k), \tag{45}$$

$$\mathbb{E}\left[\left\|\tilde{F}_i(x_i^k, v_i^k, \xi_i^k) - F_i(x_i^k, v_i^k)\right\|^2 | \mathcal{F}^k\right] \leq (\mu^k)^2, \tag{46}$$

where μ^k is some positive scalar.

Theorem 3. *Under Assumptions 1, 2, 3, 4, 5, if there exists some $T \geq 0$ such that for all $k \geq T$, γ^k and λ^k satisfy the following conditions:*

$$\sum_{k=T}^{\infty} \gamma^k = \infty, \sum_{k=T}^{\infty} \lambda^k = \infty, \sum_{k=T}^{\infty} (\gamma^k)^2 < \infty, \sum_{k=T}^{\infty} \frac{(\lambda^k)^2}{\gamma^k} < \infty,$$

and $\sum_{k=T}^{\infty}(\lambda^k\mu^k)^2 < \infty$, then Algorithm 2 converges to the Nash equilibrium of the game in (43) almost surely.

Proof. Similar to the proof of Theorem 1, the basic idea is still to apply Proposition 1 to the quantities $\sum_{i=1}^m \|x_i^{k+1}-x_i^*\|^2$ and $\sum_{i=1}^m \|v_i^{k+1}-\bar{v}^{k+1}\|^2$. Since the stochasticity in $\tilde{F}_i(x_i,\bar{x},\xi_i^k)$ does not affect the dynamics of v_i^k, the relationship for $\sum_{i=1}^m \|v_i^{k+1}-\bar{v}^{k+1}\|^2$ in Algorithm 1 still holds under Algorithm 2, i.e., we still have

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^m \|v_i^{k+1}-\bar{v}^{k+1}\|^2\mathcal{F}^k\right]\\&\leq (1-\gamma^k|\rho_2|)\sum_{i=1}^m\|v_i^k-\bar{v}^k\|^2+\frac{2(\lambda^k)^2\tilde{L}^2}{\gamma^k|\rho_2|}\sum_{i=1}^m\|v_i^k-\bar{v}^k\|^2\\&\quad+\frac{2m(\lambda^k)^2C^2}{\gamma^k|\rho_2|}+(\gamma^k)^2\|L\|^2\sum_{i=1}^m(\sigma_i^k)^2\end{aligned} \tag{47}$$

for $\mathcal{F}^k=\{x^0,\, v^0,\cdots,x^k,\, v^k\}$.

Therefore, we only characterize $\sum_{i=1}^m \|x_i^{k+1}-x_i^*\|^2$, whose evolution is affected by the replacement of $F_i(x_i^k,v_i^k)$ with $\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)$.

Using the relation $x_i^*=\Pi_{K_i}\left[x_i^*-\lambda^k F_i(x_i^*,\bar{x}^*)\right]$ with $\bar{x}^*=\frac{1}{m}\sum_{i=1}^m x_i^*$, from (44), we can arrive at

$$\begin{aligned}&\left\|x_i^{k+1}-x_i^*\right\|^2\\&=\left\|\Pi_{K_i}\left[x_i^k-\lambda^k\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)\right]-x_i^*\right\|^2\\&=\left\|\Pi_{K_i}\left[x_i^k-\lambda^k\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)\right]-\Pi_{K_i}\left[x_i^*-\lambda^kF_i(x_i^*,\bar{x}^*)\right]\right\|^2\\&\leq\left\|x_i^k-x_i^*-\lambda^k(\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^*,\bar{x}^*))\right\|^2\\&\leq\left\|x_i^k-x_i^*\right\|^2+(\lambda^k)^2\left\|\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2\\&\quad-2\left\langle x_i^k-x_i^*,\,\lambda^k(\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^*,\bar{x}^*))\right\rangle.\end{aligned} \tag{48}$$

For the second term on the right hand side of the above inequality, we can bound it by adding and subtracting $F_i(x_i^k,v_i^k)$:

$$\begin{aligned}&\left\|\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2\\&=\left\|\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^k,v_i^k)+F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2\\&\leq 2\left\|\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^k,v_i^k)\right\|^2\\&\quad+2\left\|F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2.\end{aligned} \tag{49}$$

Plugging (49) into (48) yields

$$\begin{aligned}&\left\|x_i^{k+1}-x_i^*\right\|^2\\&\leq\left\|x_i^k-x_i^*\right\|^2+2(\lambda^k)^2\left\|\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^k,v_i^k)\right\|^2\\&\quad+2(\lambda^k)^2\left\|F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2\\&\quad-2\left\langle x_i^k-x_i^*,\,\lambda^k(\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)-F_i(x_i^*,\bar{x}^*))\right\rangle.\end{aligned} \tag{50}$$

Taking the conditional expectation, given $\mathcal{F}^k=\{v^0,\,x^0,\,\ldots,v^k,\,x^k\}$, from the preceding relation we obtain for all $k\geq 0$:

$$\begin{aligned}&\mathbb{E}\left[\left\|x_i^{k+1}-x_i^*\right\|^2|\mathcal{F}^k\right]\leq\left\|x_i^k-x_i^*\right\|^2+2(\lambda^k\mu^k)^2\\&\quad+2(\lambda^k)^2\left\|F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2\\&\quad-2\left\langle x_i^k-x_i^*,\,\lambda^k(F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*))\right\rangle,\end{aligned} \tag{51}$$

where we used the assumption that $\tilde{F}_i(x_i^k,v_i^k,\xi_i^k)$ is an unbiased estimate of $F_i(x_i^k,v_i^k)$ with variance $(\mu^k)^2$ (see Assumption 5).

By adding and subtracting $F_i(x_i^k,\,\bar{v}^k)$ to the inner-product term, we arrive at

$$\begin{aligned}&\mathbb{E}\left[\left\|x_i^{k+1}-x_i^*\right\|^2|\mathcal{F}^k\right]\leq\left\|x_i^k-x_i^*\right\|^2+2(\lambda^k\mu^k)^2\\&\quad+\underbrace{2(\lambda^k)^2\left\|F_i(x_i^k,v_i^k)-F_i(x_i^*,\bar{x}^*)\right\|^2}_{\text{Term 1}}\\&\quad-\underbrace{2\left\langle x_i^k-x_i^*,\,\lambda^k(F_i(x_i^k,v_i^k)-F_i(x_i^k,\bar{v}^k))\right\rangle}_{\text{Term 2}}\\&\quad-\underbrace{2\left\langle x_i^k-x_i^*,\,\lambda^k(F_i(x_i^k,\bar{v}^k)-F_i(x_i^*,\bar{x}^*))\right\rangle}_{\text{Term 3}}.\end{aligned} \tag{52}$$

The three terms on the right hand side of (52) can be bounded in a similar way to Theorem 1:

$$\text{Term 1}\leq 12(\lambda^k)^2C^2+8(\lambda^k)^2\tilde{L}^2\|v_i^k-\bar{v}^k\|^2, \tag{53}$$

$$\text{Term 2}\geq -\frac{(\lambda^k)^2\|x_i^k-x_i^*\|^2}{\gamma^k},-\gamma^k\tilde{L}^2\left\|v_i^k-\bar{v}^k\right\|^2, \tag{54}$$

$$\text{Term 3}=2\lambda^k\left(F_i(x_i^k,\bar{x}^k)-F_i(x_i^*,\bar{x}^*)\right)^T(x_i^k-x_i^*). \tag{55}$$

Plugging (53), (54), and (55) into (52) yields

$$\begin{aligned}&\mathbb{E}\left[\left\|x_i^{k+1}-x_i^*\right\|^2|\mathcal{F}^k\right]\leq\left\|x_i^k-x_i^*\right\|^2+2(\lambda^k\mu^k)^2\\&+12(\lambda^k)^2C^2+8(\lambda^k)^2\tilde{L}^2\|v_i^k-\bar{v}^k\|^2+\frac{(\lambda^k)^2\|x_i^k-x_i^*\|^2}{\gamma^k}\\&+\gamma^k\tilde{L}^2\left\|v_i^k-\bar{v}^k\right\|^2-2\lambda^k\left(F_i(x_i^k,\bar{x}^k)-F_i(x_i^*,\bar{x}^*)\right)^T(x_i^k-x_i^*).\end{aligned} \tag{56}$$

Summing (39) from $i=1$ to $i=m$ yields

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^m\left\|x_i^{k+1}-x_i^*\right\|^2|\mathcal{F}^k\right]\\&\leq\sum_{i=1}^m\left\|x_i^k-x_i^*\right\|^2+2m(\lambda^k\mu^k)^2+12m(\lambda^k)^2C^2\\&+8(\lambda^k)^2\tilde{L}^2\sum_{i=1}^m\|v_i^k-\bar{v}^k\|^2+\frac{(\lambda^k)^2\sum_{i=1}^m\|x_i^k-x_i^*\|^2}{\gamma^k}\\&+\gamma^k\tilde{L}^2\sum_{i=1}^m\left\|v_i^k-\bar{v}^k\right\|^2-2\lambda^k\left(\phi(x^k)-\phi(x^*)\right)^T(x^k-x^*).\end{aligned} \tag{57}$$

Similar to the derivation in Theorem 1, we have the following relations from (47) and (57) for $\mathbf{v}^k=\left[\sum_{i=1}^m\|x_i^k-x_i^*\|^2,\;\sum_{i=1}^m\|v_i^k-\bar{v}^k\|^2\right]^T$:

$$\mathbb{E}\left[\mathbf{v}^{k+1}|\mathcal{F}^k\right]\leq(V^k+A^k)\mathbf{v}^k-2\lambda^k\Phi^k+B^k, \tag{58}$$

where

$$V^k = \begin{bmatrix} 1 & \tilde{L}^2\gamma^k \\ 0 & 1-\gamma^k|\rho_2| \end{bmatrix},$$

$$A^k = \begin{bmatrix} \frac{(\lambda^k)^2}{\gamma^k} & 8(\lambda^k)^2\tilde{L}^2 \\ 0 & \frac{2(\lambda^k)^2\tilde{L}^2}{\gamma^k|\rho_2|} \end{bmatrix},$$

$$\Phi^k = \begin{bmatrix} \left(\phi(x^k)-\phi(x^*)\right)^T(x^k-x^*) \\ 0 \end{bmatrix},$$

$$B^k = \begin{bmatrix} 2m(\lambda^k\mu^k)^2 + 12m(\lambda^k)^2C^2 \\ \frac{2m(\lambda^k)^2C^2}{\gamma^k|\rho_2|} + (\gamma^k)^2\|L\|^2\sum_{i=1}^m(\sigma_i^k)^2 \end{bmatrix}.$$

Using Assumption 4 and the conditions of the theorem $\sum_{k=T}^{\infty}(\gamma^k)^2 < \infty$, $\sum_{k=T}^{\infty}\frac{(\lambda^k)^2}{\gamma^k} < \infty$, and $\sum_{k=T}^{\infty}(\lambda^k\mu^k)^2 < \infty$, we have that all elements of the matrices of A^k and B^k are summable. Therefore, we have $\sum_{i=1}^m\|x_i^k - x_i^*\|^2$ and $\sum_{i=1}^m\|v_i^k - \bar{v}^k\|^2$ satisfying the conditions of Proposition 1 with $\kappa_1 = \tilde{L}^2$, $\kappa_2 = |\rho_2|$, $c^k = 2\lambda^k$, $a^k = \max\{\frac{(\lambda^k)^2}{\gamma^k}, 8(\lambda^k)^2\tilde{L}^2, \frac{2(\lambda^k)^2\tilde{L}^2}{\gamma^k|\rho_2|}\}$, and $b^k = \max\{2m(\lambda^k\mu^k)^2 + 12m(\lambda^k)^2C^2, \frac{2m(\lambda^k)^2C^2}{\gamma^k|\rho_2|} + (\gamma^k)^2\|L\|^2\sum_{i=1}^m(\sigma_i^k)^2\}$. ■

Remark 12. *Note that different from [20], [27] which deal with stochastic pseudo-gradients with decreasing variances (by increasing sample sizes), our Algorithm 2 allows the variance $(\mu^k)^2$ to be constant and even increasing with time. For example, when λ^k is set as $\frac{c_1}{1+c_2k}$, the condition $\sum_{k=T}^{\infty}(\lambda^k\mu^k)^2 < \infty$ in Theorem 3 can be satisfied for $\mu^k = c_3 + c_4k^\nu$ with any $0<\nu<0.5$ and positive constants c_1, c_2, c_3, and c_4.*

Remark 13. *We can obtain that the convergence of $\sum_{i=1}^m\|x_i^{k+1} - x_i^*\|^2$ follows*

$$\mathbb{E}\left[\sum_{i=1}^m\|x_i^{k+1}-x_i^*\|^2|\mathcal{F}^k\right] \le (1+a^k)\sum_{i=1}^m\|\bar{x}_i^k - x_i^*\|^2 + \hat{b}^k, \tag{59}$$

where all parameters are from Proposition 1 and $\hat{b}^k = \left(\kappa_1\gamma^k + a^k\right)\sum_{i=1}^m\|v_i^k - \bar{v}^k\|^2 + b^k$. Given that the stochasticty in $\tilde{F}_i(x_i^k, v_i^k, \xi_i^k)$ only increases the value of b^k but does not affect its order (still summable), we can use an argument similar to Remark 9 to obtain that the convergence of all players to the Nash equilibrium is still no slower than the order of $\mathcal{O}(\frac{\lambda^k}{\gamma^k})$. Moreover, since the evolution of v_i^k is not affected by the stochasticity in $\tilde{F}_i(x_i^k, v_i^k, \xi_i^k)$ and still follows (33), we have that the decreasing speed of $\sum_{i=1}^m\|v_i^k - \bar{v}^k\|^2$ still increases with an increase in $|\rho_2|$, which corresponds to the second largest eigenvalue of L. Therefore, the decreasing speed of $\sum_{i=1}^m\|v_i^k - \bar{v}^k\|^2$ to zero increases with an increase in the absolute value of the second largest eigenvalue of L in Assumption 3.

B. Privacy Analysis for Algorithm 2

Similar to the privacy analysis in Sec. III-B, we can also analyze the strength of differential privacy for Algorithm 2:

Theorem 4. *Under Assumptions 1, 2, 3, and 5, if $\{\lambda^k\}$, $\{\gamma^k\}$, and $\{\mu^k\}$ satisfy the conditions in Theorem 3, and all elements of ζ_i^k are drawn independently from Laplace distribution $\text{Lap}(\nu^k)$ with $(\sigma_i^k)^2 = 2(\nu^k)^2$ satisfying Assumption 4, then all players will converge almost surely to the Nash equilibrium. Moreover,*

1) *For any finite number of iterations T_0, Algorithm 2 is ϵ-differentially private with the cumulative privacy budget bounded by $\epsilon \le \sum_{k=1}^{T_0}\frac{C\varsigma^k}{\nu^k}$ where $\varsigma^k \triangleq \sum_{p=1}^{k-1}(\Pi_{q=p}^{k-1}(1-\bar{L}\gamma^q)) + 1$, $\bar{L} \triangleq \min_i\{|L_{ii}|\}$, and $C \triangleq \max_{i\in[m],0\le k\le T_0-1}\{\|x_i^{k+1} - x_i^k - (x'^{k+1}_i - x'^k_i)\|_1\}$ (note that C is always finite since the algorithm ensures convergence in both $\mathcal{P}$ and $\mathcal{P}'$);*
2) *The cumulative privacy budget is always finite for $T_0 \to \infty$ when the sequence $\{\frac{\lambda^k}{\nu^k}\}$ is summable.*

Proof. The derivation follows the proof of Theorem 2, and hence is omitted here. ■

V. Numerical Simulations

In this section, we evaluate the performance of the proposed differentially-private distributed Nash-equilibrium seeking algorithms using a networked Nash-Cournot game. More specifically, we consider m firms producing a homogeneous commodity competing over N markets, which has been considered recently in [5], [7], [9]. Fig. 1 presents a schematic of the problem involving $N = 7$ markets (represented by $M_1, \cdots, M_7$) and $m = 20$ firms (represented by circles). In the figure, an edge from circle i to M_j means that firm i participates in market M_j.

We consider the setting where a firm can only see partial decision information of the network. Namely, every firm can only communicate with its immediate neighbors and no central coordinator exists which can communicate with all firms. As in [5], [7], we allow firms to communicate with their immediate neighbors to share their production decisions. In the considered scenario, we use $x_i \in \mathbb{R}^N$ to represent the amount of firm i' products. Note that a firm i is allowed to participate in $1 \le n_i \le N$ markets, and if firm i does not participate in market j, then the jth entry of x_i will be forced to be 0 all the time. So a firm participating in $1 \le n_i \le N$ markets will have n_i nonzero entries in the production vector x_i. For the convenience of bookkeeping, we use an adjacency matrix $B_i \in \mathbb{R}^{N\times N}$ to describe the association relationship between firm i and all the markets. More specifically, B_i has zero off-diagonal elements and its $(j,j)th$ entry is 1 when firm i participates in market j, otherwise, its $(j,j)th$ entry is zero. Every firm i has a maximal capacity for each market j it participates in, which is represented by C_{ij}. Denoting $C_i \triangleq [C_{i1}, \cdots, C_{iN}]^T$, we always have $x_i \le C_i$. Represent B as $B \triangleq [B_1, \cdots, B_N]$. It can be seen that $Bx \in \mathbb{R}^N = \sum_{i=1}^N B_ix_i$ represents the total product supply to all markets, given firm i's production amount x_i. As in [7], the commodity's price in every market M_i follows a linear inverse demand function, i.e., it is a linear function of the total amount of commodity supplied to the market: $p_i(x) = \bar{P}_i - \chi_i[Bx]_i$, where $\bar{P}_i$ and $\chi_i > 0$ are constants and $[Bx]_i$ denotes the ith element of the vector Bx.

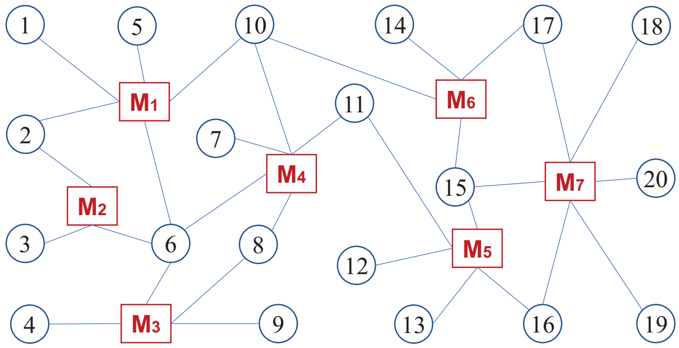

Fig. 1. Nash-Cournot game of 20 players (firms) competing over 7 locations (markets). Each firm is represented by a circular and each market is represented by a square. An edge between firm i $(1 \leq i \leq 20)$ and market j $(1 \leq j \leq 7)$ means that firm i participates in market j.

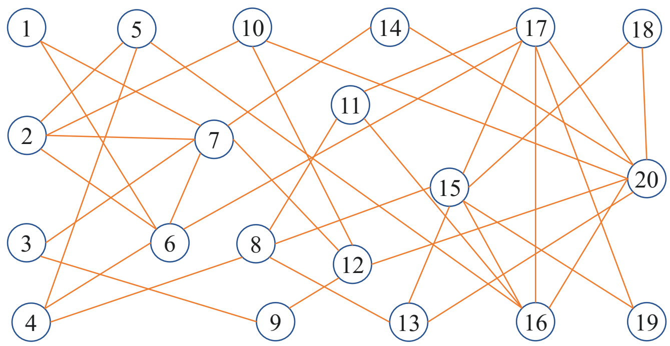

Fig. 2. The randomly generated interaction patten of the 20 firms.

It can be seen that the price decreases with an increase in the amount of supplied commodity.

We let $p \triangleq [p_1, \cdots, p_N]^T$ represent the price vector of all markets, which can be verified to satisfy $p = \bar{P} - \Xi B x$, where $\bar{P} \triangleq [\bar{P}_1, \cdots, \bar{P}_N]^T$ and $\Xi \triangleq \text{diag}(\chi_1, \cdots, \chi_N)$. The total payoff of firm i can then be expressed as $p^T B_i x_i$. Firm i's production cost is assumed to be a strongly convex, quadratic function $c_i(x_i) = x_i^T Q_i x_i + q_i^T x_i$, where $Q_i \in \mathbb{R}^{N \times N}$ is a positive definite matrix and $q_i \in \mathbb{R}^N$.

Therefore, firm i's local objective function, which is determined by its production cost c_i and payoff, is given by $f_i(x_i, x) = c_i(x_i) - (\bar{P} - \Xi B x)^T B_i^T x_i$. And it can be verified that the gradient the objective function is $F_i(x_i, x) = 2Q_i x_i + q_i + B_i^T \Xi B_i x_i - B_i(\bar{P} - \Xi B x)$. It is clear that both firm i's local objective function and gradient are dependent on other firms' actions.

In the implementation, we consider $N = 7$ markets and 20 firms. Since no firm can communicate with all the other firms, we generate local communication patterns randomly, with the interaction graph given in Fig. 2. The maximal capacities for firm i (elements in C_i) are randomly selected from the interval $[8, 10]$. Q_i in the production cost function is set as νI with ν randomly selected from $[1, 10]$. q_i in $c_i(x_i)$ is randomly selected from a uniform distribution in $[1, 2]$. In the price function, $\bar{P}_i$ and χ_i are randomly chosen from uniform distributions in $[10, 20]$ and $[1, 3]$, respectively.

To evaluate the performance of the proposed Algorithm 1, for every firm i, we inject differential-privacy noise ζ_i^k in every message it shares in all iterations. Each element of the noise vector follows Laplace distribution with parameter $\nu^k = 1 + 0.1k^{0.2}$. We set the stepsize λ^k and diminishing sequence γ^k as $\lambda^k = \frac{0.1}{1+0.1k}$ and $\gamma^k = \frac{1}{1+0.1k^{0.9}}$, respectively, which satisfy the conditions in Theorem 1 and Theorem 2. In the evaluation, we run our algorithm for 100 times and calculate the average as well as the variance of the gap $\|x^k - x^*\|$ between generated iterate x^k and the Nash equilibrium x^* as a function of the iteration index k. The result is given by the red curve and error bars in Fig. 3. For comparison, we also run the existing distributed Nash-equilibrium seeking algorithm proposed by Koshal et al. in [5] under the same noise, and the existing differential-privacy approach for networked aggregative games proposed by Ye et al. in [19] under the same cumulative privacy budget ϵ. Note that the differential-privacy approach in [19] uses geometrically decreasing stepsizes (to be summable) to ensure a finite privacy budget, but the fast decreasing stepsize also leads to the loss of guaranteed convergence to the exact Nash equilibrium. The evolution of the average error/variance of the approaches in [5] and [19] are given by the blue and black curves/error bars in Fig. 3, respectively. It is clear that the proposed algorithm has a comparable convergence speed but much better accuracy.

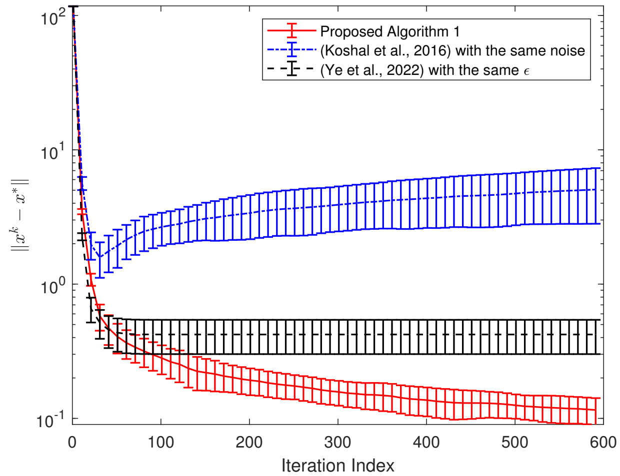

Fig. 3. Comparison of Algorithm 1 with the existing distributed Nash-equilibrium seeking algorithm by Koshal et al. in [5] (under the same noise) and the existing differential-privacy approach for distributed aggregative games by Ye et al. in [19] (under the same privacy budget ϵ).

Based a similar setup, we also test the proposed Algorithm 2 when individual players only have access to a stochastic version of the payoff functions and pseudo-gradients. More specifically, we add Gaussian noise of zero mean and unit variance in every dimension of the pseudo-gradient vector $F_i(x_i^k, v_i^k)$. The differential-privacy noise still follows Laplace distribution with parameter $\nu^k = 1 + 0.1k^{0.2}$. The stepsize λ^k and diminishing sequence γ^k are still set as $\lambda^k = \frac{0.1}{1+0.1k}$ and $\gamma^k = \frac{1}{1+0.1k^{0.9}}$, respectively, which satisfy the conditions in Theorem 3 and Theorem 4. The result is given by the red curve and error bars in Fig. 4. For comparison, we also run

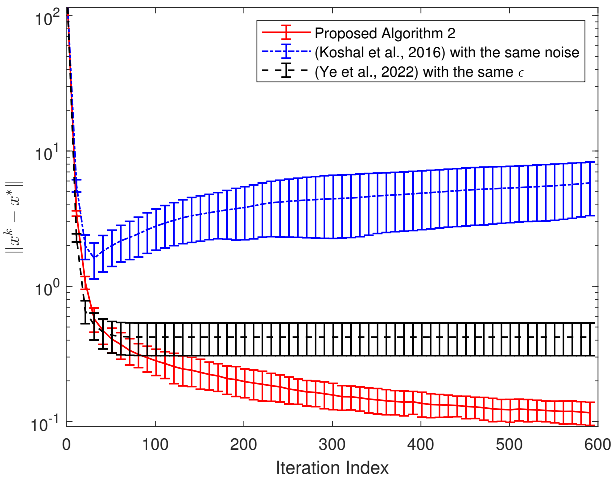

Fig. 4. Comparison of Algorithm 2 with the existing stochastic distributed Nash-equilibrium seeking algorithm by Koshal et al. in [5] (under the same noise) and the existing differential-privacy approach for distributed aggregative games by Ye et al. in [19] (under the same privacy budget ϵ).

the existing distributed Nash-equilibrium seeking algorithm proposed by Koshal et al. in [5] under the same noise, and the existing differential-privacy approach for networked aggregative games proposed by Ye et al. in [19] under the same cumulative privacy budget ϵ. The evolution of the average error/variance of the approaches in [5] and [19] are given by the blue and black curves/error bars in Fig. 4, respectively. It is clear that the proposed algorithm has a comparable convergence speed but much better accuracy.

VI. Conclusions

Although differential privacy is becoming the de facto standard for publicly sharing information, its direct incorporation into coordinator-free fully distributed aggregative games leads to errors in equilibrium computation due to the need to iteratively and repeatedly inject independent noises. This paper proposes a fully distributed Nash-equilibrium seeking approach for networked aggregative games that ensures both accurate convergence to the exact Nash equilibrium and rigorous ϵ-differential privacy with bounded cumulative privacy budget, even when the number of iterations goes to infinity. The simultaneous achievement of both goals is a sharp contrast to existing differential-privacy solutions for aggregative games that have to trade convergence accuracy for privacy, and to our knowledge, has not been achieved before. The approach can also be extended to stochastic aggregative games and is proven able to ensure both accurate convergence to the Nash equilibrium and rigorous differential privacy, even when every player's stochastic estimate of the pseudo-gradient is subject to a constant or even increasing variance. Numerical simulation results confirm that the proposed algorithms have a better accuracy compared with existing approaches.

References

[1] W. Saad, Z. Han, H. V. Poor, and T. Başar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 86–105, 2012.
[2] H. Yin, U. V. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with scaled congestion costs and shared constraints," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1702–1708, 2011.
[3] K. Okuguchi and F. Szidarovszky, *The Theory of Oligopoly with Multi-product Firms*. Springer Science & Business Media, 2012.
[4] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Nash and wardrop equilibria in aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1373–1388, 2018.
[5] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, no. 3, pp. 680–704, 2016.
[6] F. Salehisadaghiani and L. Pavel, "Distributed nash equilibrium seeking: A gossip-based algorithm," *Automatica*, vol. 72, pp. 209–216, 2016.
[7] L. Pavel, "Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, pp. 1584–1597, 2020.
[8] B. Franci and S. Grammatico, "Stochastic generalized nash equilibrium seeking in merely monotone games," *IEEE Transactions on Automatic Control*, 2021.
[9] D. T. A. Nguyen, D. T. Nguyen, and A. Nedić, "Distributed nash equilibrium seeking over time-varying directed communication networks," *arXiv preprint arXiv:2201.02323*, 2022.
[10] S. Rassenti, S. S. Reynolds, V. L. Smith, and F. Szidarovszky, "Adaptation and convergence of behavior in repeated experimental cournot games," *Journal of Economic Behavior & Organization*, vol. 41, no. 2, pp. 117–146, 2000.
[11] R. Dong, W. Krichene, A. M. Bayen, and S. S. Sastry, "Differential privacy of populations in routing games," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 2798–2803.
[12] T. A. Gerhart and A. Steinberg, "Proposition 24: Protecting California consumers by expanding protections, ensuring governmental oversight, and safeguarding the law from special interests," *California Initiative Review (CIR)*, vol. 2020, no. 1, p. 12, 2020.
[13] Y. Lu and M. Zhu, "Game-theoretic distributed control with information-theoretic security guarantees," *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 264–269, 2015.
[14] R. Cummings, M. Kearns, A. Roth, and Z. S. Wu, "Privacy and truthful equilibrium selection for aggregative games," in *International Conference on Web and Internet Economics*. Springer, 2015, pp. 286–299.
[15] I. Shilov, H. Le Cadre, and A. Busic, "Privacy impact on generalized nash equilibrium in peer-to-peer electricity market," *Operations Research Letters*, vol. 49, no. 5, pp. 759–766, 2021.
[16] S. Gade, A. Winnicki, and S. Bose, "On privatizing equilibrium computation in aggregate games over networks," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 3272–3277, 2020.
[17] M. Shakarami, C. De Persis, and N. Monshizadeh, "Distributed dynamics for aggregative games: Robustness and privacy guarantees," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 9, pp. 5048–5069, 2022.
[18] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.
[19] M. Ye, G. Hu, L. Xie, and S. Xu, "Differentially private distributed nash equilibrium seeking for aggregative games," *IEEE Transactions on Automatic Control*, vol. 67, no. 5, pp. 2451–2458, 2021.
[20] J. Wang, J.-F. Zhang, and X. He, "Differentially private distributed algorithms for stochastic aggregative games," *Automatica*, vol. 142, p. 110440, 2022.
[21] Y. Wang and A. Nedić, "Tailoring gradient methods for differentially-private distributed optimization," *IEEE Transactions on Automatic Control*, 2022.
[22] Y. Wang and T. Başar, "Quantization enabled privacy protection in decentralized stochastic optimization," *IEEE Transactions on Automatic Control*, vol. 68, no. 7, pp. 4038–4052, 2023.
[23] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2061–2075, 2021.
[24] D. Gadjov and L. Pavel, "Single-timescale distributed gne seeking for aggregative games over networks via forward–backward operator splitting," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 3259–3266, 2020.

[25] R. Zhu, J. Zhang, K. You, and T. Başar, "Asynchronous networked aggregative games," *Automatica*, vol. 136, p. 110054, 2022.
[26] S. Liang, P. Yi, and Y. Hong, "Distributed nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.
[27] J. Lei and U. V. Shanbhag, "Distributed variable sample-size gradient-response and best-response schemes for stochastic nash equilibrium problems," *SIAM Journal on Optimization*, vol. 32, no. 2, pp. 573–603, 2022.
[28] F. Yousefian, A. Nedić, and U. V. Shanbhag, "Self-tuned stochastic approximation schemes for non-lipschitzian stochastic multi-user optimization and nash games," *IEEE Transactions on Automatic Control*, vol. 61, no. 7, pp. 1753–1766, 2015.
[29] T. Tatarenko, W. Shi, and A. Nedić, "Geometric convergence of gradient play algorithms for distributed nash equilibrium seeking," *IEEE Transactions on Automatic Control*, vol. 66, no. 11, pp. 5342–5353, 2020.
[30] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.
[31] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1835–1854, 2016.
[32] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
[33] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proceedings of the forty-second ACM Symposium on Theory of Computing*, 2010, pp. 715–724.
[34] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proceedings of the 2015 International Conference on Distributed Computing and Networking*, New York, NY, USA, 2015.
[35] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," in *International Conference on Machine Learning*. PMLR, 2015, pp. 1376–1385.
[36] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Theory of Cryptography Conference*. Springer, 2016, pp. 635–658.
[37] I. Mironov, "Rényi differential privacy," in *The 30th Computer Security Foundations Symposium*. IEEE, 2017, pp. 263–275.
[38] Z. Chen and Y. Wang, "Locally differentially private distributed online learning with guaranteed optimality," *arXiv preprint arXiv:2306.14094*, 2023.
[39] D. Watling, "User equilibrium traffic network assignment with stochastic travel times and late arrival penalty," *European Journal of Operational Research*, vol. 175, no. 3, pp. 1539–1556, 2006.

Yongqiang Wang was born in Shandong, China. He received dual B.S. degrees in electrical engineering & automation and computer science & technology from Xi'an Jiaotong University, Xi'an, Shaanxi, China, in 2004, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2009. From 2007-2008, he was with the University of Duisburg-Essen, Germany, as a visiting student. He was a Project Scientist at the University of California, Santa Barbara before joining Clemson University, SC, USA, where he is currently an Associate Professor. His current research interests include distributed control, optimization, and learning, with an emphasis on privacy protection. He currently serves as an associate editor for *IEEE Transactions on Automatic Control* and *IEEE Transactions on Control of Network Systems*.

Angelia Nedić holds a Ph.D. from Moscow State University, Moscow, Russia, in Computational Mathematics and Mathematical Physics (1994), and a Ph.D. from Massachusetts Institute of Technology, Cambridge, USA in Electrical and Computer Science Engineering (2002). She has worked as a senior engineer in BAE Systems North America, Advanced Information Technology Division at Burlington, MA. She is a recipient (jointly with her co-authors) of the Best Paper Award at the Winter Simulation Conference 2013 and the Best Paper Award at the International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt) 2015 (with co-authors). Her current interest is in large-scale optimization, games, control and information processing in networks.