

Ensure Differential Privacy and Convergence Accuracy in Consensus Tracking and Aggregative Games with Coupling Constraints

Yongqiang Wang

Abstract—We address differential privacy for fully distributed aggregative games with shared coupling constraints. By co-designing the generalized Nash equilibrium (GNE) seeking mechanism and the differential-privacy noise injection mechanism, we propose the first GNE seeking algorithm that can ensure both provable convergence to the GNE and rigorous ϵ -differential privacy, even with the number of iterations tending to infinity. As a basis of the co-design, we propose a new consensus-tracking algorithm that can achieve rigorous ϵ -differential privacy while maintaining accurate tracking performance, which, to our knowledge, has not been achieved before. To facilitate the convergence analysis, we also establish a general convergence result for stochastically-perturbed nonstationary fixed-point iteration processes, which lie at the core of numerous optimization and variational problems. Numerical simulation results confirm the effectiveness of the proposed approach.

I. INTRODUCTION

In recent years, Nash equilibrium seeking in aggregative games under shared coupling constraints is gaining increased traction [1], [2], [3]. Compared with the standard Nash equilibrium seeking problem where the feasible decision sets of individual players (also called agents) are independent of each other, the incorporation of shared coupling constraints in Nash equilibrium seeking, usually called the generalized Nash equilibrium (GNE) model, captures important characteristics of noncooperative games under limited network resources. To date, the GNE seeking problem has found vast applications in economics and various engineering domains, with typical examples including power grids [4], [5], [6], optical networks [7], communication networks [8], [9], [10], plug-in electric cars [11], [12], and mobile sensor networks [13], [14], [15]. In many of these applications, due to practical constraints, no central coordinator/mediator exists to collect and disperse information, and individual players can only access or observe the decisions of their immediate neighbors. Compared with the traditional case where every player can access all its competitors' decisions to precisely evaluate its cost function, which is called the *full-decision information* setting, the case where each player can only access its neighbors' decisions is called the *partial-decision information* setting, and is more challenging in that individual players lack information to compute their cost functions or gradients.

To account for the lack of information in the partial-decision information setting, players have to exchange information among local neighbors to estimate global information necessary for GNE seeking. Since the seminal work in [16], [17],

significant inroads have been made in games with partial-decision information, both for Nash equilibrium seeking without coupling constraints (see, e.g., [18], [19], [20], [21]) and GNE seeking (see, e.g., [1], [3], [22]). A key technique underlying these fully distributed algorithms is consensus tracking [23] (also called dynamic average consensus), which allows a group of agents to locally track the average of their reference inputs while each agent only has access to its own individual reference input. However, the employment of consensus tracking in GNE seeking also entails explicit sharing of (estimated) decisions in every iteration, which is problematic when involved information is sensitive [24], [25], [26]. Given the noncooperative relationship between players, individual players' privacy should be protected to avoid exploitation by others. For example, in Nash-Cournot games, individual players' cost functions could be market sensitive, and should not be disclosed to competitors [27]. In fact, in many scenarios, privacy preservation is mandated by legislation. For example, in road routing games [28], California Privacy Rights Act forbids disclosing drivers' spatiotemporal information, which, otherwise, can reveal a person's activities [29]. Moreover, privacy preservation is also a crucial step to encourage participation in cooperative policies [30].

To address the privacy issue in games, several approaches have been proposed for the full-decision information setting where a coordinator exists (see, e.g., [31], [32], [33]). Recently, results have also emerged to protect privacy in the partial-decision information setting (see, e.g., [30], [34]). However, these approaches are restricted in that they either require the communication graph to satisfy certain properties, or can only protect the cost function value from being uniquely identifiable. As differential privacy (DP) has become the de facto standard for privacy protection due to its strong resilience against arbitrary post-processing and auxiliary information [35], recently [36] and [37] achieve DP in aggregative games at the cost of losing convergence to the exact Nash equilibrium. By co-designing the Nash-equilibrium seeking mechanism and the DP-noise injection mechanism, our prior results [38], [39] successfully achieve both ϵ -DP and accurate convergence in coupling-constraint free Nash-equilibrium seeking. However, these results are inapplicable in GNE seeking, where the shared coupling constraints increase attack surfaces, and hence, pose additional challenges to privacy protection. To our knowledge, privacy protection for GNE seeking is still an open problem in the partial-decision information setting.

In this paper, we propose a fully distributed GNE seeking algorithm that can ensure both accurate convergence and rigorous ϵ -DP, with the cumulative privacy budget guaranteed to be finite even when the number of iterations tends to infinity. Our approach is motivated by the observation that DP-noises

The work was supported in part by the National Science Foundation under Grants ECCS-1912702, CCF-2106293, 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

enter the algorithm through inter-player interaction, and hence, their influence on convergence accuracy can be attenuated by gradually weakening inter-player interaction, which would become unnecessary anyway after convergence. To ensure that every player can still estimate global information necessary for GNE seeking, we judiciously design the weakening factor sequence for inter-player interaction and the stepsize sequence, which enables the achievement of accurate convergence even in the presence of DP-noises. We prove that the algorithm is ϵ -differentially private with a finite cumulative privacy budget, even on the infinite time horizon. Given that a key component of our GNE seeking algorithm is consensus tracking but the conventional consensus-tracking algorithm will lead to cumulative and exploding variance in the presence of persistent noise [40], we first propose a new robust consensus-tracking algorithm that can ensure both provable convergence and ϵ -DP under persistent DP-noise.

The main contributions are summarized as follows:

1) The proposed new consensus-tracking approach can ensure both provable convergence accuracy and ϵ -DP, a goal that has not been achieved before to the best of our knowledge.

2) The proposed fully distributed GNE seeking algorithm, to our knowledge, is the first to achieve rigorous ϵ -DP in GNE seeking under partial-decision information. Note that compared with our prior work on differentially private NE seeking without coupling constraints [38], [39], not only is the convergence analysis completely different due to the need to use dual variables to address shared coupling constraints (we use the operator-theoretic approach to facilitate the analysis here, which is not used in [38], [39]), the DP design is also much more complicated due to the need to share more information (both the primal and dual variables) in GNE seeking. In fact, the shared coupling constraints in GNE problems create additional attack surfaces (the additional shared variables provide attackers with more data to infer sensitive information) and thus challenges in privacy protection.

3) Besides achieving rigorous ϵ -DP, the approach can simultaneously ensure provable convergence to the GNE, which is in sharp contrast to existing DP solutions for coupling-constraint free aggregative games (e.g., [36], [37]) that have to trade provable convergence for DP.

4) To facilitate convergence analysis under DP-noises, we establish convergence results for stochastically-perturbed non-stationary fixed-point iteration processes. Given that fixed-point iteration processes lie at the core of many optimization and variational problems, the results are expected to have broad ramifications.

5) Even without considering privacy protection, our proof techniques are fundamentally different from existing counterparts and are of independent interest. More specifically, existing proof techniques (in, e.g., [3], [16], [37], [41], [42], [43], [44], [45]) for partial-decision information games rely on the geometric decreasing of consensus errors among the players. In the proposed approach, the diminishing interaction makes it impossible to have such geometric decreasing of consensus errors, which entails new proof techniques.

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 new consensus-tracking algorithm. This section also proves that the algorithm can ensure both provable accurate convergence and rigorous ϵ -DP with a finite cumulative privacy budget. Sec. IV proposes a fully distributed GNE seeking algorithm, while its convergence and privacy analysis are presented in Sec. V and Sec. VI, respectively. Sec. VII presents numerical simulation results to confirm the obtained results. Finally, Sec. VIII concludes the paper.

Notations: We use \mathbb{R}^d to denote the Euclidean space of dimension d and \mathbb{R}_+^d the set of all nonnegative vectors in \mathbb{R}^d . We write I_d for the identity matrix of dimension d , and $\mathbf{1}_d$ for the d -dimensional column vector with all entries equal to 1; in both cases we suppress the dimension when it is clear from the context. A vector is viewed as a column vector, and for a vector x , $[x]_i$ denotes its i th element. 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 $\|\cdot\|$. A^T denotes the transpose of a matrix A . Given vectors x_1, \dots, x_m , we define $x = \text{col}(x_1, \dots, x_m) = [x_1^T, \dots, x_m^T]^T$, $x_{-i} = \text{col}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m)$, and $\bar{x} = \frac{\sum_{i=1}^m x_i}{m}$. We define $[m] \triangleq \{1, 2, \dots, m\}$. Often, we abbreviate *almost surely* by *a.s.* Given sets $\Omega_1, \dots, \Omega_m$, we use $\Omega_1 \times \dots \times \Omega_m$ or $\prod_{i=1}^m \Omega_i$ to represent the Cartesian product of these sets. For a variable r^k , we use $\{r^k\}$ to denote the sequence of values of r^k for all $k \geq 0$.

II. PROBLEM FORMULATION AND PRELIMINARIES

A. Monotone Operators

All of the following are adopted from [46]. Let \mathcal{X} and \mathcal{Y} be non-empty sets, and $2^{\mathcal{Y}}$ be the family of all subsets of \mathcal{Y} . An operator $T : \mathcal{X} \rightarrow \mathcal{Y}$ is a mapping that maps every point $x \in \mathcal{X}$ to a point $T(x)$ in \mathcal{Y} . Thus, the notion $T : \mathcal{X} \rightarrow 2^{\mathcal{Y}}$ means that T is a set-valued operator from \mathcal{X} to \mathcal{Y} . Id denotes the identity operator. The domain of an operator T is represented by $\text{dom}T = \{x | T(x) \neq \emptyset\}$ where \emptyset denotes the empty set. The range of an operator T is defined by $\text{ran}T = \{u | \exists x, u \in T(x)\}$. The graph of T is defined as $\text{gra}T = \{(x, u) | u \in T(x)\}$, and the inverse of an operator is defined using graph as $\text{gra}T^{-1} = (u, x) | u \in T(x)$. The zero set of T is defined as $\text{zer}T = \{x | 0 \in T(x)\}$.

We consider monotone operators, i.e., operators satisfying $\langle x - y, u - v \rangle \geq 0$ for all (x, u) and (y, v) in $\text{gra}T$. An operator T is maximally monotone if $\text{gra}T$ is not contained in the graph of any other monotone operator. T is η -strongly monotone, with $\eta > 0$, if $\langle x - y, u - v \rangle \geq \eta \|x - y\|^2$ for all (x, u) and (y, v) in $\text{gra}T$. A set-valued operator $T : \mathbb{R}^d \rightarrow 2^{\mathbb{R}^d}$ is called restricted-strictly monotone with respect to $\mathcal{U} \subset \mathbb{R}^d$ if $\langle x^* - x, u^* - u \rangle > 0$ for all $u^* \in \mathcal{U}$, $u \in \mathbb{R}^d \setminus \mathcal{U}$, $x^* \in T(u^*)$, and $x \in T(u)$. For an operator $T : \mathbb{R}^d \rightarrow 2^{\mathbb{R}^d}$, we use $(\text{Id} + T)^{-1}$ to represent its resolvent, which is single valued and has domain equal to \mathbb{R}^d if T is maximally monotone. For two operators T_1 and T_2 , their composition is denoted as $T_1 \circ T_2$, and their sum $T_1 + T_2$ is defined as $\text{gra}(T_1 + T_2) = \{(x, y + z) | (x, y) \in \text{gra}T_1, (x, z) \in \text{gra}T_2\}$. If T_2 is single valued, then we always

have $\text{zer}(T_1 + T_2) = \text{Fix}((\text{Id} + T_1)^{-1} \circ (\text{Id} - T_2))$, where $\text{Fix}(T)$ denotes the fixed point set $\{x \in \text{dom}T | T(x) = x\}$.

The subdifferential ∂f of a proper lower semicontinuous convex function $f : x \rightarrow 2^{\mathbb{R}^d}$ is defined as $x \rightarrow \{g \in \mathbb{R}^d | f(y) \geq f(x) + \langle g, y - x \rangle, \forall y \in \text{dom}f\}$. We define the indicator function of Ω as $\iota_\Omega(x)$ ($\iota_\Omega(x) = 0$ for $x \in \Omega$ and $\iota_\Omega(x) = \infty$ for $x \notin \Omega$). $\partial\iota_\Omega$ is usually called the normal cone operator of Ω , and is denoted as N_Ω .

Operator T is nonexpansive if it satisfies $\|T(x) - T(y)\| \leq \|x - y\|$ for all x and y in $\text{dom}T$. It is α -averaged for $0 < \alpha < 1$ if there exists a nonexpansive operator R such that $T = (1 - \alpha)\text{Id} + \alpha R$ holds. T being α -averaged is equivalent to $\|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha} \|(x - y) - (Tx - Ty)\|^2$ for all $x, y \in \text{dom}T$ [46]. T is called β -cocoercive for $\beta > 0$ if βT is $\frac{1}{2}$ -averaged, i.e., $\|\beta T(x) - \beta T(y)\|^2 \leq \langle x - y, Tx - Ty \rangle$. If f is convex differentiable, with its gradient ∇f being θ -Lipschitz, then ∇f is $\frac{1}{\theta}$ -cocoercive.

B. On GNE Seeking

We consider a GNE problem among a set of m players (agents) $[m] = \{1, \dots, m\}$. We index the players by $1, 2, \dots, m$. Player i is characterized by a strategy set $\Omega_i \subseteq \mathbb{R}^d$ and a cost function $J_i(x_i, \bar{x})$, which is dependent on both its own decision $x_i \in \Omega_i \subseteq \mathbb{R}^d$ and the aggregate of all players' decisions $\bar{x} \triangleq \frac{\sum_{i=1}^m x_i}{m}$. Moreover, the decisions of all players must satisfy a shared affine coupling constraint

$$\sum_{i=1}^m C_i x_i \leq \sum_{i=1}^m c_i, \quad (1)$$

where $C_i \in \mathbb{R}^{n \times d}$ and $c_i \in \mathbb{R}^n$ are local parameters only known to player i . It is worth noting that such affine coupling constraints arise in various game scenarios involving upper or lower limits of shared resources [4], [11], [47], [48], and have been a common assumption in the literature of noncooperative games [1], [2], [3], [41], [49]. For notational simplicity, we define $C \triangleq [C_1, \dots, C_m] \in \mathbb{R}^{n \times md}$ and $c \triangleq \sum_{i=1}^m c_i \in \mathbb{R}^n$.

With these notations, we can formalize the game that player i faces as the following parameterized optimization problem:

$$\begin{aligned} \min_{x_i} & J_i(x_i, \bar{x}) \\ \text{s.t.} & x_i \in \Omega_i \text{ and } C_i x_i - c_i \leq \sum_{j \neq i, j \in [m]} c_j - C_j x_j. \end{aligned} \quad (2)$$

The constraint set Ω_i , the function $J_i(\cdot, \cdot)$, and the coupling matrices C_i as well as c_i are known to player i only.

Assumption 1. For every player $i \in [m]$, $J_i(x_i, \bar{x})$ is a differentiable convex function with respect to x_i given any fixed x_{-i} . $\Omega_i \subset \mathbb{R}^d$ is nonempty, compact, and convex. Moreover, the global feasible set

$$K \triangleq \Pi_{i=1}^m \Omega_i \cap \{x \in \mathbb{R}^{md} | \sum_{i=1}^m C_i x_i \leq \sum_{i=1}^m c_i\} \quad (3)$$

is nonempty and satisfies Slater's constraint qualification.

A GNE of game (2) is a collective strategy $x^* = \text{col}(x_1^*, \dots, x_m^*)$ such that the following relationship holds for any $i \in [m]$ and $x \in K$:

$$J_i(x_i^*, \bar{x}^*) \leq J_i(z, \frac{z}{m} + \frac{\sum_{j \neq i, j \in [m]} x_j^*}{m}), \text{ s.t. } (z, x_{-i}^*) \in K.$$

Inspired by the recent success of the operator-theoretic approach to GNE seeking [1], [2], [3], [50], we will employ

a monotone-operator based approach [46] to treat the convergence to GNE under DP-noises. More specifically, according to [1], [2], [3], a set of strategies x^* is a GNE of the game in (2) if and only if the following conditions are satisfied for some $\lambda_1^*, \dots, \lambda_m^* \in \mathbb{R}_+^n$:

$$\forall i \in [m], \quad \begin{cases} 0 \in \nabla_{x_i} J_i(x_i^*, \bar{x}^*) + N_{\Omega_i}(x_i^*) + C_i^T \lambda_i^* \\ 0 \leq \lambda_i^* \perp -(Cx^* - c) \geq 0 \end{cases}, \quad (4)$$

where the symbol \perp means perpendicular (namely, given $u, v \in \mathbb{R}^n$, $u \perp v$ means $u^T v = 0$).

Similar to [1], [2], [3], we are interested in a type of GNE called variational GNE, which has the economic interpretation of no price discrimination [51], and corresponds to a solution set of (4) with equal dual variables $\lambda_1^* = \dots = \lambda_m^*$ [1], [52]. Recently, [3] shows that the variational GNE corresponds to the zero set of the following set-valued mapping:

$$T : \begin{bmatrix} x \\ \lambda \end{bmatrix} \mapsto \begin{bmatrix} N_\Omega(x) + F(x) + \frac{C_f^T \lambda}{m} \\ N_{\mathbb{R}_+^{mn}}(\lambda) + \Pi_f \lambda - \frac{C_f x - c_f}{m} \end{bmatrix}, \quad (5)$$

where $x = \text{col}(x_1, \dots, x_m)$, $\lambda = \text{col}(\lambda_1, \dots, \lambda_m)$, $C_f = \mathbf{1}_m \otimes C$, $c_f = \mathbf{1}_m \otimes c$, $\Pi_f \triangleq (I_m - \frac{\mathbf{1}\mathbf{1}^T}{m}) \otimes I_n$, and $F(x)$ is the pseudogradient of the game defined as

$$F(x) = \text{col}(\nabla_{x_1} J(x_1, \bar{x}), \dots, \nabla_{x_m} J(x_m, \bar{x})). \quad (6)$$

Assumption 2. $F(\cdot)$ in (6) is μ -strongly monotone over \mathbb{R}^{md} .

Using the definition of $F(\cdot)$, we can further define $F_i(v, u) \triangleq (\frac{\partial}{\partial z_1} J_i(z_1, z_2) + \frac{1}{m} \frac{\partial}{\partial z_2} J_i(z_1, z_2))|_{z_1=v, z_2=u}$ and $F(v, u) \triangleq \text{col}(F_1(v, u), \dots, F_m(v, u))$. It can be verified that $F(x, \mathbf{1} \otimes \bar{x})$ is equal to $F(x)$ defined in (6).

Moreover, following [1], [3], [16], we make the following assumption on the mapping $F(\cdot, \cdot)$:

Assumption 3. There exists a constant \tilde{L} such that $F(\cdot, \cdot)$ satisfies $\|F(u_1, v_1) - F(u_2, v_2)\| \leq \tilde{L} \left\| \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} - \begin{bmatrix} u_2 \\ v_2 \end{bmatrix} \right\|$ for all $u_1, u_2 \in \mathbb{R}^{md}$ and $v_1, v_2 \in \mathbb{R}^{md}$.

By combining Assumption 2 and Assumption (3), we have that F is $\frac{\mu}{L^2}$ -cocoercive over \mathbb{R}^{md} [46].

Using operator splitting [46], the mapping T in (5) can be split into the sum of two mappings:

$$T_1 : \begin{bmatrix} x \\ \lambda \end{bmatrix} \mapsto \begin{bmatrix} F(x) \\ \Pi_f \lambda + \frac{c_f}{m} \end{bmatrix}, \quad T_2 : \begin{bmatrix} x \\ \lambda \end{bmatrix} \mapsto \begin{bmatrix} N_\Omega(x) + \frac{C_f^T \lambda}{m} \\ N_{\mathbb{R}_+^{mn}}(\lambda) - \frac{C_f x}{m} \end{bmatrix}. \quad (7)$$

It has been proven in [3] that the operator T_2 is maximally monotone, and that the operator T_1 is δ -cocoercive ($0 < \delta \leq \min\{1, \frac{\mu}{L^2}\}$ with μ from Assumption 2 and \tilde{L} from Assumption 3) and restricted-strictly monotone with respect to $\mathbb{R}^{md} \times E^\parallel$ where E^\parallel denotes the consensus subspace of dual variables. Furthermore, [3] also showed that when T is restricted-strictly monotone with respect to $\mathbb{R}^{md} \times E^\parallel$, the variational GNE of the game in (2) is fully characterized by the zeros of the operator T :

Lemma 1. [3] Under Assumption 1, 1) if $\text{zer}(T) \neq \emptyset$ and $\text{col}(x^*, \lambda^*) \in \text{zer}(T)$, then x^* is a variational GNE and $\lambda_1^* = \lambda_2^* = \dots = \lambda_m^* \in \mathbb{R}_+^n$; 2) if a variational GNE exists, then $\text{zer}(T)$ is nonempty.

We consider distributed algorithms for the game in (2), where no player has direct access to the average decision \bar{x} . Instead, each player has to locally estimate \bar{x} through interactions with its neighbors. We describe the local interaction using a weight matrix $L = \{L_{ij}\}$, where $L_{ij} > 0$ if player j can directly communicate with player i , and $L_{ij} = 0$ otherwise. For player $i \in [m]$, its neighbor set \mathbb{N}_i is the collection of players j such that $L_{ij} > 0$. We define $L_{ii} \triangleq -\sum_{j \in \mathbb{N}_i} L_{ij}$ for all $i \in [m]$. We make the following assumption on L :

Assumption 4. 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 $\|I + L - \frac{\mathbf{1}\mathbf{1}^T}{m}\| < 1$.

We also need the following lemma for convergence analysis:

Lemma 2. [53] Let $\{v^k\}$, $\{\alpha^k\}$, $\{p^k\}$ be random nonnegative scalar sequences, and $\{q^k\}$ be a deterministic nonnegative scalar sequence satisfying $\sum_{k=0}^{\infty} \alpha^k < \infty$ a.s., $\sum_{k=0}^{\infty} q^k = \infty$, $\sum_{k=0}^{\infty} p^k < \infty$ a.s. If there exists a $k_0 \geq 0$ such that

$$\mathbb{E}[v^{k+1} | \mathcal{F}^k] \leq (1 + \alpha^k - q^k)v^k + p^k, \quad \forall k \geq k_0 \quad \text{a.s.}$$

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 $\lim_{k \rightarrow \infty} v^k = 0$ hold almost surely.

Lemma 3. [53] Let $\{v^k\}$ be a nonnegative sequence, and $\{\alpha^k\}$ and $\{\beta^k\}$ be positive non-increasing sequences satisfying $\sum_{k=0}^{\infty} \alpha^k = \infty$, $\lim_{k \rightarrow \infty} \alpha^k = 0$, and $\lim_{k \rightarrow \infty} \frac{\beta^k}{\alpha^k} = 0$. If there exists a $k_0 \geq 0$ such that $v^{k+1} \leq (1 - \alpha^k)v^k + \beta^k$ holds for all $k \geq k_0$, then $v^k \leq C \frac{\beta^k}{\alpha^k}$ holds for all k , where C is some constant.

C. On Differential Privacy

We use the notion of ϵ -DP for continuous bit streams [54], which has recently been applied to distributed optimization (see [55] as well as our own work [53]). To enable ϵ -DP, we add Laplace noise to all shared messages. For a constant $\nu > 0$, we use $\text{Lap}(\nu)$ to denote a Laplace distribution of a scalar random variable with the probability density function $x \mapsto \frac{1}{2\nu} e^{-\frac{|x|}{\nu}}$. $\text{Lap}(\nu)$'s mean is zero and its variance is $2\nu^2$. To facilitate DP analysis, we represent the game \mathcal{P} in (2) by three parameters (K, \mathcal{J}, L) , where K defined in (3) is the domain of decision variables, $\mathcal{J} \triangleq \{J_1, \dots, J_m\}$, and L is the weight matrix. Then we define adjacent games as follows:

Definition 1. Two games $\mathcal{P} = (K, \mathcal{J}, L)$ and $\mathcal{P}' = (K', \mathcal{J}', L')$ are adjacent if the following conditions hold:

- $K = K'$ and $L = L'$, i.e., the domains of decision variables and the weight matrices are identical;
- there exists an $i \in [m]$ such that $J_i \neq J'_i$ but $J_j = J'_j$ for all $j \in [m]$, $j \neq i$;
- the different functions J_i and J'_i have similar behaviors around x^* , the GNE of \mathcal{P} . More specifically, there exists some $\delta > 0$ such that for all $v = \text{col}(v_1, \dots, v_m)$ and $v' = \text{col}(v'_1, \dots, v'_m)$ in $B_\delta(x^*) \triangleq \{u : u \in \mathbb{R}^{md}, \|u - x^*\| < \delta\}$, we have $\nabla F_i(v_i, \bar{v}) = \nabla F'_i(v'_i, \bar{v}')$ where $\bar{v} = \frac{\sum_{i=1}^m v_i}{m}$ and $\bar{v}' = \frac{\sum_{i=1}^m v'_i}{m}$.

Definition 1 implies that two GNE problems are adjacent if one player changes its cost function while all other game characteristics are identical.

Remark 1. In Definition 1, since the change from J_i to J'_i in the second condition can be arbitrary, the third condition is added to restrict the change magnitude. It is necessary to enabling rigorous ϵ -DP while maintaining accurate convergence. This is because DP aims to make observations statistically indistinguishable while accurate convergence means that the state will stop changing and remain time-invariant after a transient period. Hence, to make the observations of \mathcal{P} and \mathcal{P}' the same after their states converge and remain at their respective converging points, we have to require \mathcal{P} and \mathcal{P}' to have identical converging points.

Given a distributed GNE seeking algorithm, we represent an execution of such an algorithm as \mathcal{A} , which is an infinite sequence of the iteration variable ϑ , i.e., $\mathcal{A} = \{\vartheta^0, \vartheta^1, \dots\}$. We consider adversaries that can observe all communicated messages among the players. Therefore, the observation part of an execution is the infinite sequence of shared messages, which is denoted as \mathcal{O} . Given a distributed GNE problem \mathcal{P} and an initial state ϑ^0 , we define the observation mapping as $\mathcal{R}_{\mathcal{P}, \vartheta^0}(\mathcal{A}) \triangleq \mathcal{O}$. Given a GNE problem \mathcal{P} , observation sequence \mathcal{O} , and an initial state ϑ^0 , $\mathcal{R}_{\mathcal{P}, \vartheta^0}^{-1}(\mathcal{O})$ is the set of executions \mathcal{A} that can generate the observation \mathcal{O} .

Definition 2. (ϵ -differential privacy, adapted from [55]). For a given $\epsilon > 0$, an iterative distributed algorithm is ϵ -differentially private if for any two adjacent network games \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 , the following relationship always holds

$$\mathbb{P}[\mathcal{R}_{\mathcal{P}, \vartheta^0} \in \mathcal{O}_s] \leq e^\epsilon \mathbb{P}[\mathcal{R}_{\mathcal{P}', \vartheta^0} \in \mathcal{O}_s], \quad (8)$$

with the probability \mathbb{P} taken over randomness in iterations.

Since the observation sequence in Definition 2 involves observation values in multiple iterations, the privacy budget ϵ in (8) is a cumulative privacy budget for these multiple iterations. Sometimes, by fixing the length of the observation sequence to one (a single iteration k) in Definition 2, a privacy budget ϵ^k can be calculated for this particular iteration k [35]. In this case, according to the sequential composition property of DP [35], $\sum_{k=1}^{\bar{k}} \epsilon^k$ corresponds to ϵ in our Definition 2 using observation sequences of length \bar{k} (from $k = 1$ to $k = \bar{k}$).

ϵ -DP ensures that an adversary having access to all shared messages cannot gain information with a significant probability of any participating player's cost function. It can be seen that a smaller ϵ means a higher level of privacy protection.

III. DIFFERENTIALLY-PRIVATE CONSENSUS TRACKING

Lying at the core of our distributed GNE seeking algorithm is the consensus-tracking technique, which enables multiple agents to cooperative track the average of multiple time-varying reference signals (i.e., $\frac{\sum_{i=1}^m r_i^k}{m}$) while each individual agent i can only access one reference signal r_i^k . (Note that in this section, r_i^k is regarded as a given signal, and an instantiation will be given in the GNE seeking application in Sec. IV.) However, the conventional consensus-tracking algorithm (see, e.g., [23]) is sensitive to noise in that it leads to the accumulation and explosion of noise variance in the

presence of persistent information-sharing noise [40]. Hence, we first propose a new robust consensus-tracking algorithm. Our basic idea is to tailor consensus tracking for ϵ -DP by adding one stepsize parameter γ^k and a weakening factor χ^k (see Algorithm 1), which enable us to suppress the influence of persistent DP-noise, and hence, ensure both goals of rigorous ϵ -DP and guaranteed accuracy. Following the convention, we call a participant node an agent here.

Algorithm 1: Differentially-private consensus tracking with guaranteed convergence accuracy

Parameters: Weakening factor $\chi^k > 0$ and stepsize $\gamma^k > 0$.

Every agent i 's reference is r_i^k . Every agent i maintains one state variable x_i^k , which is initialized as $x_i^0 = r_i^0$.

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

- a) Every agent j adds persistent DP-noise ζ_j^k to its state x_j^k , and then sends the obscured state $x_j^k + \zeta_j^k$ to agent $i \in \mathbb{N}_j$.
- b) After receiving $x_j^k + \zeta_j^k$ from all $j \in \mathbb{N}_i$, agent i updates its state as follows:

$$x_i^{k+1} = (1 - \gamma^k)x_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij}(x_j^k + \zeta_j^k - x_i^k) + r_i^{k+1} - (1 - \gamma^k)r_i^k. \quad (9)$$

c) end

Remark 2. Our prior result in [38] also discusses enabling DP in a particular consensus-tracking application in aggregative games. However, the approach there leaks information of agent i to a neighboring agent j if agent j is the only neighbor of agent i . This is because at any time instant k , the noises entering into all agents' updates in [38] always add up to zero. Under such an obfuscation mechanism, when agent i only has one neighbor (say, agent j), its information will be leaked to agent j : When r_i^k is a zero signal, Theorem 3 in [56] shows that the initial state of agent i is inferrable by agent j ; When r_i^k is the local gradient of agent i (typical in gradient-tracking based distributed optimization), Theorem 3 in [57] proves that the final gradient information of agent i is inferrable by its only neighbor agent j . By employing an additional stepsize factor γ^k (besides the weakening factor χ^k for inter-agent coupling), our Algorithm 1 removes the requirement that the network-level sum of noises entering into agents' updates has to be zero, and hence, can avoid the privacy leakage in [38].

Assumption 5. For every $i \in [m]$ and every k , the noises in $\{\zeta_i^k\}$ are zero-mean independent random variables, and are independent of $\{x_i^0; i \in [m]\}$. The noise variances $\mathbb{E}[\|\zeta_i^k\|^2] = (\sigma_i^k)^2$ satisfy

$$\sum_{k=0}^{\infty} (\chi^k)^2 \max_{i \in [m]} (\sigma_i^k)^2 < \infty, \quad (10)$$

where $\{\chi^k\}$ is the weakening sequence from Algorithm 1. The initial random vectors satisfy $\mathbb{E}[\|x_i^0\|^2] < \infty, \forall i \in [m]$.

A. Convergence Analysis

To prove that Algorithm 1 can ensure every x_i^k to track the average target signal $\bar{r}^k \triangleq \frac{1}{m} \sum_{i=1}^m r_i^k$, we first prove that \bar{x}^k converges almost surely to \bar{r}^k under the following assumption:

Assumption 6. For every $i \in [m]$, there exist some nonnegative sequence $\{\beta^k\}$ and a constant C such that

$$\|r_i^{k+1} - (1 - \gamma^k)r_i^k\| \leq \beta^k C \quad (11)$$

holds, where γ^k is from Algorithm 1 and $\{\beta^k\}$ satisfies $\lim_{k \rightarrow \infty} \frac{\gamma^k}{\beta^k} < \infty$.

Note that the condition $\lim_{k \rightarrow \infty} \frac{\gamma^k}{\beta^k} < \infty$ is necessary since otherwise (11) will not hold when $\{r_i^k\}$ is a constant signal.

Lemma 4. Under Assumptions 4, 5, \bar{x}^k in Algorithm 1 converges a.s. to \bar{r}^k if the following conditions hold:

$$\sum_{k=0}^{\infty} \gamma^k = \infty, \quad \sum_{k=0}^{\infty} (\gamma^k)^2 < \infty.$$

Proof. According to the definitions of \bar{x}^k and \bar{r}^k , we have the following relationship based on (9):

$$\bar{x}^{k+1} = (1 - \gamma^k)\bar{x}^k + \chi^k \bar{\zeta}^k + \bar{r}^{k+1} - (1 - \gamma^k)\bar{r}^k, \quad (12)$$

where $\bar{\zeta}^k \triangleq \frac{\sum_{i=1}^m |L_{ii}| \zeta_i^k}{m}$ and we have used the symmetric property of L and the fact $L_{ii} \triangleq -\sum_{j \in \mathbb{N}_i} L_{ij}$.

The preceding relationship implies

$$\begin{aligned} \|\bar{x}^{k+1} - \bar{r}^{k+1}\|^2 &= (1 - \gamma^k)^2 \|\bar{x}^k - \bar{r}^k\|^2 + (\chi^k)^2 \|\bar{\zeta}^k\|^2 \\ &\quad + 2 \langle (1 - \gamma^k)(\bar{x}^k - \bar{r}^k), \chi^k \bar{\zeta}^k \rangle. \end{aligned}$$

Using the assumption on the DP-noise ζ_i^k in Assumption 5, taking the conditional expectation, given $\mathcal{F}^k = \{x^0, \dots, x^k\}$, we obtain the following inequality for all $k \geq 0$:

$$\begin{aligned} \mathbb{E}[\|\bar{x}^{k+1} - \bar{r}^{k+1}\|^2 | \mathcal{F}^k] &\leq (1 + (\gamma^k)^2 - 2\gamma^k) \|\bar{x}^k - \bar{r}^k\|^2 \\ &\quad + \frac{\sum_{i=1}^m L_{ii}^2 (\sigma_i^k)^2 (\chi^k)^2}{m}, \end{aligned}$$

where we have used the relationship $\|\bar{\zeta}^k\|^2 \leq \frac{\sum_{i=1}^m L_{ii}^2 \|\zeta_i^k\|^2}{m}$.

Under the conditions in the lemma statement, it can be seen that $\|\bar{x}^{k+1} - \bar{r}^{k+1}\|^2$ satisfies the conditions for v^k in Lemma 2, with $o^k = (\gamma^k)^2$, $q^k = 2\gamma^k$, and $p^k = \frac{\sum_{i=1}^m L_{ii}^2 (\sigma_i^k)^2 (\chi^k)^2}{m}$. Therefore, we have $\sum_{k=0}^{\infty} \gamma^k \|\bar{x}^k - \bar{r}^k\|^2 < \infty$ and $\|\bar{x}^{k+1} - \bar{r}^{k+1}\|^2$ converging a.s. to zero, and hence, \bar{x}^k converging a.s. to \bar{r}^k . ■

Remark 3. From the derivation of (12), it can be seen that χ^k has to be the same for all agents to make sure that it only attenuates the noise input to \bar{x}^k and does not introduce additional input terms on individual states.

Lemma 5. Under Assumption 4 and two positive sequences $\{\chi^k\}$ and $\{\gamma^k\}$ satisfying $\sum_{k=0}^{\infty} (\gamma^k)^2 < \infty$ and $\sum_{k=0}^{\infty} (\chi^k)^2 < \infty$, there always exists a $k_0 \geq 0$ such that $\|(1 - \gamma^k)(I - \frac{11}{m}L) + \chi^k L\| < 1 - \chi^k |\rho_2|$ holds for all $k \geq k_0$, where ρ_2 is the second largest eigenvalue of L .

Proof. Under Assumption 4 and the definition $L_{ii} = -\sum_{j \in \mathbb{N}_i} L_{ij}$, the Gershgorin circle theorem implies that all eigenvalues of L are non-positive, with one (and only one) of them being equal to 0. Arrange the eigenvalues of L as $\rho_m \leq \rho_{m-1} \leq \dots \leq \rho_2 < \rho_1 = 0$. It can be verified that the eigenvalues of $(1 - \gamma^k)I + \chi^k L$ are equal to $1 - \gamma^k + \chi^k \rho_m \leq 1 - \gamma^k + \chi^k \rho_{m-1} \leq \dots \leq 1 - \gamma^k + \chi^k \rho_2 < 1 - \gamma^k + \chi^k \rho_1 = 1 - \gamma^k$, and the eigenvalues of $(1 - \gamma^k)(I - \frac{11}{m}L) + \chi^k L = (1 - \gamma^k)I + \chi^k L + (1 - \gamma^k)\frac{11}{m}L$

are given by $\{1 - \gamma^k + \chi^k \rho_m, 1 - \gamma^k + \chi^k \rho_{m-1}, \dots, 1 - \gamma^k + \chi^k \rho_2, 0\}$, with $|\rho_m| \geq |\rho_{m-1}| \geq \dots \geq |\rho_2| > 0$. Hence, the norm of $\|(1 - \gamma^k)(I - \frac{\mathbf{1}\mathbf{1}^T}{m}) + \chi^k L\|$ is no larger than $\max\{|1 - \gamma^k + \chi^k \rho_m|, |1 - \gamma^k + \chi^k \rho_2|\}$. Further taking into account the fact that γ^k and χ^k decay to zero because they are square summable, we conclude that there always exists a $k_0 \geq 0$ such that $|1 - \gamma^k + \chi^k \rho_m| = 1 - \gamma^k - \chi^k |\rho_m| < 1 - \chi^k |\rho_m|$ and $|1 - \gamma^k + \chi^k \rho_2| = 1 - \gamma^k - \chi^k |\rho_2| < 1 - \chi^k |\rho_2|$ hold for all $k \geq k_0$. Given $|\rho_m| \geq |\rho_2|$, we have the stated result. ■

Theorem 1. Under Assumptions 4, 5, 6, if the nonnegative sequences $\{\gamma^k\}$ and $\{\chi^k\}$ in Algorithm 1 and the nonnegative sequence $\{\beta^k\}$ in Assumption 6 satisfy

$$\sum_{k=0}^{\infty} \gamma^k = \infty, \sum_{k=0}^{\infty} \chi^k = \infty, \sum_{k=0}^{\infty} (\chi^k)^2 < \infty, \sum_{k=0}^{\infty} \frac{(\beta^k)^2}{\chi^k} < \infty, \quad (13)$$

then, the following results hold almost surely:

- 1) every x_i^k in Algorithm 1 converges to $\bar{r}^k = \frac{\sum_{i=1}^m r_i^k}{m}$;
- 2) $\sum_{k=0}^{\infty} \chi^k \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2 < \infty$;
- 3) $\sum_{k=0}^{\infty} \beta^k \sum_{i=1}^m \|x_i^k - \bar{x}^k\| < \infty$.

Proof. For the convenience of analysis, we write the iterates of x_i^k on per-coordinate expressions. More specifically, for all $\ell = 1, \dots, d$, and $k \geq 0$, we define $x^k(\ell) = [x_1^k(\ell), \dots, x_m^k(\ell)]^T$ where $[x_i^k(\ell)]$ represents the ℓ th element of the vector x_i^k . Similarly, we define $r^k(\ell) = [r_1^k(\ell), \dots, r_m^k(\ell)]^T$ and $\zeta^k(\ell) = [\zeta_1^k(\ell), \dots, \zeta_m^k(\ell)]^T$. In this per-coordinate view, (9) has the following form for all $\ell = 1, \dots, d$, and $k \geq 0$:

$$x^{k+1}(\ell) = (1 - \gamma^k)x^k(\ell) + \chi^k L x^k(\ell) + \chi^k L^0 \zeta^k(\ell) + r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell), \quad (14)$$

where L^0 is obtained by replacing all diagonal entries of L with zero.

The dynamics of \bar{x}^k is (noting $\mathbf{1}^T L = 0$ from Assumption 4)

$$[\bar{x}^{k+1}]_\ell = \frac{1}{m} ((1 - \gamma^k)x^k(\ell) + \chi^k L^0 \zeta^k(\ell) + r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)). \quad (15)$$

Combining (14) and (15) yields (with $\Pi \triangleq I - \frac{\mathbf{1}\mathbf{1}^T}{m}$)

$$x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell = ((1 - \gamma^k)\Pi + \chi^k L)x^k(\ell) + \chi^k \Pi L^0 \zeta^k(\ell) + \Pi(r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)). \quad (16)$$

To simplify notations, we define $W^k \triangleq (1 - \gamma^k)\Pi + \chi^k L = (1 - \gamma^k)(I - \frac{\mathbf{1}\mathbf{1}^T}{m}) + \chi^k L$. Assumption 4 ensures $W^k \mathbf{1} = 0$, and hence $W^k \mathbf{1}[\bar{x}^k]_\ell = 0$ for any $1 \leq \ell \leq d$. Subtracting $W^k \mathbf{1}[\bar{x}^k]_\ell = 0$ from the right hand side of (16) yields

$$x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell = W^k(x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell) + \chi^k \Pi L^0 \zeta^k(\ell) + \Pi(r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)), \quad (17)$$

which, further implies (noting $\|\Pi\| = 1$)

$$\begin{aligned} & \|x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell\|^2 \\ & \leq \|W^k(x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell) + \Pi(r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell))\|^2 \\ & + 2\langle W^k(x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell) + \Pi(r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)), \\ & \quad \chi^k \Pi L^0 \zeta^k(\ell) \rangle + (\chi^k)^2 \|L^0\|^2 \|\zeta^k(\ell)\|^2. \end{aligned} \quad (18)$$

Using Assumption 5, taking the conditional expectation, given $\mathcal{F}^k = \{x^0, \dots, x^k\}$, we obtain the following inequality for all $k \geq 0$:

$$\begin{aligned} & \mathbb{E}[\|x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell\|^2 | \mathcal{F}^k] \\ & \leq (\|W^k\| \|x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell\| + \|r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)\|)^2 \\ & + (\chi^k)^2 \|L^0\|^2 \mathbb{E}[\|\zeta^k(\ell)\|^2]. \end{aligned} \quad (19)$$

For the first term on the right hand side of the preceding inequality, we bound it using the fact that there exists a $k_0 \geq 0$ such that $0 < \|W^k\| \leq 1 - \chi^k |\rho_2|$ holds for all $k \geq k_0$ (see Lemma 5). Hence, equation (19) implies the following relationship for all $k \geq k_0$:

$$\begin{aligned} & \mathbb{E}[\|x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell\|^2 | \mathcal{F}^k] \leq \\ & ((1 - \chi^k |\rho_2|) \|x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell\| + \|r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)\|)^2 \\ & + (\chi^k)^2 \|L^0\|^2 \mathbb{E}[\|\zeta^k(\ell)\|^2]. \end{aligned} \quad (20)$$

Next, we apply the inequality $(a + b)^2 \leq (1 + \epsilon)a^2 + (1 + \epsilon^{-1})b^2$, which is valid for any scalars a, b , and $\epsilon > 0$, to (20). More specifically, setting $\epsilon = \frac{\chi^k |\rho_2|}{1 - \chi^k |\rho_2|}$ (which leads to $(1 + \epsilon) = \frac{1}{1 - \chi^k |\rho_2|}$ and $1 + \epsilon^{-1} = \frac{1}{\chi^k |\rho_2|}$) yields

$$\begin{aligned} & \mathbb{E}[\|x^{k+1}(\ell) - \mathbf{1}[\bar{x}^{k+1}]_\ell\|^2 | \mathcal{F}^k] \leq (\chi^k)^2 \|L^0\|^2 \mathbb{E}[\|\zeta^k(\ell)\|^2] \\ & + (1 - \chi^k |\rho_2|) \|x^k(\ell) - \mathbf{1}[\bar{x}^k]_\ell\|^2 \\ & + \frac{1}{\chi^k |\rho_2|} \|r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)\|^2. \end{aligned} \quad (21)$$

Note that the following relations always hold: $\sum_{\ell=1}^d \|x^k(\ell) - [\bar{x}^k]_\ell \mathbf{1}\|^2 = \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2$, $\sum_{\ell=1}^d \|r^{k+1}(\ell) - (1 - \gamma^k)r^k(\ell)\|^2 = \sum_{i=1}^m \|r_i^{k+1} - (1 - \gamma^k)r_i^k\|^2$, $\sum_{\ell=1}^d \|\zeta^k(\ell)\|^2 = \sum_{i=1}^m \|\zeta_i^k\|^2$.

Hence, summing (21) over $\ell = 1, \dots, d$ leads to

$$\begin{aligned} & \mathbb{E} \left[\sum_{i=1}^m \|x_i^{k+1} - \bar{x}^{k+1}\|^2 | \mathcal{F}^k \right] \\ & \leq (\chi^k)^2 \|L^0\|^2 \sum_{i=1}^m (\sigma_i^k)^2 + (1 - \chi^k |\rho_2|) \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2 \\ & + \frac{1}{\chi^k |\rho_2|} \sum_{i=1}^m \|r_i^{k+1} - (1 - \gamma^k)r_i^k\|^2. \end{aligned} \quad (22)$$

Assumption 6 implies $\frac{1}{\chi^k |\rho_2|} \sum_{i=1}^m \|r_i^{k+1} - (1 - \gamma^k)r_i^k\|^2 \leq \frac{(\beta^k)^2}{\chi^k |\rho_2|} m C^2$. Submitting the relationship into (22) yields

$$\begin{aligned} & \mathbb{E} \left[\sum_{i=1}^m \|x_i^{k+1} - \bar{x}^{k+1}\|^2 | \mathcal{F}^k \right] \leq (\chi^k)^2 \|L^0\|^2 \sum_{i=1}^m (\sigma_i^k)^2 \\ & + (1 - \chi^k |\rho_2|) \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2 + \frac{(\beta^k)^2}{\chi^k |\rho_2|} m C^2. \end{aligned} \quad (23)$$

Therefore, under Assumption 5 and the conditions for χ^k and β^k in (13), we have that the sequence $\{\sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2\}$ satisfies the conditions for $\{v^k\}$ in Lemma 2, and hence, converges to zero almost surely. So x_i^k converges to \bar{x}^k almost surely. Further recalling \bar{x}^k converging a.s. to \bar{r}^k in Lemma 4 yields that x_i^k converges a.s. to \bar{r}^k .

Moreover, Lemma 2 also implies the following relation a.s.:

$$\sum_{k=0}^{\infty} \chi^k \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2 < \infty. \quad (24)$$

To prove the last statement, we invoke the Cauchy-Schwarz inequality, which ensures $\sum_{k=0}^{\infty} \sqrt{\chi^k \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2} \sqrt{\frac{(\beta^k)^2}{\chi^k}} \leq \sqrt{\sum_{k=0}^{\infty} \chi^k \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2} \sqrt{\sum_{k=0}^{\infty} \frac{(\beta^k)^2}{\chi^k}}$. Noting that the summand in the left hand side of the preceding inequality is actually $\beta^k \sqrt{\sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2}$, and the right hand side of the preceding inequality is less than infinity almost surely under the proven result in (24) and the assumption in (13), we have that $\sum_{k=0}^{\infty} \beta^k \sqrt{\sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2} < \infty$ holds almost surely. Further utilizing the relationship $\sum_{i=1}^m \|x_i^k - \bar{x}^k\| \leq \sqrt{m \sum_{i=1}^m \|x_i^k - \bar{x}^k\|^2}$ yields the stated result. ■

B. Differential-privacy Analysis

To analyze the level of DP protection on individual signals $\{r_i^k\}$, we first define adjacency for the consensus-tracking problem. Representing a consensus-tracking problem by two parameters (r, L) , where $r \triangleq \{\{r_1^k\}, \dots, \{r_m^k\}\}$ and L is the interaction weight matrix, we define adjacency between two consensus-tracking problems as follows:

Definition 3. Two consensus-tracking problems $\mathcal{P} = (r, L)$ and $\mathcal{P}' = (r', L')$ are adjacent if the following conditions hold:

- $L = L'$, i.e., the interaction weight matrices are identical;
- there exists an $i \in [m]$ such that $\{r_i^k\} \neq \{r_i^{k'}\}$ but $\{r_j^k\} = \{r_j^{k'}\}$ for all $j \in [m]$, $j \neq i$;
- the different sequences $\{r_i^k\}$ and $\{r_i^{k'}\}$ have similar asymptotic behaviors, i.e., there exists some $C_r > 0$ such that

$$\|r_i^k - r_i^{k'}\|_1 \leq C_r \chi^k \beta^k \quad (25)$$

holds for all $k \geq 0$, where χ^k and β^k are from Algorithm 1 and Assumption 6, respectively.

Definition 4. (ϵ -differential privacy). For a given $\epsilon > 0$, an iterative consensus-tracking algorithm is ϵ -differentially private if for any two adjacent consensus-tracking problems \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 , the following relationship always holds

$$\mathbb{P}[\mathcal{R}_{\mathcal{P}, \vartheta^0} \in \mathcal{O}_s] \leq e^\epsilon \mathbb{P}[\mathcal{R}_{\mathcal{P}', \vartheta^0} \in \mathcal{O}_s], \quad (26)$$

with the probability \mathbb{P} taken over the randomness over iteration processes.

From the theory of ϵ -DP, we have to characterize the sensitivity of an algorithm in order to quantify the amount of DP-noise that has to be injected to shared messages to enable a certain level of ϵ -DP. Similar to the sensitivity definition of iterative optimization algorithms in [55], we define the sensitivity of a consensus-tracking algorithm as follows:

Definition 5. At each iteration k , for any initial state ϑ^0 and any adjacent consensus-tracking problems \mathcal{P} and \mathcal{P}' , the sensitivity of Algorithm 1 is

$$\Delta^k \triangleq \sup_{\mathcal{O} \in \mathbb{O}} \left\{ \sup_{\vartheta \in \mathcal{R}_{\mathcal{P}, \vartheta^0}^{-1}(\mathcal{O}), \vartheta' \in \mathcal{R}_{\mathcal{P}', \vartheta^0}^{-1}(\mathcal{O})} \|\vartheta^k - \vartheta'^k\|_1 \right\}. \quad (27)$$

Based on the sensitivity definition, if we just focus on a specific iteration k and use the Laplace noise mechanism [35], the privacy budget for this particular iteration k will be $\epsilon^k = \frac{\Delta^k}{\nu^k}$ when the Laplace noise is with parameter ν^k . Using the sequential composition property of DP [35], the cumulative privacy budget $\sum_{k=1}^{\bar{k}} \epsilon^k$ for iterations from $k = 1$ to $k = \bar{k}$ can also be obtained:

Lemma 6. In Algorithm 1, at each iteration k , if each agent's DP-noise vector $\zeta_i^k \in \mathbb{R}^d$ consists of d independent Laplace noises with parameter ν^k such that $\sum_{k=1}^{\bar{k}} \frac{\Delta^k}{\nu^k} \leq \bar{\epsilon}$, then Algorithm 1 is $\bar{\epsilon}$ -differentially private with the cumulative privacy budget from iteration $k = 1$ to $k = \bar{k}$ less than $\bar{\epsilon}$.

Proof. The lemma follows the same line of reasoning of Lemma 2 in [55] (also see Theorem 3 in [36]). It can be intuitively understood as follows: At any single iteration k , since the algorithm sensitivity is given by Δ^k , Laplace noise with parameter ν^k leads to DP protection for this single iteration with a single-iteration privacy budget $\frac{\Delta^k}{\nu^k}$ (note that a smaller privacy budget means a stronger privacy protection). Given that the injected noises are independent across different iterations, according to the sequential composition property of DP [35], the cumulative privacy budget can be obtained by adding single-iteration privacy budgets in individual iterations. Therefore, for iterations from $k = 1$ to $k = \bar{k}$, the cumulative privacy budget is less than $\bar{\epsilon}$ when $\sum_{k=1}^{\bar{k}} \frac{\Delta^k}{\nu^k} \leq \bar{\epsilon}$ holds. ■

Theorem 2. Under the conditions of Theorem 1, if 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 5, then all agents will converge a.s. to the average tracking target \bar{r}^k . Moreover,

- 1) For any finite number of iterations \bar{k} , Algorithm 1 is ϵ -differentially private with the cumulative privacy budget bounded by $\epsilon \leq \sum_{k=1}^{\bar{k}} \frac{C_r \zeta^k}{\nu^k}$ where $\zeta^k \triangleq \sum_{p=1}^{k-1} \prod_{q=p}^{k-1} (1 - \gamma^q - \bar{L} \beta^q) \Lambda^{p-1} + \Lambda^{k-1}$, $\bar{L} \triangleq \min_i \{L_{ii}\}$, $\Lambda^k \triangleq \beta^{k+1} \chi^{k+1} + (1 - \gamma^k) \beta^k \chi^k$, and C_r is given in (25);
- 2) The cumulative privacy budget is finite for $\bar{k} \rightarrow \infty$ when the sequence $\{\frac{\beta^k}{\nu^k}\}$ is summable.

Proof. Since the Laplace noise satisfies Assumption 5 and we assume that all conditions of Theorem 1 hold, the convergence result follows directly from Theorem 1.

To prove the two statements on the strength of ϵ -DP, we first analyze the sensitivity of the algorithm. Given two adjacent consensus-tracking problems \mathcal{P} and \mathcal{P}' , for any given fixed observation \mathcal{O} and initial state $\vartheta^0 = x^0$, the sensitivity depends on $\|x^k - x'^k\|_1$ according to Definition 5. Since in \mathcal{P} and \mathcal{P}' , only one reference signal is different, we represent this different reference signal as the i th one, i.e., r_i in \mathcal{P} and r'_i in \mathcal{P}' , without loss of generality.

Because the initial conditions, reference signals, and observations of \mathcal{P} and \mathcal{P}' are identical for $j \neq i$, we have $x_j^k = x_j'^k$ for all $j \neq i$ and k . Therefore, $\|x^k - x'^k\|_1$ is always equal to $\|x_i^k - x_i'^k\|_1$.

Algorithm 1 implies

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

Note that we have used the fact that the observations $x_j^k + \zeta_j^k$ and $x_j'^k + \zeta_j'^k$ are the same.

Hence, using the third condition in Definition 1, the sensitivity Δ^k satisfies

$$\begin{aligned} \Delta^{k+1} &\leq (1 - \gamma^k - |L_{ii}|\chi^k)\Delta^k + C_r\beta^{k+1}\chi^{k+1} + C_r(1 - \gamma^k)\beta^k\chi^k \\ &\leq (1 - \gamma^k - \bar{L}\chi^k)\Delta^k + C_r\Lambda^k, \end{aligned} \quad (28)$$

where \bar{L} and Λ^k are defined in the theorem statement. Then, by iteration, we can arrive at the first privacy statement using Lemma 6 (noting $\Delta^0 = 0$ as \mathcal{P} and \mathcal{P}' have identical initial conditions).

For the infinity-time-horizon result in the second statement, we exploit Lemma 3 and the third condition in Definition 3. More specifically, from (28), according to Lemma 3, we can always find some \bar{C} such that $\Delta^k \leq \bar{C}\beta^k$ holds (note that γ^k decays faster than χ^k). Using Lemma 6, we can easily obtain $\epsilon \leq \sum_{k=1}^{\bar{k}} \frac{\bar{C}\beta^k}{\nu^k}$. Hence, ϵ will be finite even when \bar{k} tends to infinity if the sequence $\{\frac{\beta^k}{\nu^k}\}$ is summable. ■

Note that in consensus-tracking applications such as constraint-free Nash equilibrium seeking, [36] achieves ϵ -DP by enforcing the tracking reference to be summable (geometrically-decreasing, more specifically), which, however, also makes it impossible to ensure accurate convergence to the desired equilibrium. In our approach, by allowing the tracking reference $r_i^{k+1} - (1 - \gamma^k)r_i^k$ to be non-summable (since Assumption 6 allows $\sum_{k=0}^{\infty} \beta^k = \infty$), we achieve both accurate convergence and finite cumulative privacy budget. To our knowledge, this is the first consensus-tracking algorithm that can achieve both *a.s.* convergence to a tracking target and rigorous ϵ -DP, even on the infinite time horizon.

Remark 4. To ensure that the cumulative privacy budget $\epsilon = \sum_{k=1}^{\infty} \frac{\bar{C}\beta^k}{\nu^k}$ is bounded, we employ Laplace noise with parameter ν^k increasing with time (since we require the sequence $\{\frac{\beta^k}{\nu^k}\}$ to be summable while the sequence $\{\beta^k\}$ is non-summable). Because when sending signals, every agent i always uses x_i^k , the amplitude of which is time-invariant, an increasing ν^k makes the noise-to-signal amplitude ratio increase with time. However, since it is $\chi^k \text{Lap}(\nu^k)$ that is actually fed into the algorithm, and the increase in the noise level ν^k is outweighed by the decrease of χ^k (see Assumption 5), the actual noise fed into the algorithm still decays with time, which explains why Algorithm 1 can ensure every agent's accurate convergence. Moreover, according to Theorem 1, the convergence will not be affected if we scale ν^k by any constant $\frac{1}{\epsilon} > 0$ to achieve any desired level of ϵ -DP, as long as ν^k (with associated variance $(\sigma_i^k)^2 = 2(\nu^k)^2$) satisfies Assumption 5.

IV. DIFFERENTIALLY-PRIVATE GNE SEEKING

Based on our differentially-private consensus-tracking algorithm, we propose a differentially-private GNE seeking algorithm that can ensure both rigorous ϵ -DP and provable convergence to the GNE, which is summarized in Algorithm 2. Inspired by the distributed GNE seeking algorithm in [3], we let each player maintain three estimates of the global information that are not locally available. More specifically, σ_i^k is used to track the average \bar{x}^k of the decision variables x_i^k of all players, z_i^k is used to track the average $\bar{\lambda}^k$ of the dual variables λ_i^k , and y_i^k is used to track the average \bar{d}^k of the violation of the coupling constraint, which is defined as

$$d_i^k \triangleq 2C_i\tilde{x}_i^k - C_i x_i^k - c_i \quad (29)$$

for player i (where \tilde{x}_i^k is an auxiliary variable with update rule given in the first line in (30)).

Algorithm 2: Differentially-private GNE seeking algorithm with guaranteed convergence accuracy

Parameters: Weakening factor $\chi^k > 0$ and stepsizes $\alpha^k, \gamma^k > 0$.

Every player i maintains a decision variable x_i^k (with values at $k = -1$ and $k = 0$, i.e., x_i^{-1} and x_i^0 , selected randomly in $\Omega_i \subseteq \mathbb{R}^d$) and an auxiliary variable \tilde{x}_i^k (with value at $k = 0$, i.e., \tilde{x}_i^0 , selected randomly in $\Omega_i \subseteq \mathbb{R}^d$). Player i also maintains a dual variable λ_i^k , which is initialized randomly in \mathbb{R}_+^n . Every player i maintains local estimates (represented by σ_i^k , z_i^k , and y_i^k) of global information \bar{x}^k , $\bar{\lambda}^k$, and \bar{d}^k . These estimates are initialized as $\sigma_i^0 = x_i^0$, $z_i^0 = \lambda_i^0$, and $y_i^0 = 2C_i\tilde{x}_i^{-1} - C_i x_i^{-1} - c_i$, respectively.

for $k = 0, 1, \dots$ **do**

- Every player j adds persistent DP-noises ζ_j^k , ξ_j^k , and v_j^k to σ_j^k , y_j^k , and z_j^k , respectively, and then sends the obscured estimates $\sigma_j^k + \zeta_j^k$, $y_j^k + \xi_j^k$, and $z_j^k + v_j^k$ to player $i \in \mathbb{N}_j$.
- After receiving $\sigma_j^k + \zeta_j^k$, $y_j^k + \xi_j^k$, and $z_j^k + v_j^k$ from all $j \in \mathbb{N}_i$, player i updates its decision variable and estimates as follows:

$$\begin{aligned} \tilde{x}_i^k &= \Pi_{\Omega_i} [x_i^k - \alpha^k(F_i(x_i^k, \sigma_i^k) + C_i^T z_i^k)], \\ y_i^{k+1} &= (1 - \gamma^k)y_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij}(y_j^k + \xi_j^k - y_i^k) \\ &\quad + d_i^k - (1 - \gamma^k)d_i^{k-1}, \\ \tilde{\lambda}_i^k &= \Pi_{\mathbb{R}_+^n} [\lambda_i^k + \alpha^k(y_i^{k+1} - \lambda_i^k + z_i^k)], \\ x_i^{k+1} &= x_i^k + \gamma^k(\tilde{x}_i^k - x_i^k), \\ \lambda_i^{k+1} &= \lambda_i^k + \gamma^k(\tilde{\lambda}_i^k - \lambda_i^k), \\ \sigma_i^{k+1} &= (1 - \gamma^k)\sigma_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij}(\sigma_j^k + \zeta_j^k - \sigma_i^k) \\ &\quad + x_i^{k+1} - (1 - \gamma^k)x_i^k, \\ z_i^{k+1} &= (1 - \gamma^k)z_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij}(z_j^k + v_j^k - z_i^k) \\ &\quad + \lambda_i^{k+1} - (1 - \gamma^k)\lambda_i^k. \end{aligned} \quad (30)$$

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

c) end

Remark 5. In many applications, constraint parameters contain sensitive information (e.g., local power generation in smart grid [58]) and should be kept private. This motivates us to mask all transmitted variables with noises, including dual variables corresponding to constraints. In fact, in many cases, the iteration trajectory of dual variables still bears information of the primal variable, which makes it necessary to mask these variables. This is particularly the case in primal-dual subgradient methods like [59] and distributed dual subgradient methods like [60] where disclosing dual updates in two consecutive iterations allows an observer to directly calculate some function value of the primal variable.

$\{\chi^k\}$, which diminishes with time, is used to suppress the influence of DP-noises ζ_i^k , ξ_i^k , and v_i^k on the convergence point. The stepsizes $\{\alpha^k\}$, $\{\gamma^k\}$ and attenuation sequence $\{\chi^k\}$ have to be designed appropriately to guarantee the accurate convergence of the iterate vector $x^k \triangleq \text{col}(x_1^k, \dots, x_m^k)$ to the GNE $x^* \triangleq \text{col}(x_1^*, \dots, x_m^*)$. The DP noise sequences $\{\zeta_i^k\}$, $\{\xi_i^k\}$, $\{v_i^k\}$, $i \in [m]$ satisfy the assumption below:

Assumption 7. In Algorithm 2, for every $i \in [m]$, the noise sequences $\{\zeta_i^k\}$, $\{\xi_i^k\}$, and $\{v_i^k\}$ are zero-mean independent random variables, and are independent of $\{\sigma_i^0, z_i^0, y_i^0; i \in [m]\}$. Also, for every k , the noise collection $\{\zeta_j^k, \xi_j^k, v_j^k; j \in [m]\}$ is independent. The noise variances $(\sigma_{\zeta,i}^k)^2 = \mathbb{E}[\|\zeta_i^k\|^2]$, $(\sigma_{\xi,i}^k)^2 = \mathbb{E}[\|\xi_i^k\|^2]$, and $(\sigma_{v,i}^k)^2 = \mathbb{E}[\|v_i^k\|^2]$ satisfy $\sum_{k=0}^{\infty} (\chi^k)^2 \max_{i \in [m]} \{(\sigma_{\zeta,i}^k)^2, (\sigma_{\xi,i}^k)^2, (\sigma_{v,i}^k)^2\} < \infty$.

(31)

The initial random vectors satisfy $\max\{\mathbb{E}[\|x_i^0\|^2], \mathbb{E}[\|z_i^0\|^2], \mathbb{E}[\|y_i^0\|^2]\} < \infty, \forall i \in [m]$.

V. CONVERGENCE ANALYSIS OF ALGORITHM 2

Our convergence analysis leverages stochastically-perturbed nonstationary fixed-point iteration processes. Recently, the authors in [61] show that stochastically-perturbed nonstationary fixed-point iteration processes with nonexpansive operators can still converge to a fixed point if the iteration stepsize sequence $\{\gamma^k\}$ is lower bounded away from 0, i.e., $\liminf_{k \rightarrow \infty} \gamma^k > 0$. In our case, since the DP-noises are persistent to ensure a strong privacy, the stepsize has to decay to zero to ensure a finite privacy budget. Hence, we extend the result in [61] to the case where the iteration stepsize decays to zero. Interestingly, we prove that *a.s.* convergence to a fixed point can still be achieved in this case. It is worth noting that our results generalize the Krasnosel'skiĭ-Mann iteration process in [46] which addresses a time-invariant operator, and the inexact Krasnosel'skiĭ-Mann iteration process in [3], [62] which addresses deterministic errors.

We first prove that a stochastically-perturbed nonstationary fixed-point iteration process can still ensure *a.s.* convergence to a fixed point, even when the stepsize decays to zero:

Lemma 7. If there exists some $k_0 \geq 0$ such that for every $k \geq k_0$, $R_k : \mathbb{R}^p \rightarrow \mathbb{R}^p$ are nonexpansive operators and $\mathbf{F} \triangleq \bigcap_{k \geq k_0} \text{Fix}(R_k) \neq \emptyset$, then, under iteration

$$\omega^{k+1} = \omega^k + \gamma^k (R_k(\omega^k) + e^k - \omega^k), \quad (32)$$

ω^k converges *a.s.* to some $\omega^* \in \mathbf{F}$ when the following conditions are satisfied:

- 1) The nonnegative sequence $\{\gamma^k\}$ satisfies $\sum_{k=0}^{\infty} \gamma^k (1 - \gamma^k) = +\infty$;
- 2) The error sequence $\{e^k\}$ satisfies $\sum_{k=0}^{\infty} \gamma^k \sqrt{\mathbb{E}[\|e^k\|^2 | \mathcal{F}^k]} < \infty$ almost surely, where $\mathcal{F}^k = \{\omega^\ell; 0 \leq \ell \leq k\}$;
- 3) For every $\{\omega^k\}$, $\lim_{k \rightarrow \infty} \|R_k(\omega^k) - \omega^k\|$ exists almost surely;
- 4) The cluster points of $\{\omega^k\}$ belong to \mathbf{F} almost surely.

Proof. This lemma is an adaptation of Corollary 2.7 and Theorem 3.2 of [61]. More specifically, we use the *a.s.* convergence of $\lim_{k \rightarrow \infty} \|R_k(x^k) - x^k\|$ in condition 3) to replace the requirement of $\liminf_{k \rightarrow \infty} \gamma^k > 0$ in Theorem 3.2 of [61], which serves the same purpose in the derivation. It is worth noting that since the results in Corollary 2.7 and Theorem 3.2 of [61] are asymptotic, they remain valid when the starting index is shifted from $k = 0$ to $k = k_0$, for an arbitrary $k_0 \geq 0$. ■

Remark 6. Since we do not impose any extra conditions on R_k except being nonexpansive, we introduce Condition 3) and Condition 4) to ensure convergence behaviors. The two conditions can be proven to be satisfied for the particular R_k corresponding to our proposed algorithms.

Remark 7. Different from [61] which requires the stepsize in nonstationary fixed-point iteration processes to be bounded away from 0, Lemma 7 allows the stepsize to decay to zero. This difference is significant in that allowing the stepsize to decay to zero is key to ensuring the privacy budget to be bounded, as detailed later in Theorem 4.

In the remainder of the convergence analysis, we show that Algorithm 2 is indeed a stochastically-perturbed version of a nonstationary fixed-point iteration process. Hence, we can leverage Lemma 7 to establish the *a.s.* convergence of Algorithm 2 under persistent DP-noises. The main challenges lie in 1) finding a sequence of deterministic nonexpansive operators $R_k(\cdot)$ that have convergent properties as required in the third and fourth conditions in Lemma 7, and 2) proving that Algorithm 2 is a stochastically-perturbed version of $R_k(\cdot)$ with the amplitude of perturbation e^k satisfying the *a.s.* summability condition in condition 2) of Lemma 7. We address the two challenges in Sec. V-A and Sec. V-B, respectively.

A. Finding a non-perturbed fixed-point iteration process

In this subsection, we show that when the estimates σ_i^k , y_i^k , and z_i^k in Algorithm 2 are replaced with their target signals that they are designed to track, i.e., \bar{x}^k , \bar{d}^k , and $\bar{\lambda}^k$, then the iterations in Algorithm 2 reduce to a nonstationary fixed-point iteration process with involved operators being nonexpansive and satisfying the conditions 3) and 4) in Lemma 7, if $\{\alpha^k\}$ is set as a nonnegative sequence satisfying $\sum_{k=0}^{\infty} \alpha^k = \infty$ and $\sum_{k=0}^{\infty} (\alpha^k)^2 < \infty$. Note that replacing σ_i^k , y_i^k , and z_i^k with \bar{x}^k , \bar{d}^k , and $\bar{\lambda}^k$ corresponds to the full-decision information setting where a central coordinator has direct access to all x_i^k , d_i^k , and λ_i^k for $i \in [m]$, and hence, can compute and disperse

the respective averages to all players. The reduced algorithm is summarized in Algorithm 3:

Algorithm 3: GNE seeking with full-decision information

Parameters: Stepsizes $\alpha^k > 0$ and $\gamma^k > 0$.

Every player i maintains one decision variable x_i^k , which is initialized randomly in $\Omega_i \subseteq \mathbb{R}^d$. Every player i also maintains a dual variable λ_i^k , which is initialized randomly in \mathbb{R}_+^n .

for $k = 0, 1, \dots$ **do**

a) Every player i calculates

$$\begin{aligned} \tilde{x}_i^k &= \Pi_{\Omega_i} [x_i^k - \alpha^k (F_i(x_i^k, \bar{x}^k) + C_i^T \bar{\lambda}^k)], \\ d_i^k &= 2C_i \tilde{x}_i^k - C_i x_i^k - c_i, \\ \tilde{\lambda}_i^k &= \Pi_{\mathbb{R}_+^n} [\lambda_i^k + \alpha^k (\bar{d}_i^k - \lambda_i^k + \bar{\lambda}^k)]. \end{aligned} \quad (33)$$

b) Player i updates its variables as follows:

$$x_i^{k+1} = x_i^k + \gamma^k (\tilde{x}_i^k - x_i^k), \quad \lambda_i^{k+1} = \lambda_i^k + \gamma^k (\tilde{\lambda}_i^k - \lambda_i^k). \quad (34)$$

c) **end**

Following [3], we have the following results:

Proposition 1. *In both Algorithm 2 and Algorithm 3, x_i^k , \tilde{x}_i^k , $\tilde{\lambda}_i^k$, λ_i^k can always be made bounded.*

Proof. One can obtain that \tilde{x}_i^k and x_i^k are always within Ω_i since \tilde{x}_i^k is always obtained by projection onto Ω_i , and the update $x_i^k = x_i^{k-1} + \gamma^k (\tilde{x}_i^{k-1} - x_i^{k-1})$ amounts to a convex combination of elements in the convex set Ω_i . Since Ω_i is compact, we have that \tilde{x}_i^k and x_i^k are bounded. The boundedness of $\tilde{\lambda}_i^k$ can be easily enforced in a distributed manner by a superset based technique (commonly used in distributed optimization [63] and distributed GNE seeking [3]), which enables us to replace the projection set \mathbb{R}_+^n in the update rule of $\tilde{\lambda}_i^k$ with a convex compact set. It is worth noting that this replacement does not affect the convergence analysis or the DP-design. Under the update rule $\lambda_i^k = \lambda_i^{k-1} + \gamma^k (\tilde{\lambda}_i^{k-1} - \lambda_i^{k-1})$, a bounded $\tilde{\lambda}_i^k$ implies a bounded λ_i^k . ■

Lemma 8. *Algorithm 3 corresponds to a nonstationary fixed-point iteration process*

$$\omega^{k+1} = \omega^k + \gamma^k (R_k(\omega^k) - \omega^k), \quad (35)$$

where $\omega = \text{col}(x^k, \lambda^k)$, and the operators R_k are given by

$$R_k = (\text{Id} + (\Phi^k)^{-1} T_2)^{-1} \circ (\text{Id} - (\Phi^k)^{-1} T_1), \quad (36)$$

with T_1 and T_2 given in (7) and

$$\Phi^k = \begin{bmatrix} (\alpha^k)^{-1} I_{md} & -\frac{C_f^T}{m} \\ -\frac{C_f}{m} & (\alpha^k)^{-1} I_{mn} \end{bmatrix}. \quad (37)$$

Moreover, when the non-increasing sequence $\{\alpha^k\}$ satisfies $\sum_{k=0}^{\infty} \alpha^k = \infty$, $\sum_{k=0}^{\infty} (\alpha^k)^2 < \infty$, and

$$\alpha^k \leq \frac{m}{2 \max_{i \in [m]} \|C_i\|}, \quad (38)$$

there exists some $k_0 \geq 0$ such that for all $k \geq k_0$, $\{R_k\}$ are nonexpansive operators and the conditions 3) and 4) in Lemma 7 are satisfied.

Proof. Algorithm 3 can be written in the following compact form:

$$\begin{aligned} \tilde{x}^k &= \Pi_{\Omega} [x^k - \alpha^k (F(x^k, \bar{x}_1^k) + C_d^T \bar{\lambda}_1^k)], \\ \tilde{\lambda}^k &= \Pi_{\mathbb{R}_+^{mn}} [\lambda^k + \alpha^k (\bar{d}_1^k - \lambda^k + \bar{\lambda}_1^k)], \\ x^{k+1} &= x^k + \gamma^k (\tilde{x}^k - x^k), \\ \lambda^{k+1} &= \lambda^k + \gamma^k (\tilde{\lambda}^k - \lambda^k), \end{aligned} \quad (39)$$

where $\bar{x}_1^k = \mathbf{1} \otimes \bar{x}^k$, $\bar{\lambda}_1^k = \mathbf{1} \otimes \bar{\lambda}^k$, $\bar{d}_1^k = \mathbf{1} \otimes \bar{d}^k$, and $C_d = \text{diag}(C_1, \dots, C_m)$. Then, following the forward-backward operator splitting approach in [2], [3], we can obtain that Algorithm 3 corresponds to a nonstationary fixed-point iteration process (35).

Next, we prove that there exists a $k_0 \geq 0$ such that R_k s are nonexpansive for all $k \geq k_0$. According to the Gershgorin circle theorem, the eigenvalues of Φ^k are no less than $\min_{i \in [m]} \{\frac{1}{\alpha^k} - \frac{\|C_i\|}{m}\}$. Hence, we have that the eigenvalues of Φ^k are no less than $\frac{1}{2\alpha^k}$ according to (38), implying $\|\Phi^k\| \geq \frac{1}{2\alpha^k}$ and $\|(\Phi^k)^{-1}\| \leq 2\alpha^k$. Using this property of $(\Phi^k)^{-1}$, we proceed to prove that there exists a $k_1 \geq 0$ such that for all $k \geq k_1$, $(\Phi^k)^{-1} T_1$ is $\frac{\delta}{2\alpha^k}$ -cocoercive in the Euclidean norm for all $0 < \delta \leq \min\{1, \frac{\mu}{L^2}\}$, where μ and \tilde{L} are from Assumption 2 and Assumption 3, respectively. To this end, using the matrix inversion lemma ([64], page 258), we rewrite $(\Phi^k)^{-1}$ explicitly as follows:

$$(\Phi^k)^{-1} = I_{\alpha^k} \left(I - U (I_{\diamond}^{-1} + I_{\alpha^k} U)^{-1} I_{\alpha^k} \right), \quad (40)$$

where $I_{\alpha^k} \triangleq \alpha^k I_{m(d+n)}$, U is a block diagonal matrix $U \triangleq \text{diag}(-\frac{C_f^T}{m}, -\frac{C_f}{m})$, and I_{\diamond} is an anti-diagonal block matrix $I_{\diamond} \triangleq [0, I_{md}; I_{mn}, 0]$.

Defining M^k as $U(I_{\diamond}^{-1} + \alpha^k U)^{-1}$, we can obtain the following relation from (40):

$$(\Phi^k)^{-1} = \alpha^k (I - \alpha^k M^k). \quad (41)$$

According to the properties of cocoercive mapping, the relationship in (41) implies that to prove $(\Phi^k)^{-1} T_1$ is $\frac{\delta}{2\alpha^k}$ -cocoercive, we only need to prove $(I - \alpha^k M^k) T_1$ being $\frac{\delta}{2}$ -cocoercive, i.e.,

$$\begin{aligned} &\langle (I - \alpha^k M^k)(T_1(\omega_1) - T_1(\omega_2)), \omega_1 - \omega_2 \rangle \\ &\geq \frac{\delta}{2} \|(I - \alpha^k M^k)(T_1(\omega_1) - T_1(\omega_2))\|^2 \end{aligned} \quad (42)$$

for all $\omega_1, \omega_2 \in \Omega \times E$ where $\Omega = \Omega_1 \times \dots \times \Omega_m$ and E denotes the space of augmented dual variables $\text{col}(\lambda_1, \dots, \lambda_m)$.

Since $F(\cdot)$ is $\frac{\mu}{L^2}$ -cocoercive in the Euclidean norm, following the same proof of Lemma 3(ii) in [3], T_1 is δ -cocoercive for any $0 < \delta \leq \min\{1, \frac{\mu}{L^2}\}$. Hence, the left hand side of (42) satisfies

$$\begin{aligned} &\langle (I - \alpha^k M^k)(T_1(\omega_1) - T_1(\omega_2)), \omega_1 - \omega_2 \rangle \\ &= \langle T_1(\omega_1) - T_1(\omega_2), \omega_1 - \omega_2 \rangle \\ &\quad - \alpha^k \langle M^k(T_1(\omega_1) - T_1(\omega_2)), \omega_1 - \omega_2 \rangle \\ &\geq \delta \|T_1(\omega_1) - T_1(\omega_2)\|^2 - \alpha^k \langle M^k(T_1(\omega_1) - T_1(\omega_2)), \omega_1 - \omega_2 \rangle \\ &\geq \delta \|T_1(\omega_1) - T_1(\omega_2)\|^2 - \alpha^k \|M^k\| \|T_1(\omega_1) - T_1(\omega_2)\| \|\omega_1 - \omega_2\|. \end{aligned} \quad (43)$$

The right hand side of (42) can be verified to satisfy

$$\begin{aligned} & \frac{\delta}{2} \|(I - \alpha^k M^k)(T_1(\omega_1) - T_1(\omega_2))\|^2 \\ &= \frac{\delta}{2} \|T_1(\omega_1) - T_1(\omega_2)\|^2 + \frac{\delta(\alpha^k)^2}{2} \|M^k(T_1(\omega_1) - T_1(\omega_2))\|^2 \\ & \quad + \delta\alpha^k \langle T_1(\omega_1) - T_1(\omega_2), M^k(T_1(\omega_1) - T_1(\omega_2)) \rangle \\ &\leq \frac{\delta}{2} (1 + 2\alpha^k \|M^k\| + (\alpha^k)^2 \|M^k\|^2) \|T_1(\omega_1) - T_1(\omega_2)\|^2. \end{aligned} \quad (44)$$

Therefore, to prove the relation in (42), we only need to prove that (43) is no smaller than (44), or (after combining like terms) the following inequality

$$\begin{aligned} & \frac{\delta}{2} (1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2) \|T_1(\omega_1) - T_1(\omega_2)\|^2 \geq \\ & \geq \alpha^k \|M^k\| \|T_1(\omega_1) - T_1(\omega_2)\| \|\omega_1 - \omega_2\|. \end{aligned} \quad (45)$$

Since (45) always holds under $\|T_1(\omega_1) - T_1(\omega_2)\| = 0$, proving (45) is equivalent to proving

$$\begin{aligned} & \frac{\delta}{2} (1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2) \|T_1(\omega_1) - T_1(\omega_2)\| \\ & \geq \alpha^k \|M^k\| \|\omega_1 - \omega_2\| \end{aligned} \quad (46)$$

under the condition $\|T_1(\omega_1) - T_1(\omega_2)\| \neq 0$.

Hence, proving $(\Phi^k)^{-1}T_1$ to be $\frac{\delta}{2\alpha^k}$ -cocoercive in the Euclidean norm when k is larger than some k_1 reduces to proving

$$\begin{aligned} & \|T_1(\omega_1) - T_1(\omega_2)\| \\ & \geq \frac{\alpha^k \|M^k\|}{\frac{\delta}{2} (1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2)} \|\omega_1 - \omega_2\| \end{aligned} \quad (47)$$

for all $k \geq k_1$ under the condition $\|T_1(\omega_1) - T_1(\omega_2)\| \neq 0$ (note that $\|M^k\|$ is always bounded and α^k tends to zero under the Lemma conditions, which ensure $1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2 > 0$ when k_1 is sufficiently large).

Recalling the definition of ω , i.e., $\omega_1 \triangleq \text{col}(x_1, \lambda_1)$ and $\omega_2 \triangleq \text{col}(x_2, \lambda_2)$, the condition $\|T_1(\omega_1) - T_1(\omega_2)\| \neq 0$ implies either 1) $x_1 \neq x_2$, or 2) $\lambda_1 - \lambda_2 \notin E^\parallel$ where E^\parallel denotes the consensus subspace of the dual variables $\lambda = \text{col}(\lambda_1, \dots, \lambda_m)$. Denoting E^\perp as the disagreement subspace of the dual variables λ , we can decompose $\lambda_1 - \lambda_2$ as $\lambda_1 - \lambda_2 = (\lambda_1 - \lambda_2)_\parallel + (\lambda_1 - \lambda_2)_\perp$ following the argument in the proof of Lemma 3 in [3], where $(\lambda_1 - \lambda_2)_\parallel \in E^\parallel$ and $(\lambda_1 - \lambda_2)_\perp \in E^\perp$. Therefore, the two conditions under which $\|T_1(\omega_1) - T_1(\omega_2)\| \neq 0$ can be restated as: case 1) $x_1 \neq x_2$, and case 2) $(\lambda_1 - \lambda_2)_\perp \neq 0$. Next we show that in both cases, (47) can be satisfied when k is sufficiently large.

- 1) When $x_1 \neq x_2$ is true, the strongly monotone condition in Assumption 2 implies $\|T_1(\omega_1) - T_1(\omega_2)\| \geq \mu \|x_1 - x_2\| = \mu \|\omega_1 - \omega_2\| \frac{\|x_1 - x_2\|}{\|\omega_1 - \omega_2\|}$. Because the primal variables x_1 and x_2 are always within the compact set Ω , and the dual variables are bounded (see Proposition 1), $\|\omega_1 - \omega_2\|$ is always bounded, implying the existence of some k_1 such that for all $k \geq k_1$ (when α^k is sufficiently small), $\mu \frac{\|x_1 - x_2\|}{\|\omega_1 - \omega_2\|} \geq \frac{\alpha^k \|M^k\|}{\frac{\delta}{2} (1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2)}$ holds (hence (47) holds) under $x_1 \neq x_2$.

- 2) When $(\lambda_1 - \lambda_2)_\perp \neq 0$ is true, using Assumption 2 and an argument similar to Lemma 3 of [3], we have
$$\begin{aligned} & \|T_1(\omega_1) - T_1(\omega_2)\| \geq \mu \|x_1 - x_2\| + \|\Pi_f(\lambda_1 - \lambda_2)_\perp\| \\ & \geq \mu \|x_1 - x_2\| + \text{eig}_2(\Pi_f) \|(\lambda_1 - \lambda_2)_\perp\| \\ & \geq \min \left\{ \mu, \frac{\text{eig}_2(\Pi_f) \|(\lambda_1 - \lambda_2)_\perp\|}{\|\lambda_1 - \lambda_2\|} \right\} (\|x_1 - x_2\| + \|\lambda_1 - \lambda_2\|) \\ & \geq \min \left\{ \mu, \frac{\text{eig}_2(\Pi_f) \|(\lambda_1 - \lambda_2)_\perp\|}{\|\lambda_1 - \lambda_2\|} \right\} \|\omega_1 - \omega_2\|, \end{aligned}$$

where $\Pi_f \triangleq (I_m - \frac{11^T}{m}) \otimes I_n$ and $\text{eig}_2(\Pi_f)$ denotes the second smallest eigenvalue of Π_f and is equal to 1. Because $(\lambda_1 - \lambda_2)_\perp \neq 0$, and $\|\lambda_1 - \lambda_2\|$ is always bounded, we can always find some k_1 such that for all $k \geq k_1$ (when α^k is sufficiently small), $\min \left\{ \mu, \frac{\text{eig}_2(\Pi_f) \|(\lambda_1 - \lambda_2)_\perp\|}{\|\lambda_1 - \lambda_2\|} \right\} \geq \frac{\alpha^k \|M^k\|}{\frac{\delta}{2} (1 - 2\alpha^k \|M^k\| - (\alpha^k)^2 \|M^k\|^2)}$ holds (hence (47) holds) under $(\lambda_1 - \lambda_2)_\perp \neq 0$.

In summary, (47) always holds for $k \geq k_1$, implying that for all $k \geq k_1$, $(I - \alpha^k M^k)T_1$ is $\frac{\delta}{2}$ -cocoercive in the Euclidean norm, and hence, $(\Phi^k)^{-1}T_1 = \alpha^k(I - \alpha^k M^k)T_1$ is $\frac{\delta}{2\alpha^k}$ -cocoercive in the Euclidean norm. Using Proposition 4.39 of [46], we have that $\text{Id} - (\Phi^k)^{-1}T_1$ is $\frac{\alpha^k}{\delta}$ -averaged in the Euclidean norm for all $k \geq k_1$.

Using a similar argument, we can prove that there exists a $k_2 \geq 0$ such that for all $k \geq k_2$, $(\Phi^k)^{-1}T_2$ is maximally monotone in the Euclidean norm since T_2 is maximally monotone (see Lemma 3 in [3]). Hence, the mapping $(\text{Id} - (\Phi^k)^{-1}T_2)^{-1}$ is $\frac{1}{2}$ -averaged in the Euclidean norm for all $k \geq k_2$.

Then, since for an a_1 -averaged operator T_a and an a_2 -averaged operator T_b , the composition $T_a \circ T_b$ is $\frac{a_1 + a_2 - 2a_1 a_2}{1 - a_1 a_2}$ -averaged (Proposition 4.44 of [46]), there exists a $k_0 = \max\{k_1, k_2\}$ such that for all $k \geq k_0$, R_k in (36) is $\frac{1}{2 - 2\alpha^k}$ -averaged, and hence, is nonexpansive (Remark 3.43 of [46]).

To prove that for every $\{\omega^k\}$, $\lim_{k \rightarrow \infty} \|R_k(\omega^k) - \omega^k\|$ exists almost surely, we use the expression of $(\Phi^k)^{-1}$ in (41). Given $\sum_{k=0}^{\infty} (\alpha^k)^2 < \infty$, it can be easily seen that $(\Phi^k)^{-1}$ will converge to the zero matrix, which, in turn, implies that R^k will converge to the identity operator. Proposition 1 ensures that ω^k is always bounded. Therefore, we always have that $\lim_{k \rightarrow \infty} \|R_k(\omega^k) - \omega^k\|$ exists almost surely for every ω^k . The fact that Condition 4) in Lemma 7 can be guaranteed by $\sum_{k=0}^{\infty} \alpha^k = \infty$ follows [5], [9]. ■

Remark 8. Different from [3] which proves that $\Phi^{-1}T_1$ (resp. $\Phi^{-1}T_2$) is cocoercive (resp. maximally monotone) in a Φ -induced norm (Φ is time-invariant therein), here we prove that $(\Phi^k)^{-1}T_1$ (resp. $(\Phi^k)^{-1}T_2$) is $\frac{\delta}{2\alpha^k}$ -cocoercive (resp. maximally monotone) in the Euclidean norm. The results are reminiscent to Theorem 5.2 of [65], which proves that transformations that are small perturbations of the identity matrix do not affect cocoercive properties. The difference is that Φ^k here is time-varying, and it is close to an α^k -scaled identity matrix rather than an identity matrix in [65].

B. Algorithm 2 is a stochastically-perturbed version of (35)

In this subsection, we prove that Algorithm 2 is a stochastically perturbed version of (35) with the perturbation $e(k)$

satisfying the conditions in Lemma 7. To this end, we first prove that the averages of σ_i^k , y_i^k , and z_i^k in Algorithm 2 always converge to \bar{x}^k , \bar{d}^k , and $\bar{\lambda}^k$, respectively:

Lemma 9. *Under Assumptions 4, 7, $\bar{\sigma}^k = \frac{\sum_{i=1}^m \sigma_i^k}{m}$, $\bar{y}^k = \frac{\sum_{i=1}^m y_i^k}{m}$, and $\bar{z}^k = \frac{\sum_{i=1}^m z_i^k}{m}$ in Algorithm 2 converge a.s. to \bar{x}^k , $\bar{d}^k = \frac{\sum_{i=1}^m d_i^k}{m}$ with d_i^k defined in (29), and $\bar{\lambda}^k$, respectively, if $\sum_{k=0}^{\infty} \gamma^k = \infty$ and $\sum_{k=0}^{\infty} (\gamma^k)^2 < \infty$ hold. Moreover, $\sum_{k=0}^{\infty} \alpha^k \|\bar{\sigma}^k - \bar{x}^k\| < \infty$, $\sum_{k=0}^{\infty} \alpha^k \|\bar{z}^k - \bar{\lambda}^k\| < \infty$, and $\sum_{k=0}^{\infty} \alpha^k \|\bar{y}^k - \bar{d}^k\| < \infty$ hold a.s. if $\sum_{k=0}^{\infty} \frac{(\alpha^k)^2}{\gamma^k} < \infty$ holds.*

Proof. The derivation of the first part follows the line of reasoning in Lemma 4, and the derivation of the second part follows the last paragraph of the proof of Theorem 1. Hence, we do not include the proof here. ■

To show that Algorithm 2 corresponds to a stochastically-perturbed version of (35), we rewrite the updates in Algorithm 2 in the following more compact form:

$$\begin{aligned}\tilde{x}^k &= \Pi_{\Omega} [x^k - \alpha^k (F(x^k, \sigma^k) + C_d^T z^k)], \\ \tilde{\lambda}^k &= \Pi_{\mathbb{R}^m} [\lambda^k + \alpha^k (y^k - \lambda^k + z^k)], \\ x^{k+1} &= x^k + \gamma^k (\tilde{x}^k - x^k), \\ \lambda^{k+1} &= \lambda^k + \gamma^k (\tilde{\lambda}^k - \lambda^k),\end{aligned}\quad (48)$$

with

$$\begin{aligned}y^{k+1} &= (1 - \gamma^k)y^k + \chi^k (L \otimes I_n)y^k + \chi^k (L^0 \otimes I_n)\xi^k \\ &\quad + (2C_d \tilde{x}^k - C_d x^k - c_d) \\ &\quad - (1 - \gamma^k)(2C_d \tilde{x}^{k-1} - C_d x^{k-1} - c_d),\end{aligned}\quad (49)$$

where $c_d \triangleq \text{col}(c_1, \dots, c_m)$.

Recalling that (39) corresponds to the nonstationary iteration process in (35), comparing (48) and (39) yields that Algorithm 2 corresponds to the following stochastically-perturbed nonstationary fixed-point iteration process

$$\omega^{k+1} = \omega^k + \gamma^k (R_k(\omega^k) + e^k - \omega^k), \quad (50)$$

with $e^k \triangleq \text{col}(e_x^k, e_{\lambda}^k)$ given by

$$\begin{aligned}e_x^k &= \Pi_{\Omega} [x^k - \alpha^k (F(x^k, \sigma^k) + C_d^T z^k)] \\ &\quad - \Pi_{\Omega} [x^k - \alpha^k (F(x^k, \bar{x}_1^k) + C_d^T \bar{\lambda}_1^k)], \\ e_{\lambda}^k &= \Pi_{\mathbb{R}^m} [\lambda^k + \alpha^k (y^k - \lambda^k + z^k)] \\ &\quad - \Pi_{\mathbb{R}^m} [\lambda^k + \alpha^k (\bar{d}_1^k - \lambda^k + \bar{\lambda}_1^k)].\end{aligned}\quad (51)$$

C. Combine preceding two subsections to prove the final result

In this subsection, we combine the preceding two subsections to prove that Algorithm 2 can ensure the almost sure convergence of all agents to the GNE.

Theorem 3. *Under Assumptions 1,2,3,4,7, when the nonnegative non-increasing sequence $\{\alpha^k\}$ satisfies*

$$\sum_{k=0}^{\infty} \alpha^k = \infty, \sum_{k=0}^{\infty} (\alpha^k)^2 < \infty, \quad (52)$$

and the nonnegative sequences $\{\gamma^k\}$ and $\{\chi^k\}$ satisfy

$$\sum_{k=0}^{\infty} \chi^k = \infty, \sum_{k=0}^{\infty} \frac{(\gamma^k)^2}{\chi^k} < \infty, \sum_{k=0}^{\infty} \frac{(\alpha^k)^2}{\gamma^k} < \infty, \quad (53)$$

then the iterates x_i^k for $i \in [m]$ in Algorithm 2 converge almost surely to a GNE x^ of (2).*

Proof. The basic idea is to leverage Lemma 7. To this end, we characterize $\|e^k\|$, which can be verified to satisfy

$$\|e^k\| \leq \|e_x^k\| + \|e_{\lambda}^k\|. \quad (54)$$

Using the nonexpansive property of the projection operator, we have the following relationship for $\|e_x^k\|$:

$$\begin{aligned}\|e_x^k\| &\leq \|x^k - \alpha^k (F(x^k, \sigma^k) + C_d^T z^k) \\ &\quad - (x^k - \alpha^k (F(x^k, \bar{x}_1^k) + C_d^T \bar{\lambda}_1^k))\| \\ &\leq \alpha^k \|F(x^k, \sigma^k) - F(x^k, \bar{x}_1^k)\| + \alpha^k \|C_d^T\| \|z^k - \bar{\lambda}_1^k\| \\ &= \alpha^k \|F(x^k, \sigma^k) - F(x^k, \bar{\sigma}_1^k) + F(x^k, \bar{\sigma}_1^k) - F(x^k, \bar{x}_1^k)\| \\ &\quad + \alpha^k \|C_d^T\| \|z^k - \bar{z}_1^k + \bar{z}_1^k - \bar{\lambda}_1^k\| \\ &\leq \alpha^k \tilde{L} \|\sigma^k - \bar{\sigma}_1^k\| + \alpha^k \sqrt{m} \tilde{L} \|\bar{\sigma}^k - \bar{x}^k\| \\ &\quad + \alpha^k \|C_d^T\| \|z^k - \bar{z}_1^k\| + \alpha^k \sqrt{m} \|C_d^T\| \|\bar{z}^k - \bar{\lambda}^k\|,\end{aligned}\quad (55)$$

where we have used the Lipschitz condition in Assumption 3 in the last inequality.

Lemma 9 implies $\sum_{k=0}^{\infty} \alpha^k \|\bar{\sigma}^k - \bar{x}^k\| < \infty$ and $\sum_{k=0}^{\infty} \alpha^k \|\bar{z}^k - \bar{\lambda}^k\| < \infty$ a.s. Hence, we only need to consider the first and third terms on the right hand side of (55).

According to (30), the evolutions of σ_i^k and z_i^k follow

$$\sigma_i^{k+1} = (1 - \gamma^k) \sigma_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij} (\sigma_j^k + \zeta_j^k - \sigma_i^k) + \gamma^k (\tilde{x}_i^k),$$

and

$$z_i^{k+1} = (1 - \gamma^k) z_i^k + \chi^k \sum_{j \in \mathbb{N}_i} L_{ij} (z_j^k + v_j^k - z_i^k) + \gamma^k (\tilde{\lambda}_i^k),$$

respectively. Therefore, given that x_i^k , \tilde{x}_i^k , \bar{x}_1^k , λ_i^k are bounded from Proposition 1, Theorem 1 implies $\sigma_i^k \rightarrow \bar{x}^k$ and $z_i^k \rightarrow \bar{\lambda}^k$ almost surely, and the following relations hold a.s.:

$$\sum_{k=0}^{\infty} \gamma^k \sum_{i=1}^m \|\sigma_i^k - \bar{\sigma}^k\| < \infty, \sum_{k=0}^{\infty} \gamma^k \sum_{i=1}^m \|z_i^k - \bar{z}^k\| < \infty. \quad (56)$$

Combining (55) and (56) implies the following relation a.s.:

$$\sum_{k=0}^{\infty} \gamma^k \|e_x^k\| < \infty. \quad (57)$$

Following the same line of derivation, we have

$$\begin{aligned}\|e_{\lambda}^k\| &\leq \|\lambda^k + \alpha^k (y^k - \lambda^k + z^k) \\ &\quad - (\lambda^k + \alpha^k (\bar{d}_1^k - \lambda^k + \bar{\lambda}_1^k))\| \\ &\leq \alpha^k \|y^k - \bar{y}_1^k\| + \alpha^k \|\sqrt{m}\| \|\bar{y}^k - \bar{d}^k\| + \alpha^k \|z^k - \bar{z}_1^k\| \\ &\quad + \alpha^k \|\sqrt{m}\| \|\bar{z}^k - \bar{\lambda}^k\|.\end{aligned}\quad (58)$$

Lemma 9 implies $\sum_{k=0}^{\infty} \alpha^k \|\bar{z}^k - \bar{\lambda}^k\| < \infty$ and $\sum_{k=0}^{\infty} \alpha^k \|\bar{y}^k - \bar{d}^k\| < \infty$ a.s. under $\sum_{k=0}^{\infty} \frac{(\alpha^k)^2}{\gamma^k} < \infty$. Hence, we only need to consider the first and third terms on the right hand side of (58).

According to Theorem 1, for the dynamics y_i^k in (49), if we can prove that $\|(2C_d \tilde{x}^k - C_d x^k - c_d) - (1 - \gamma^k)(2C_d \tilde{x}^{k-1} - C_d x^{k-1} - c_d)\| \leq \gamma^k C$ holds for some C , then we have $\sum_{k=0}^{\infty} \gamma^k \sum_{i=1}^m \|y_i^k - \bar{y}^k\| < \infty$ and hence, $\sum_{k=0}^{\infty} \alpha^k \sum_{i=1}^m \|y_i^k - \bar{y}^k\| < \infty$.

Noticing $(2C_d\tilde{x}^k - C_dx^k - c_d) - (1 - \gamma^k)(2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d) = 2C_d(\tilde{x}^k - \tilde{x}^{k-1}) - C_d(x^k - x^{k-1}) + \gamma^k(2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d)$, we have

$$\begin{aligned} & \|2C_d\tilde{x}^k - C_dx^k - c_d - (1 - \gamma^k)(2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d)\| \\ & \leq \|2C_d\|\|\tilde{x}^k - \tilde{x}^{k-1}\| + \|C_d\|\|x^k - x^{k-1}\| \\ & \quad + \gamma^k\|2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d\| \\ & \leq \|2C_d\|\|\tilde{x}^k - \tilde{x}^{k-1}\| + \gamma^k\|C_d\|\|\tilde{x}^{k-1} - x^{k-1}\| \\ & \quad + \gamma^k\|2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d\|, \end{aligned} \quad (59)$$

where we have used the update rule of x_i^k in (30) in the last inequality.

To characterize $\|\tilde{x}^k - \tilde{x}^{k-1}\|$, we employ the update rule of \tilde{x}_i^k in (30). Using the nonexpansive property of projection operators, we have

$$\begin{aligned} \|\tilde{x}^k - \tilde{x}^{k-1}\| &= \|\Pi_\Omega [x^k - \alpha^k (F(x^k, \sigma^k) + C_d^T z^k)] \\ &\quad - \Pi_\Omega [x^{k-1} - \alpha^{k-1} (F(x^{k-1}, \sigma^{k-1}) + C_d^T z^{k-1})]\| \\ &\leq \|x^k - \alpha^k (F(x^k, \sigma^k) + C_d^T z^k) \\ &\quad - (x^{k-1} - \alpha^{k-1} (F(x^{k-1}, \sigma^{k-1}) + C_d^T z^{k-1}))\| \\ &\leq \|x^k - x^{k-1}\| + \alpha^k \|F(x^k, \sigma^k) + C_d^T z^k\| \\ &\quad + \alpha^{k-1} \|F(x^{k-1}, \sigma^{k-1}) + C_d^T z^{k-1}\|. \end{aligned} \quad (60)$$

Using the proven results that $\sigma_i^k \rightarrow \bar{x}^k$ and $z_i^k \rightarrow \bar{\lambda}^k$ hold almost surely, and \bar{x}^k and $\bar{\lambda}^k$ are bounded (see Proposition 1), we have that z_i^k and σ_i^k are bounded almost surely. Therefore, the term $F(x^k, \sigma^k) + C_d^T z^k$ is bounded almost surely. Without loss of generality, we assume $\|F(x^k, \sigma^k) + C_d^T z^k\| \leq \bar{C}$ for all k almost surely.

Plugging the preceding relationship into (60) leads to

$$\begin{aligned} \|\tilde{x}^k - \tilde{x}^{k-1}\| &\leq \|x^k - x^{k-1}\| + (\alpha^k + \alpha^{k-1})\bar{C} \\ &\leq \gamma^k \|\tilde{x}^k - x^k\| + (\alpha^k + \alpha^{k-1})\bar{C}, \end{aligned} \quad (61)$$

where we have used the update rule of x_i^k in (30) in the last inequality. Given $\sum_{k=0}^{\infty} \frac{(\alpha^k)^2}{\gamma^k} < \infty$ in the theory statement, we have $(\alpha^k + \alpha^{k-1}) \leq \tilde{C}\gamma^k$ for some constant \tilde{C} , and hence $\|\tilde{x}^k - \tilde{x}^{k-1}\| \leq \gamma^k(\|\tilde{x}^k - x^k\| + \tilde{C})$ and further

$$\begin{aligned} & \|C_d(2\tilde{x}^k - x^k - c_d) - (1 - \gamma^k)C_d(2\tilde{x}^{k-1} - x^{k-1} - c_d)\| \\ & \leq \gamma^k\|C_d\|(\|\tilde{x}^{k-1} - x^{k-1}\| + 2\|\tilde{x}^k - x^k\| + 2\tilde{C}) \\ & \quad + \gamma^k\|2C_d\tilde{x}^{k-1} - C_dx^{k-1} - c_d\| \end{aligned} \quad (62)$$

by using (59).

Combining (49) and (62), we have that the evolution of y_i^k satisfies the conditions of Theorem 1, and hence, $\sum_{k=0}^{\infty} \gamma^k \sum_{i=1}^m \|y_i^k - \bar{y}^k\| < \infty$ holds almost surely. Further invoking (56) and (58) yields the following result *a.s.*:

$$\sum_{k=0}^{\infty} \gamma^k \|e_\lambda^k\| < \infty. \quad (63)$$

In summary, (54), (57), and (63) mean that $\sum_{k=0}^{\infty} \gamma^k \|e^k\| < \infty$ holds under the theorem statement almost surely. Following the arguments in [5], [9], under $\sum_{k=0}^{\infty} \alpha^k = \infty$, this implies that the Condition 4) in Lemma 7 will also be guaranteed almost surely for the perturbed algorithm 2. Moreover, the

condition $\sum_{k=0}^{\infty} \gamma^k \|e^k\| < \infty$ means $\sum_{k=0}^{\infty} \gamma^k \mathbb{E}[\|e^k\|] < \infty$ almost surely. Given $\sqrt{\mathbb{E}[\|e^k\|^2]} \leq \mathbb{E}[\|e^k\|]$ due to Jensen's inequality, invoking Lemma 7 yields that Algorithm 2 guarantees the convergence of all players to the GNE almost surely. ■

VI. PRIVACY ANALYSIS OF ALGORITHM 2

Similar to Definition 5, we define the sensitivity of a distributed GNE seeking algorithm to problem (2) as follows:

Definition 6. At each iteration k , for any initial state ϑ^0 and any adjacent distributed GNE problems \mathcal{P} and \mathcal{P}' , the sensitivity of a GNE seeking algorithm is

$$\Delta^k \triangleq \sup_{\Theta \in \mathcal{O}} \left\{ \sup_{\Theta \in \mathcal{R}_{\mathcal{P}, \vartheta^0}^{-1}(\mathcal{O}), \Theta' \in \mathcal{R}_{\mathcal{P}', \vartheta^0}^{-1}(\mathcal{O})} \|\Theta^k - \Theta'^k\|_1 \right\}, \quad (64)$$

where $\Theta^k = \text{col}(\sigma^k, y^k, z^k)$.

Then, similar to Lemma 6, we have the following lemma:

Lemma 10. In Algorithm 2, at each iteration k , if each player adds noise vectors ζ_i^k , v_i^k , and ξ_i^k , to its shared messages σ_i^k , z_i^k , and y_i^k , respectively, with every noise vector consisting of independent Laplace scalar noises with parameter ν^k , such that $\sum_{k=1}^{\bar{k}} \frac{\Delta^k}{\nu^k} \leq \bar{\epsilon}$, then Algorithm 2 is $\bar{\epsilon}$ -differentially private with the cumulative privacy budget for iterations from $k = 1$ to $k = \bar{k}$ less than $\bar{\epsilon}$.

Proof. The lemma can be obtained following the same line of reasoning of Lemma 2 in [55] (also see Theorem 3 in [36]). Please also see the explanations in the proof of Lemma 6. ■

Theorem 4. Under the conditions of Theorem 3, if all elements of ζ_i^k , ξ_i^k , and v_i^k are drawn independently from Laplace distribution $\text{Lap}(\nu^k)$ with $(\sigma_i^k)^2 = 2(\nu^k)^2$ satisfying Assumption 7, then all players in Algorithm 2 will converge almost surely to the GNE. Moreover,

- 1) For any finite number of iterations \bar{k} , Algorithm 1 is ϵ -differentially private with the cumulative privacy budget bounded by $\epsilon \leq \sum_{k=1}^{\bar{k}} \frac{(\zeta_y^k + \zeta_\sigma^k + \zeta_z^k)}{\nu^k}$ where $\zeta_y^k \triangleq C_y(\sum_{p=1}^{k-1} (\Pi_{q=p}^{k-1} (1 - \gamma^q - \bar{L}\chi^q)(2 - \gamma^{p-1})) + 2 - \gamma^{k-1})$, $\zeta_\sigma^k \triangleq C_\sigma(\sum_{p=1}^{k-1} (\Pi_{q=p}^{k-1} (1 - \gamma^q - \bar{L}\chi^q))\gamma^{p-1} + \gamma^{k-1})$, $\zeta_z^k \triangleq C_z(\sum_{p=1}^{k-1} (\Pi_{q=p}^{k-1} (1 - \gamma^q - \bar{L}\chi^q))\gamma^{p-1} + \gamma^{k-1})$, $\bar{L} \triangleq \min_i \{L_{ii}\}$, $C_y \triangleq \max_{i \in [m], 0 \leq k \leq \bar{k}} \|d_i^k - d_i^k\|_1$, $C_\sigma \triangleq \max_{i \in [m], 0 \leq k \leq \bar{k}} \|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1$, $C_z \triangleq \max_{i \in [m], 0 \leq k \leq \bar{k}} \|\tilde{\lambda}_i^k - (\tilde{\lambda}_i^k)'\|_1$;
- 2) The cumulative privacy budget is finite for $\bar{k} \rightarrow \infty$ when the sequence $\{\frac{\Delta^k}{\nu^k}\}$ is summable.

Proof. Since the Laplace noises satisfy Assumption 7, the convergence follows directly from Theorem 3.

According to the definition of sensitivity in Definition 6, we can obtain $\Delta^k = \Delta_\sigma^k + \Delta_y^k + \Delta_z^k$, where Δ_σ^k , Δ_y^k , and Δ_z^k are obtained by replacing Θ^k in (64) with σ^k , y^k , and z^k , respectively (note that the norm is L_1 norm).

Given two adjacent networked games \mathcal{P} and \mathcal{P}' , we represent the different cost functions as J_i in \mathcal{P} and J'_i in \mathcal{P}' without loss of generality.

Here, we only derive the result for Δ_σ^k , but Δ_y^k and Δ_z^k can be obtained using the same argument.

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

According to Algorithm 2, we can arrive at

$$\begin{aligned} \|\sigma_i^{k+1} - \sigma_i'^{k+1}\|_1 &\leq (1 - \gamma^k - |L_{ii}|\chi^k) \|\sigma_i^k - \sigma_i'^k\| \\ &\quad + \gamma^k \|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1, \end{aligned}$$

where we have used the relationship $x_i^{k+1} - (1 - \gamma^k)x_i^k = \gamma^k \tilde{x}_i^k$ and the fact that the observations $\sigma_j^k + \sigma_j^k$ and $\sigma_j'^k + \sigma_j'^k$ are the same.

Hence, Δ_σ^k satisfies $\Delta_\sigma^{k+1} \leq (1 - \gamma^k - |L_{ii}|\chi^k) \Delta_\sigma^k + \gamma^k \|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1$.

Using a similar line of argument, we can obtain the iteration relations for Δ_y^k and Δ_z^k , and hence, arrive at the first privacy statement by iteration.

For the infinite-horizon result in the second privacy statement, we exploit the fact that our algorithm ensures convergence of both \mathcal{P} and \mathcal{P}' to their respective GNE points, which are the same under the third requirement in Definition 1. This means that $\|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1 = 0$, $\|\tilde{\lambda}_i^k - (\tilde{\lambda}_i^k)'\|_1 = 0$, and $\|d_i^k - d_i'^k\|_1 = 0$ will hold when k is large enough. Furthermore, the ensured convergence also means that $\|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1$, $\|\tilde{\lambda}_i^k - (\tilde{\lambda}_i^k)'\|_1$, and $\|d_i^k - d_i'^k\|_1$ are always bounded. Hence, there always exists some constant C such that the sequences $\{\|\tilde{x}_i^k - (\tilde{x}_i^k)'\|_1\}$, $\{\|\tilde{\lambda}_i^k - (\tilde{\lambda}_i^k)'\|_1\}$, and $\{\|d_i^k - d_i'^k\|_1\}$ are upper bounded by the sequence $\{C\chi^k\gamma^k\}$.

Therefore, according to Lemma 3, there always exists a constant \bar{C} such that $\Delta^k \leq \bar{C}\gamma^k$ holds (note that $\{\gamma^k\}$ decays faster than $\{\chi^k\}$). Using Lemma 6, we can easily obtain $\epsilon \leq \sum_{k=1}^{\bar{k}} \frac{\bar{C}\gamma^k}{\nu^k}$. Hence, ϵ will be finite even when \bar{k} tends to infinity if the sequence $\{\frac{\gamma^k}{\nu^k}\}$ is summable, i.e., $\sum_{k=0}^{\infty} \frac{\gamma^k}{\nu^k} < \infty$. ■

Remark 9. Similar to the consensus-tracking case, to ensure that the cumulative privacy budget $\epsilon = \sum_{k=1}^{\infty} \frac{\bar{C}\gamma^k}{\nu^k}$ is bounded when $k \rightarrow \infty$, our algorithm uses Laplace noise with parameter ν^k that increases with time (since we require $\{\frac{\gamma^k}{\nu^k}\}$ to be summable while $\{\gamma^k\}$ is non-summable). In addition, according to Theorem 3, the convergence is not affected by scaling ν^k by any constant coefficient $\frac{1}{\epsilon} > 0$ to achieve any desired level of ϵ -DP, as long as the DP-noise parameter ν^k satisfies Assumption 7.

VII. NUMERICAL SIMULATIONS

We evaluate the performance of the proposed GNE seeking algorithm using a Nash-Cournot game recently considered in [1], [16], [66]. In the game, we consider m firms producing a homogeneous commodity competing over N markets, with a schematic presented in Fig. 1. In the schematic, we plot $N = 7$ markets (represented by M_1, \dots, M_7) and $m = 20$ firms (represented by circles). We use an edge from circle i to M_j to denote that firm i participates in market M_j .

We use $x_i \in \mathbb{R}^N$ to represent the amount of firm i 's products. If firm i does not participate in market j , then the j th entry

of x_i will be forced to be 0 all the time. Hence, if firm i participates in $1 \leq n_i \leq N$ markets, then its production vector x_i will have n_i non-zero entries. For notational simplicity, we use an adjacency matrix $B_i \in \mathbb{R}^{N \times N}$ to describe the association relationship between firm i and all markets. More specifically, B_i 's off-diagonal elements are all zero, and its j th diagonal entry is one if firm i participates in market j , otherwise, its j th diagonal entry is zero. Every firm i has a maximal capacity for each market j it participates in, which is represented by C_{ij} . Defining $\bar{C}_i \triangleq [C_{i1}, \dots, C_{iN}]^T$, the capacity constraint can be formulated as $x_i \leq \bar{C}_i$. Defining $B \triangleq [B_1, \dots, B_N]$, $Bx \in \mathbb{R}^N = \sum_{i=1}^N B_i x_i$ represents the total product supply to all markets, given firm i 's production amount x_i . We assume that every market has a maximum capacity \bar{c}_i . Defining $\bar{c} \triangleq [\bar{c}_1, \dots, \bar{c}_N]^T \in \mathbb{R}^N$, the shared coupling constraints can be formulated as $Bx \leq \bar{c}$.

As in [1], we let the commodity's price in every market M_i follow a linear inverse demand function: $p_i(x) = \bar{P}_i - \varsigma_i [Bx]_i$, where \bar{P}_i and $\varsigma_i > 0$ are constants and $[Bx]_i$ denotes the i th element of the vector Bx . It can be verified that the price value decreases when the amount of supplied commodity increases.

Representing the price vector of all markets as $p \triangleq [p_1, \dots, p_N]^T$, we have $p = \bar{P} - \Xi Bx$, where $\bar{P} \triangleq [\bar{P}_1, \dots, \bar{P}_N]^T$ and $\Xi \triangleq \text{diag}(\varsigma_1, \dots, \varsigma_N)$. 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 quadratic function $\phi_i(x_i) = x_i^T Q_i x_i + q_i^T x_i$, where $Q_i \in \mathbb{R}^{N \times N}$ is positive definite and $q_i \in \mathbb{R}^N$.

Firm i 's local cost function, which is determined by its production cost ϕ_i and payoff, is given by $J_i(x_i, x) = \phi_i(x_i) - (\bar{P} - \Xi Bx)^T B_i^T x_i$. One can verify that the gradient of the cost function is $F_i(x_i, x) = 2Q_i x_i + q_i + B_i^T \Xi B_i x_i - B_i(\bar{P} - \Xi Bx)$. It is clear that both firm i 's local cost function and gradient are dependent on other firms' actions.

In our simulation, we consider $m = 20$ firms competing over $N = 7$ markets. Since in the partial-decision information setting, each firm can only communicate with its immediate neighboring firms, we use a randomly generated local communication pattern given in Fig. 2. The maximal capacities for firm i (elements in \bar{C}_i) are randomly selected from the interval $[8, 10]$. The maximal capacities for all markets are set as $\kappa \sum_{i=1}^{20} \bar{C}_i$, with κ randomly selected from a uniform distribution on $(0, 1)$. Q_i in the production cost function is set as νI with ν randomly selected from $[1, 10]$. q_i in $\phi_i(x_i)$ is randomly selected from a uniform distribution in $[1, 2]$. In the price function, \bar{P}_i and ς_i are randomly chosen from uniform distributions in $[10, 20]$ and $[1, 3]$, respectively.

To evaluate the proposed Algorithm 2, for every firm i , we inject DP-noises ζ_i^k , ξ_i^k , and v_i^k in every shared σ_i^k , y_i^k , and z_i^k in all iterations. Each element of the noise vectors follows Laplace distribution with parameter $\nu^k = 1 + 0.1k^{0.2}$. We set the stepsizes and diminishing sequences as $\alpha^k = \frac{0.1}{1+0.1k}$, $\gamma^k = \frac{0.01}{1+0.1k^{0.98}}$ and $\chi^k = \frac{1}{1+0.1k^{0.9}}$, respectively, which satisfy the conditions in Theorem 3. In the evaluation, we run our algorithm for 100 times, and calculate the average and the variance of $\|x^k - x^*\|$ against 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 GNE seeking algorithm proposed by Belgioioso et al. in [3] under the same noise, and

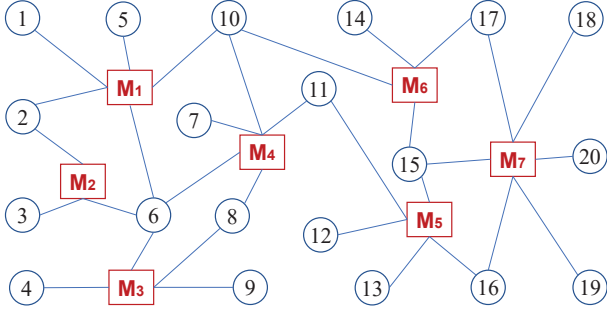


Fig. 1. Nash-Cournot game of 20 players (firms) competing over 7 locations (markets). Each firm is represented by a circular and each market is represented by a square. An edge between firm i ($1 \leq i \leq 20$) and market j ($1 \leq j \leq 7$) means that firm i participates in market j .

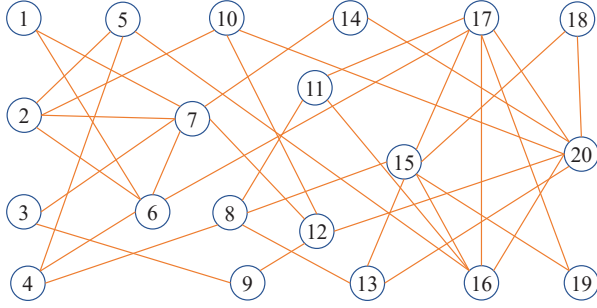


Fig. 2. The randomly generated interaction pattern of the 20 firms.

the existing DP approach for Nash equilibrium seeking proposed by Ye et al. in [36] under the same cumulative privacy budget ϵ . Note that the DP approach in [36] addresses Nash equilibrium seeking problems without shared coupling constraints. We adapt its DP mechanism (geometrically decreasing stepsizes and DP-noises) to the GNE seeking problem. The evolution of the average error/variance of the approaches in [3] and [36] are given by the blue and black curves/error bars in Fig. 3, respectively. It can be seen that the proposed algorithm has a much better accuracy.

VIII. CONCLUSIONS

This paper proposes a differentially-private fully distributed algorithm for generalized Nash equilibrium seeking. Different from existing privacy solutions for coupling-constraint free aggregative games, the proposed approach allows the existence of shared coupling constraints, which increase attack surfaces, and hence, pose additional challenges to privacy protection. More interestingly, the proposed approach can ensure both provable convergence to the generalized Nash equilibrium point and rigorous ϵ -differential privacy. As a basis of the differentially private generalized Nash equilibrium seeking algorithm, we also propose a new consensus-tracking algorithm that can ensure both provable convergence accuracy and rigorous ϵ -differential privacy. The convergence analysis generalizes existing results of stochastically-perturbed non-stationary fixed-point iteration processes to the diminishing-stepsize case, which is crucial to ensure a finite privacy budget,

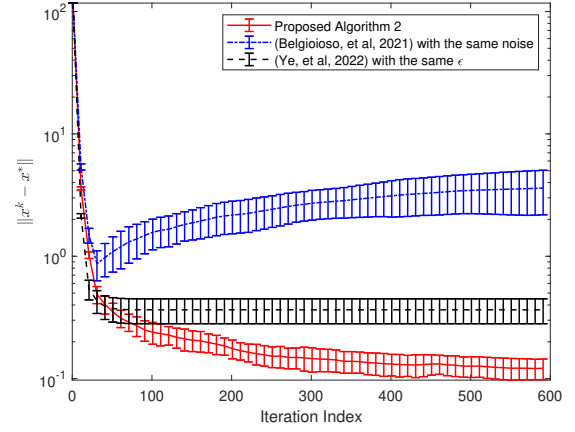


Fig. 3. Comparison of Algorithm 2 with the existing GNE seeking algorithm by Belgioioso et al. in [3] (under the same noise) and the differential-privacy approach from Ye et al. in [36] after adaptation to GNE seeking (under the same privacy budget ϵ).

and hence, rigorous differential privacy. Numerical simulation results confirm the effectiveness of the proposed algorithm.

REFERENCES

- [1] L. Pavel, "Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *IEEE Trans. Autom. Control*, vol. 65, no. 4, pp. 1584–1597, 2020.
- [2] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.
- [3] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2061–2075, 2021.
- [4] 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 Process. Mag.*, vol. 29, no. 5, pp. 86–105, 2012.
- [5] M. Zhu and E. Frazzoli, "Distributed robust adaptive equilibrium computation for generalized convex games," *Automatica*, vol. 63, pp. 82–91, 2016.
- [6] M. Ye and G. Hu, "Game design and analysis for price-based demand response: An aggregate game approach," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 720–730, 2016.
- [7] Y. Pan and L. Pavel, "Games with coupled propagated constraints in optical networks with multi-link topologies," *Automatica*, vol. 45, no. 4, pp. 871–880, 2009.
- [8] T. Alpcan and T. Başar, "Distributed algorithms for Nash equilibria of flow control games," in *Advances in Dynamic Games*. Springer, 2005, pp. 473–498.
- [9] H. Yin, U. V. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with scaled congestion costs and shared constraints," *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1702–1708, 2011.
- [10] W. Yong-Qiang, Y. Hao, X. S. Ding, and W. Gui-Zeng, "Fault detection of networked control systems based on optimal robust fault detection filter," *Acta Automatica Sinica*, vol. 34, no. 12, pp. 1534–1539, 2008.
- [11] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 1, pp. 67–78, 2011.
- [12] Z. Ma, S. Zou, L. Ran, X. Shi, and I. A. Hiskens, "Efficient decentralized coordination of large-scale plug-in electric vehicle charging," *Automatica*, vol. 69, pp. 35–47, 2016.
- [13] M. S. Stankovic, K. H. Johansson, and D. M. Stipanovic, "Distributed seeking of Nash equilibria with applications to mobile sensor networks," *IEEE Trans. Autom. Control*, vol. 57, no. 4, pp. 904–919, 2011.
- [14] Y. Wang, F. Núñez, and F. J. Doyle, "Statistical analysis of the pulse-coupled synchronization strategy for wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5193–5204, 2013.

- [15] Y. Wang and F. J. Doyle III, "On influences of global and local cues on the rate of synchronization of oscillator networks," *Automatica*, vol. 47, no. 6, pp. 1236–1242, 2011.
- [16] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Oper. Res.*, vol. 64, no. 3, pp. 680–704, 2016.
- [17] F. Salehisadaghiani and L. Pavel, "Distributed Nash equilibrium seeking: A gossip-based algorithm," *Automatica*, vol. 72, pp. 209–216, 2016.
- [18] T. Tatarenko, W. Shi, and A. Nedić, "Geometric convergence of gradient play algorithms for distributed Nash equilibrium seeking," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5342–5353, 2020.
- [19] D. Gadjov and L. Pavel, "A passivity-based approach to Nash equilibrium seeking over networks," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1077–1092, 2018.
- [20] M. Ye and G. Hu, "Distributed Nash equilibrium seeking by a consensus based approach," *IEEE Trans. Autom. Control*, vol. 62, no. 9, pp. 4811–4818, 2017.
- [21] F. Salehisadaghiani, W. Shi, and L. Pavel, "Distributed Nash equilibrium seeking under partial-decision information via the alternating direction method of multipliers," *Automatica*, vol. 103, pp. 27–35, 2019.
- [22] M. Bianchi, G. Belgioioso, and S. Grammatico, "Fast generalized Nash equilibrium seeking under partial-decision information," *Automatica*, vol. 136, p. 110080, 2022.
- [23] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.
- [24] K. Zhang, Z. Li, Y. Wang, A. Louati, and J. Chen, "Privacy-preserving dynamic average consensus via state decomposition: Case study on multi-robot formation control," *Automatica*, vol. 139, p. 110182, 2022.
- [25] 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.
- [26] Y. Wang and H. V. Poor, "Decentralized stochastic optimization with inherent privacy protection," *IEEE Transactions on Automatic Control*, vol. 68, no. 4, pp. 2293–2308, 2022.
- [27] S. Rassenti, S. S. Reynolds, V. L. Smith, and F. Szidarovszky, "Adaptation and convergence of behavior in repeated experimental cournot games," *J. Econ. Behav. Organ.*, vol. 41, no. 2, pp. 117–146, 2000.
- [28] R. Dong, W. Krichene, A. M. Bayen, and S. S. Sastry, "Differential privacy of populations in routing games," in *Proc. IEEE Conf. Decis. Control*. IEEE, 2015, pp. 2798–2803.
- [29] 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.
- [30] M. Shakarami, C. De Persis, and N. Monshizadeh, "Distributed dynamics for aggregative games: Robustness and privacy guarantees," *Int. J. Robust Nonlinear Control*, vol. 32, no. 9, pp. 5048–5069, 2022.
- [31] Y. Lu and M. Zhu, "Game-theoretic distributed control with information-theoretic security guarantees," *IFAC-PapersOnLine*, vol. 48, no. 22, pp. 264–269, 2015.
- [32] R. Cummings, M. Kearns, A. Roth, and Z. S. Wu, "Privacy and truthful equilibrium selection for aggregative games," in *Conf. Web Internet Econ.* Springer, 2015, pp. 286–299.
- [33] I. Shilov, H. Le Cadre, and A. Busic, "Privacy impact on generalized Nash equilibrium in peer-to-peer electricity market," *Oper. Res. Lett.*, vol. 49, no. 5, pp. 759–766, 2021.
- [34] 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.
- [35] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theo. Crypto. Conf.* Springer, 2006, pp. 265–284.
- [36] M. Ye, G. Hu, L. Xie, and S. Xu, "Differentially private distributed Nash equilibrium seeking for aggregative games," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2451–2458, 2021.
- [37] J. Wang, J.-F. Zhang, and X. He, "Differentially private distributed algorithms for stochastic aggregative games," *Automatica*, vol. 142, p. 110440, 2022.
- [38] Y. Wang and A. Nedić, "Differentially-private distributed algorithms for aggregative games with guaranteed convergence," *arXiv preprint arXiv:2209.01486*, 2022.
- [39] Y. Wang and T. Başar, "Ensuring both accurate convergence and differential privacy in Nash equilibrium seeking on directed graphs," *arXiv preprint arXiv:2209.04938*, 2022.
- [40] —, "Gradient-tracking based distributed optimization with guaranteed optimality under noisy information sharing," *IEEE Trans. Autom. Control*, 2022.
- [41] F. Parise, B. Gentile, and J. Lygeros, "A distributed algorithm for almost-Nash equilibria of average aggregative games with coupling constraints," *IEEE Trans. Control. Netw. Syst.*, vol. 7, no. 2, pp. 770–782, 2019.
- [42] D. Gadjov and L. Pavel, "Single-timescale distributed GNE seeking for aggregative games over networks via forward-backward operator splitting," *IEEE Trans. Autom. Control*, vol. 66, no. 7, pp. 3259–3266, 2020.
- [43] R. Zhu, J. Zhang, K. You, and T. Başar, "Asynchronous networked aggregative games," *Automatica*, vol. 136, p. 110054, 2022.
- [44] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.
- [45] J. Lei and U. V. Shanbhag, "Distributed variable sample-size gradient-response and best-response schemes for stochastic Nash equilibrium problems," *SIAM J. Optim.*, vol. 32, no. 2, pp. 573–603, 2022.
- [46] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2011, vol. 408.
- [47] N. Li, L. Chen, and M. A. Dahleh, "Demand response using linear supply function bidding," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1827–1838, 2015.
- [48] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Trans. Autom. Control*, vol. 60, no. 2, pp. 299–310, 2014.
- [49] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Nash and wardrop equilibria in aggregative games with coupling constraints," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1373–1388, 2018.
- [50] G. Belgioioso, P. Yi, S. Grammatico, and L. Pavel, "Distributed generalized Nash equilibrium seeking: An operator-theoretic perspective," *IEEE Control Syst.*, vol. 42, no. 4, pp. 87–102, 2022.
- [51] A. A. Kulkarni and U. V. Shanbhag, "On the variational equilibrium as a refinement of the generalized Nash equilibrium," *Automatica*, vol. 48, no. 1, pp. 45–55, 2012.
- [52] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.
- [53] Y. Wang and A. Nedić, "Tailoring gradient methods for differentially-private distributed optimization," *IEEE Trans. Autom. Control*, 2023.
- [54] C. Dwork, M. Naor, T. Pitassi, and G. N. Rothblum, "Differential privacy under continual observation," in *Proc. Annu. ACM Symp. Theory Comput.*, 2010, pp. 715–724.
- [55] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proc. Int. Conf. Distrib. Comput. Syst.*, NY, USA, 2015.
- [56] H. Gao and Y. Wang, "Algorithm-level confidentiality for average consensus on time-varying directed graphs," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 2, pp. 918–931, 2022.
- [57] H. Gao, Y. Wang, and A. Nedić, "Dynamics based privacy preservation in decentralized optimization," *Automatica*, vol. 151, p. 110878, 2023.
- [58] W. Chen, L. Liu, and G.-P. Liu, "Privacy-preserving distributed economic dispatch of microgrids: A dynamic quantization-based consensus scheme with homomorphic encryption," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 701–713, 2022.
- [59] K. Tjell and R. Wisniewski, "Privacy preservation in distributed optimization via dual decomposition and admm," in *Proc. IEEE Conf. Decis. Control*. IEEE, 2019, pp. 7203–7208.
- [60] D. Han, K. Liu, H. Sandberg, S. Chai, and Y. Xia, "Privacy-preserving dual averaging with arbitrary initial conditions for distributed optimization," *IEEE Trans. Autom. Control*, vol. 67, no. 6, pp. 3172–3179, 2021.
- [61] P. L. Combettes and J.-C. Pesquet, "Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping," *SIAM J. Optim.*, vol. 25, no. 2, pp. 1221–1248, 2015.
- [62] P. L. Combettes, "Quasi-Fejérian analysis of some optimization algorithms," in *Studies in Computational Mathematics*. Elsevier, 2001, vol. 8, pp. 115–152.
- [63] T.-H. Chang, A. Nedić, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1524–1538, 2014.
- [64] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*. SIAM, 2002.
- [65] D. Mateos-Núñez and J. Cortés, "Noise-to-state exponentially stable distributed convex optimization on weight-balanced digraphs," *SIAM J. Control Optim.*, vol. 54, no. 1, pp. 266–290, 2016.
- [66] 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.