Geometric Convergence of Gradient Play Algorithms for Distributed Nash Equilibrium Seeking

Tatiana Tatarenko, Wei Shi and Angelia Nedić

Abstract—We study distributed algorithms for seeking a Nash equilibrium in a class of convex **networked Nash games with strongly monotone mappings**. Each player has access to her own smooth local cost function and can communicate to her neighbors in some undirected graph. To deal with fast distributed learning of Nash equilibria under such settings, we introduce a so called augmented game mapping and provide conditions under which this mapping is strongly monotone. We consider a distributed gradient play algorithm for determining a Nash equilibrium (GRANE). The algorithm involves every player performing a gradient step to minimize her own cost function while sharing and retrieving information locally among her neighbors in the network. Using the reformulation of the Nash equilibrium problem based on the strong monotone augmented game mapping, we prove the convergence of this algorithm to a Nash equilibrium with a geometric rate. Further, we introduce the Nesterov type acceleration for the gradient play algorithm. We demonstrate that, similarly to the accelerated algorithms in centralized optimization and variational inequality problems, our accelerated algorithm outperforms GRANE in the convergence rate. Moreover, to relax assumptions required to guarantee the strongly monotone augmented mapping, we analyze the restricted strongly monotone property of this mapping and prove geometric convergence of the distributed gradient play under milder assumptions.

I. Introduction

Game theory provides a framework to deal with optimization problems arising in multi-agent systems, where agents' local objective functions are coupled through decision variables of all agents in a system. The applications of game theoretic approach include, for example, electricity markets, flow control problems, communication networks, and the future smart grids [1], [18], [21]. Nash equilibria in games characterize desirable and stable solutions to the corresponding multi-agent optimization problems. Due to emergence of the large-scale systems and agents' access only to some partial information in such systems, fast distributed optimization procedures have gained a lot of attention over the recent years. The focus of our paper is on discrete-time algorithms applicable to fast distributed Nash equilibrium seeking in a class of networked Nash games with strongly monotone mappings.

There is a large body of work on Nash equilibrium computation in games. Each approach is based on information available to agents in a systems and takes into account some structural properties of agents' cost functions. For example, in the case of the convex *potential game* structure, a central optimization problem can be formulated whose minimizers are Nash equilibria in the game. To compute the minima of the potential function, some distributed communication-based algorithms can be set up [8]. Moreover, in absence of communication, so called *oracle-based* algorithms can be applied to Nash equilibrium seeking in potential games. In systems with oracle information each agent can calculate her current output given any action from her action set. Various versions of the logit dynamics have been presented to compute Nash equilibria in such setting in both discrete action [10], [27] and continuous action [17], [28] games. In some practical situations agents do not have access to functional forms of their objectives. Rather, each agent (player) can only observe obtained payoffs and be aware of her own actions. Such information structure is usually referred to as *payoff-based*. A payoff-based Nash equilibrium learning is proposed in [11], [32] for discrete action potential games and in [5], [25], [29] for continuous action convex games. A convergent asynchronous best-response algorithm is presented in [22]. For implementation of this procedure agents need to collect the full information about the actions of their opponents at some instants of time.

In our work we deal with distributed setting, where each agent has access only to its own cost function and can get some partial information on the others' decisions by means of communication. Based on this information agents construct their estimations of the joint decision variables which they use in local optimization procedures to achieve a common goal, namely to find a Nash equilibrium state. Many works study convex networked Nash games where players can exchange local information on the decision variables with neighbors in some undirected connected communication graph. The papers [6], [31] provide continuous-time algorithms for Nash equilibrium seeking with exponential convergence rates. However, the theoretical tools used for the analysis in the works above cannot be applied to discrete-time setting. Distributed discrete-time algorithms are proposed for *aggregative games* [4], [7], [14]. Communication protocols are applied to general convex games with some convergence guarantees [19], [20], [26]. The work [19] studies distributed Nash equilibrium seeking of general networked Nash games and proposes a gradient based gossip algorithm. Under strict convexity, Lipschitz continuity and bounded gradient assumptions, this algorithm converges almost surely to the Nash equilibrium, given a diminishing step size. Under further assumption of strong convexity, with

some constant step size α, the algorithm converges to an $O(\alpha)$ neighborhood of the Nash equilibrium in average. The work [20] develops an algorithm within the framework of inexact-ADMM and proves its convergence to the Nash equilibrium with the rate $o(1/k)$ under cocoercivity of the game mapping. Communication-based algorithms for generalized Nash equilibrium problems are presented in papers [4], [16].

All aforementioned papers in discrete-time domain focus on convergence guarantees in different game settings. However, no distributed algorithm for Nash equilibrium seeking has been presented so far that provably possesses a geometric convergence rate. The contributions of this paper are as follows. To the best of our knowledge, we provide a novel approach for distributed Nash equilibrium seeking based on equivalent condition for Nash equilibrium states which uses so called *augmented game mapping*. The augmented game mapping takes into account not only partial derivatives of the cost functions but also agents' local information that is exchanged over a communication graph. Further, similarly to majority of research on distributed Nash equilibrium seeking, we consider undirected connected communication graphs and determine under which assumptions the augmented game mapping is strongly monotone and Lipschitz continuous. Next, we apply the standard algorithm to the extended variational inequality to obtain a distributed gradient play algorithm for determining a Nash equilibrium (GRANE). We demonstrate that GRANE converges to the Nash equilibrium with a geometric rate, given strongly monotone and Lipschitz continuous augmented game mapping. Moreover, using the results for strongly monotone variational inequalities presented in [13], we consider the Nesterov type acceleration for the gradient play algorithm. Similarly to the classical methods for solving variational inequalities, the corresponding accelerated gradient play algorithm (Acc-GRANE) outperforms GRANE in the convergence rate. However, the assumptions under which the augmented game mapping is strongly monotone restrict the class of games where Acc-GRANE can be applied. To rectify this issue we consider a relaxed assumption implying the restricted strongly monotone game mapping. Although Acc-GRANE is not applicable in this case, we demonstrate that GRANE converges to the Nash equilibrium with a geometric rate under this milder assumption. A preliminary and brief version of our work will appear in Conference on Decision and Control 2018 [30]. In the current paper, we extend our past work in the following ways. Firstly, we provide all proofs omitted in the conference paper. Secondly, we relax the assumption for the augmented mapping to be able to consider a broader class of games where geometric convergence to a Nash equivilibrium takes place. Finally, we discuss how communication topology affects the convergence rate of the distributed algorithms.

Notations. The set $\{1, \ldots, n\}$ is denoted by $[n]$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote a vector or a matrix with all entries equal to 0 and 1 respectively. For any function $f : K \to \mathbb{R}$, $K \subseteq \mathbb{R}^n$, $\nabla_i f(x) = \frac{\partial f(x)}{\partial x_i}$ is the partial derivative taken in respect to the ith coordinate of the vector variable $x \in \mathbb{R}^n$. For any real vector space $\tilde{E}$ its dual space is denoted by $\tilde{E}^*$ and the inner product is denoted by $\langle u, v \rangle$, $u \in \tilde{E}^*$, $v \in \tilde{E}$. Some

operator $B : \tilde{E} \to \tilde{E}^*$ is positive definite if $\langle Bv, v \rangle > 0$ for all $v \in \tilde{E} \setminus \{0\}$. Some operator $B : \tilde{E} \to \tilde{E}^*$ is self-adjoint if $\langle Bv, v' \rangle = \langle Bv', v \rangle$ for all $v', v \in \tilde{E}$. Given a positive definite and self-adjoint operator B, we define the Euclidean norm on $\tilde{E}$ as $\|v\| = \langle Bv, v \rangle^{1/2}$. Any mapping $g : \tilde{E} \to \tilde{E}^*$ is said to be *strongly monotone with the constant* $\mu > 0$ on $Q \subseteq \tilde{E}$, if $\langle g(u) - g(v), u - v \rangle \geq \mu \|u - v\|^2$ for any $u, v \in Q$, where $\|\cdot\|$ is the corresponding norm in $\tilde{E}$. We consider real vector space E, which is either space of real vectors $E = E^* = \mathbb{R}^n$ or the space of real matrices $E = E^* = \mathbb{R}^{n \times n}$. In the case $E = \mathbb{R}^{n \times n}$ the inner product $\langle u, v \rangle \triangleq \sqrt{\text{trace}(u^T v)}$ is the Frobenius inner product on $\mathbb{R}^{n \times n}$. In the case $E = \mathbb{R}^n$ we use $\|\cdot\|_2$ to denote the Euclidean norm induced by the standard dot product in $\mathbb{R}^n$, whereas in the case $E = \mathbb{R}^{n \times n}$ we use $\|\cdot\|_{\text{Fro}}$ to denote the Frobenius norm induced by the Frobenius inner product, i.e., $\|v\|_{\text{Fro}} \triangleq \sqrt{\text{trace}(v^T v)}$. We use $\mathcal{P}_\Omega \{v\}$ to denote the projection of $v \in E$ to a set $\Omega \subseteq E$. The largest singular value of a matrix A is denoted by $\sigma_{\max}\{A\}$. The smallest *nonzero* eigenvalue of a positive semidefinite matrix $A \neq \mathbf{0}$ is denoted by $\tilde{\lambda}_{\min}\{A\}$, which is positive. For a matrix $A \in \mathbb{R}^{m \times n}$, $\text{null}\{A\} \triangleq \{x \in \mathbb{R}^n | Ax = \mathbf{0}\}$ is the null space of A and $\text{span}\{A\} \triangleq \{y \in \mathbb{R}^m | y = Ax, \forall x \in \mathbb{R}^n\}$ is the linear span of all the columns of A. For any matrix $A \in \mathbb{R}^{n \times n}$ we use $\text{diag}(A)$ to denote its diagonal vector, i.e. $\text{diag}(A) = (a_{11}, \ldots, a_{nn})$. For any vector $a \in \mathbb{R}^n$ we use $\text{Diag}(a)$ to denote the diagonal matrix with the vector a on its diagonal. We call a matrix A *consensual*, if it has equal row vectors. Assuming that $\Omega \subseteq \mathbb{R}^n$, we define the indicator function $\mathcal{I}_\Omega(x)$ of the set Ω such that $\mathcal{I}_\Omega(x) = 0$, if $x \in \Omega$, and $\mathcal{I}_\Omega(x) = +\infty$, otherwise.

II. NASH EQUILIBRIUM SEEKING

A. Problem Formulation

We consider a networked Nash game between n players. Let J_i and $\Omega_i \subseteq \mathbb{R}$[1] denote respectively the cost function and the feasible action set of the player i. We denote the joint action set by $\Omega = \Omega_1 \times \ldots \times \Omega_n$. Each function $J_i(x_i, x_{-i})$, $i \in [n]$, depends on x_i and x_{-i}, where $x_i \in \Omega_i$ is the action of the player i and $x_{-i} \in \Omega_{-i} = \Omega_1 \times \ldots \times \Omega_{i-1} \times \Omega_{i+1} \times \Omega_n$ denotes the joint action of all players except for the player i. We assume that the players can interact over an undirected communication graph $\mathcal{G}([n], \mathcal{A})$, where the set of nodes represents the set of the player $[n]$ and the set of arcs $\mathcal{A}$ is such that $(i, j) \in \mathcal{A}$ if and only if $(j, i) \in \mathcal{A}$. Thus, some information (message) can be passed from the player i to the player j and vice versa. For each player i the set $\mathcal{N}_i$ is the set of neighbors in the graph $\mathcal{G}([n], \mathcal{A})$, namely $\mathcal{N}_i \triangleq \{j \in [n] : (i, j) \in \mathcal{A}\}$. We assume each player to be the neighbor for itself. Let us denote the game introduced above by $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$. We make the following assumptions regarding the game Γ.

Assumption 1. *The game under consideration is* convex. *Namely, for all $i \in [n]$, the set Ω_i is convex and closed, the cost function $J_i(x_i, x_{-i})$ is convex in x_i over Ω_i and continuously*

---

[1] All results below are applicable for games with different dimensions $\{d_i\}$ of the action sets $\{\Omega_i\}$. The one-dimensional case is considered for the sake of notation simplicity.

differentiable in x_i for each fixed x_{-i}. Moreover, the partial derivatives $\nabla_i J_i(x_i, x_{-i})$, $i \in [n]$, are continuous functions in the whole x.

Under Assumption 1, we can define the game mapping.

Definition 1. *The* game mapping $\mathbf{F}(x) : \Omega \to \mathbb{R}^n$ *is defined as follows:*

$$\mathbf{F}(x) \triangleq [\nabla_1 J_1(x_1, x_{-1}), \ldots, \nabla_n J_n(x_n, x_{-n})]^T. \quad (1)$$

Assumption 2. *The game mapping is* strongly monotone *on Ω with the constant $\mu_{\mathbf{F}}$. Moreover, each cost function $J_i(x_i, x_{-i})$, $i \in [n]$, is defined on the set $\mathbb{R} \times \ldots \times \Omega_i \times \ldots \times \mathbb{R}$ and is strongly convex in x_i over Ω_i for any fixed $x_{-i} \in \mathbb{R}^{n-1}$ with the constant $\mu_{\mathbf{F}}$.*

Remark 1. *Note that Assumption 2 above implies strong convexity of each cost function $J_i(x_i, x_{-i})$ in x_i over Ω_i for any fixed x_{-i} with the constant $\mu_{\mathbf{F}}$. Thus, the convexity condition in Assumption 1 holds. However, we consider Assumptions 1 and 2 separately, to be able to emphasize what conditions are required in each of the results presented further in the paper. An example of the cost functions satisfying Assumptions 1 and 2 can be found in the class of aggregative games (see [4], [7], [14]) with $J_i(x_i, x_{-i}) = c_i(x_i) + l_i(x_{-i})x_i$, where c_i is a strongly convex function over Ω_i and l_i is a linear function defined on the set $\mathbb{R}^{n-1}$.*

If the cost functions J_i, $i \in [n]$, are defined only on the set Ω, we assume that for each fixed $x_i \in \Omega_i$ they can be extended to the whole $\mathbb{R}^{n-1}$ and the strongly convex property in respect to x_i can be preserved.

Assumption 3. *For every $i \in [n]$ the function $\nabla_i J_i(x_i, x_{-i})$ is uniformly Lipschitz continuous in x_i on Ω_i, that is, there exists some constant $L_i \geq 0$ such that for any $x_{-i} \in \mathbb{R}^{n-1}$ we have for all $x_i, y_i \in \Omega_i$,*

$$|\nabla_i J_i(x_i, x_{-i}) - \nabla_i J_i(y_i, x_{-i})| \leq L_i |x_i - y_i|.$$

Moreover, for every $i \in [n]$ the function $\nabla_i J_i(x_i, x_{-i})$ is uniformly Lipschitz continuous in x_{-i} on $\mathbb{R}^{n-1}$, that is, there exists some constant $L_{-i} \geq 0$ such that for every fixed $x_i \in \Omega_i$ we have

$$|\nabla_i J_i(x_i, x_{-i}) - \nabla_i J_i(x_i, y_{-i})| \leq L_{-i} \|x_{-i} - y_{-i}\|_2,$$
$$\forall\, x_{-i}, y_{-i} \in \mathbb{R}^{n-1}.$$

We consider a symmetric non-negative mixing matrix $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ associated with the communication graph $\mathcal{G}([n], \mathcal{A})$. Its elements define the weights on the undirected arcs such that $w_{ij} = w_{ji} > 0$ if and only if $(i, j) \in \mathcal{A}$. Finally, we make the following assumption on the communication graph and the associated matrix, which guarantees sufficient and balanced information "mixing" in the network.

Assumption 4. *The underlying undirected communication graph $\mathcal{G}([n], \mathcal{A})$ is connected. Moreover, the associated matrix W is symmetric and column-stochastic, namely $\sum_{i=1}^{n} w_{ij} = 1$, $\forall j \in [n]$.*

Remark 2. *Note that there are some simple strategies for generating symmetric mixing matrices for which Assumption 4 holds (see Section 2.4 in [23] for a summary of strategies).*

Assumption 4 implies the following properties of the mixing matrix W (see [3]):

$$\sum_{i=1}^{n} w_{ij} = \sum_{j=1}^{n} w_{ij} = 1, \ \forall i, j \in [n] \quad (2a)$$

$$I - W \succeq \mathbf{0}; \quad (2b)$$

$$\text{null}\{I - W\} = \text{span}\{\mathbf{1}\}, \quad (2c)$$

One of the solution concepts for a game Γ is a Nash equilibrium defined below.

Definition 2. *A vector $x^* = [x_1^*, x_2^*, \cdots, x_n^*]^T \in \Omega$ is a* Nash equilibrium *if for any $i \in [n]$ and $x_i \in \Omega_i$*

$$J_i(x_i^*, x_{-i}^*) \leq J_i(x_i, x_{-i}^*).$$

In this work, we are interested in *distributed seeking of a Nash equilibrum* in a game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ for which Assumptions 1-4 hold.

B. Existence and Uniqueness of the Nash Equilibrium

In this subsection, we demonstrate the existence and uniqueness of the Nash equilibrium for $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ under Assumptions 1 and 2. For this purpose we recall the results connecting Nash equilibria and solutions of variational inequalities [15].

Definition 3. *For a set $Q \subseteq \mathbb{R}^d$ and a mapping $g: Q \to \mathbb{R}^d$ the* variational inequality *problem $VI(Q, g)$ is formulated as follows:*

$$\text{Find } q^* \in Q: \quad \langle g(q^*), q - q^* \rangle \geq 0 \text{ for all } q \in Q.$$

The set of solutions *to the variational inequality problem $VI(Q, g)$ is denoted by $SOL(Q, g)$.*

Remark 3. *One of the equivalent reformulation of the solution q^* in the definition above is the following one (see [15], Chapter 1.1):*

$$\text{Find } q^* \in Q: \quad \mathbf{0} \in g(q^*) + \mathcal{N}(q^*, Q),$$

where $\mathcal{N}(q^*, Q)$ is a normal cone to Q at q^*, which is defined as the set of the subgradients of the indicator function $\mathcal{I}_Q(q^*)$ for the set Q at the point q^*.

The following theorem is the well-known result on the connection between Nash equilibria in games and solutions of a definite variational inequality (see Proposition 1.4.2 in [15]).

Theorem 1. *Consider a networked Nash game Γ. Suppose that the action sets of the players $\{\Omega_i\}$ are closed and convex, the cost functions $\{J_i(x_i, x_{-i})\}$ are continuously differentiable and convex in x_i for every fixed x_{-i} from Ω_{-i}. Then, a vector $x^* \in \Omega$ is a Nash equilibrium for the Γ if and only if $x^* \in SOL(\Omega, \mathbf{F})$, where $\mathbf{F}$ is the game mapping defined by (1).*

The following result holds for variational inequalities with strongly monotone mappings (see Proposition 2.3.3 in [15]).

Theorem 2. *Given the $VI(Q, g)$, suppose that Q is a closed convex set and the mapping g is continuous and strongly monotone. Then, the solution set $SOL(Q, g)$ is nonempty and is a singleton.*

Taking into account Theorems 1 and 2, we obtain the following results.

Theorem 3. *Let $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ be a game for which Assumptions 1-2 hold. Then, there exists a unique Nash equilibrium in Γ. Moreover, the Nash equilibrium in Γ is the solution of $VI(\Omega, \mathbf{F})$, where $\mathbf{F}$ is the game mapping (see (1)).*

Finally, if Assumption 2 does not hold, but in addition to Assumption 1 all action sets are compact, then, Corollary 2.2.5 in [15] guarantees only existence of a Nash equilibrium in the game under consideration:

Theorem 4 (Corollary 2.2.5 in [15]). *Let $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ be a game for which Assumption 1 holds and Ω_i is a compact set for all $i \in [n]$. Then, there exists a Nash equilibrium in Γ. Moreover, any Nash equilibrium in Γ is the solution of $VI(\Omega, \mathbf{F})$, where $\mathbf{F}$ is the game mapping (see (1)).*

Thus, if Assumptions 1-2 hold, we can guarantee existence and uniqueness of the Nash equilibrium in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ under consideration. However, the reformulation of the Nash equilibrium in terms of the solution of the variational inequality $VI(\Omega, \mathbf{F})$ does not take into account the distributed setting of the problem presented in the previous subsection. Hence, to let the players learn the unique Nash equilibrium in a distributed way, we need to find an alternative condition which allows us to determine the Nash equilibrium in presence of partial information exchange over the communication graph $\mathcal{G}$ with the associated mixing matrix W. In the next section, we provide such a condition and present distributed algorithms for computing the Nash Equilibrium.

III. Equivalent Condition for Nash Equilibria and Augmented Game Mapping

A. Nash Equilibria in Distributed Settings

To deal with the partial information available to players and exchanged among them over the communication graph, we assume that each player i maintains a *local variable*

$$x_{(i)} = [\tilde{x}_{(i)1}, \cdots, \tilde{x}_{(i)i-1}, x_i, \tilde{x}_{(i)i+1}, \cdots, \tilde{x}_{(i)n}]^T \in \mathbb{R}^n, \quad (3)$$

which is her estimation of the joint action $x = [x_1, x_2, \cdots, x_n]^T \in \Omega$. Here $\tilde{x}_{(i)j} \in \mathbb{R}$ is the player i's estimate of x_j and $\tilde{x}_{(i)i} = x_i \in \Omega_i$. Also, we compactly denote the estimates of other players' actions by the player i as

$$\tilde{x}_{-i} = [\tilde{x}_{(i)1}, \cdots, \tilde{x}_{(i)i-1}, \tilde{x}_{(i)i+1}, \cdots, \tilde{x}_{(i)n}]^T \in \mathbb{R}^{n-1}, \quad (4)$$

and the estimates of the player j's action x_j by all players as

$$\tilde{x}_{(:)j} = [\tilde{x}_{(1)j}, \cdots, \tilde{x}_{(j-1)j}, x_j, \tilde{x}_{(j+1)j}, \cdots, \tilde{x}_{(n)j}]^T \in \mathbb{R}^n.$$

Thus, we can define the estimation matrix $\mathbf{x} \in \mathbb{R}^{n \times n}$, where the ith row is equal to the estimation vector $x_{(i)}$, $i \in [n]$, namely

$$\mathbf{x} \triangleq \begin{pmatrix} — & x_{(1)}^T & — \\ — & x_{(2)}^T & — \\ & \vdots & \\ — & x_{(n)}^T & — \end{pmatrix}.$$

Note that the estimation matrix belongs to a subset of the space $\mathbb{R}^{n \times n}$ that we denote by Ω_a. The set Ω_a consists of matrices whose diagonal vectors are from the set of joint action set Ω, i.e. $\Omega_a \triangleq \{\mathbf{x} \in \mathbb{R}^{n \times n} : \text{diag}(\mathbf{x}) \in \Omega\}$. Indeed, the definition of the domain Ω_a for estimation matrices $\mathbf{x}$ is based on the assumption that each agent i knows its own action set Ω_i, but not the action set Ω_j of other agents $j \in [n] \setminus i$. Finally, for any given estimation matrix, we define the diagonal matrix $\tilde{\mathbf{F}}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ with $\tilde{\mathbf{F}}(\mathbf{x})_{ii} = \nabla_i J_i(x_{(i)})$, $i \in [n]$, namely

$$\tilde{\mathbf{F}}(\mathbf{x}) \triangleq \text{Diag}(\nabla_1 J_1(x_{(1)}), \dots, \nabla_n J_n(x_{(n)})). \quad (5)$$

Now we are ready to state the proposition providing a necessary and sufficient condition for a joint action x^* to be a Nash equilibrium in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$.

Proposition 1. *Let us consider the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ for which Assumptions 1 and 4 hold. Then the following statements are equivalent*

1) *The vector $x^* = [x_1^*, \dots, x_n^*]^T$ is a Nash equilibrium in $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$.*
2) *There exists an estimation matrix $\mathbf{x}^* \in \Omega_a$ with the diagonal vector $x^* = [x_1^*, \dots, x_n^*]^T \in \Omega$ and corresponding diagonal matrix $\tilde{\mathbf{F}}(\mathbf{x}^*)$ (see (5)) such that for any $\mathbf{x} \in \Omega_a$ the following holds:*

$$\langle (I - W)\mathbf{x}^* + \Lambda \tilde{\mathbf{F}}(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0,$$

where $\Lambda = \text{Diag}(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^{n \times n}$ is an arbitrary diagonal matrix with $\alpha_i > 0$, $\forall i \in [n]$.

Proof. From the definition of Nash equilibrium (see Definition 2) and Assumption 1 we conclude that $x^* = [x_1^*, \dots, x_n^*]^T$ is a Nash equilibrium in $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ if and only if there exists the estimation vector $x_{(i)}^* \in \mathbb{R}^n, \forall i \in [n]$, such that

$$\langle \nabla_i J_i(x_{(i)}^*), x_i - x_i^* \rangle \geq 0, \forall x_i \in \Omega_i, \quad (6a)$$

$$x_{(1)}^* = x_{(2)}^* = \dots = x_{(n)}^*. \quad (6b)$$

Now we proceed with showing the equivalence (6) $\Leftrightarrow$ 2).

The implication (6) $\Rightarrow$ 2) holds, if we take $\mathbf{x}^*$ to be the matrix with each row equal to transpose of x^*. Indeed, in this case $(I - W)\mathbf{x}^* = 0$ (since $\sum_{i=1}^{n} w_{ij} = 1$, $\forall j \in [n]$, due to Assumption 4) and, according to (6), $\nabla_i J_i(x^*)(x_i - x_i^*) \geq 0$ for all $i \in [n]$.

To show 2) $\Rightarrow$ (6) we refer to Remark 3. According to this remark, the variational inequality in 2) implies existence of a subgradient $\tilde{\nabla} \mathcal{I}_{\Omega_i}(x_i^*)$ of the indicator function $\mathcal{I}_\Omega(x_i^*)$ of the set Ω_i at x_i^* for all $i \in [n]$ such that, given the diagonal matrix $\mathbf{G}(x^*) = \text{Diag}(\tilde{\nabla}\mathcal{I}_{\Omega_1}(x_1^*), \dots, \tilde{\nabla}\mathcal{I}_{\Omega_n}(x_n^*))$, the following equality holds:

$$(I - W)\mathbf{x}^* + \Lambda(\tilde{\mathbf{F}}(\mathbf{x}^*) + \mathbf{G}(x^*)) = \mathbf{0}. \quad (7)$$

Recall that the j-th column of $\mathbf{x}^*$ is denoted by $\tilde{x}^*_{(:)j}$. Thus, for any $j \in [n]$ the matrix equation (7) implies the following system of the linear equations

$$\begin{cases} [I - W]_{-j}\tilde{x}^*_{(:)j} = \mathbf{0}, \\ \langle [I - W]_j, \tilde{x}^*_{(:)j} \rangle + \alpha_j(\nabla_j J_j(x^*_{(j)}) + \mathcal{I}_{\Omega_j}(x^*_j)) = 0, \end{cases}$$
(8)

where $[I - W]_{-j} \in \mathbb{R}^{(n-1) \times n}$ denotes the matrix obtained from $I - W$ by discarding the jth row and $[I - W]_j \in \mathbb{R}^n$ denotes the jth row of the matrix $I - W$. By discarding the last equation in (8) we obtain the subsystem

$$[I - W]_{-j}\tilde{x}^*_{(:)j} = \mathbf{0},$$

which holds for all $j \in [n]$. According to (2c), the matrix $[I - W]_{-j}$ has full row rank for any $j \in [n]$. Thus, only vectors $\tilde{x}^*_{(:)j} \in \text{span}\{\mathbf{1}\}$, $j \in [n]$, are solutions to the systems of the linear equations above.

Therefore, $(I - W)\mathbf{x}^* = \mathbf{0}$ and (6b) holds. From (7) we have $\tilde{\mathbf{F}}(\mathbf{x}^*) + \mathbf{G}(\mathbf{x}^*) = \mathbf{0}$ and, thus, (6a) follows. ∎

Note that, regarding conditions for the communication graph, to prove the proposition above we used only assumption on stochastic rows of the matrix W and condition (2c), which can correspond to any directed strongly connected communication graph. Thus, the equivalent reformulation of Nash equilibria presented in Proposition 1 holds also for any convex game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$, where the graph $\mathcal{G}$ is possibly directed, strongly connected and the mixing matrix W is row stochastic, i.e. $\sum_{i=1}^n w_{ij} = 1, \forall j \in [n]$.

Moreover, in proof of Proposition 1 we did not use Assumption 2 requiring strongly monotone game mapping. However, we need this assumption in the next subsections, where we present distributed algorithms and prove their convergence to the Nash Equilibrium with a geometric rate.

B. Augmented Mapping

According to Proposition 1, to determine a Nash equilibrium in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ under consideration we need to find an estimation matrix $\mathbf{x}^*$ that solves the following variational inequality:

$$\langle (I - W)\mathbf{x}^* + \Lambda\tilde{\mathbf{F}}(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0, \quad \forall \mathbf{x} \in \Omega_a. \quad (9)$$

where $\Lambda \in \mathbb{R}^{n \times n}$ is an arbitrary diagonal matrix $\Lambda = \text{Diag}(\alpha_1, \ldots, \alpha_n)$, $\alpha_i > 0$, $\forall i \in [n]$. Due to Proposition 1, any solution matrix $\mathbf{x}^*$ of the variational inequality above is consensual and its diagonal vector, as well as any row, represents a Nash equilibrium in Γ. Further we call such estimation matrix $\mathbf{x}^*$ the *Nash equilibrium matrix*. Note that (9) is defined in terms of the matrix mapping $\mathbf{F}_a : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$:

$$\mathbf{F}_a(\mathbf{x}) = (I - W)\mathbf{x} + \Lambda\tilde{\mathbf{F}}(\mathbf{x}). \quad (10)$$

We refer to $\mathbf{F}_a(\mathbf{x})$ above as to the *augmented mapping* of the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$. In contrast to the game mapping $\tilde{\mathbf{F}}(\mathbf{x})$ (see (1)), the augmented mapping allows us to take into account the partial information exchange among the players over the graph $\mathcal{G}$. To set up efficient distributed

procedures with fast convergence rates, we leverage the results on centralized algorithms for classical variational inequalities with strongly monotone and Lipschitz continuous mappings obtained in [13]. The summary of these results and corresponding algorithms for distributed settings are presented in the next subsections. We conclude this subsection by formulating conditions under which the augmented mapping is Lipschitz continuous, strongly monotone, or restricted strongly monotone and, thus, the results from [13] can be applied to the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$.

Lemma 1. *Under Assumption 3, the augmented mapping* $\mathbf{F}_a(\mathbf{x})$ *of the game* $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ *is Lipschitz continuous on* Ω_a *with the Lipschitz constant* $L_{\Lambda\mathbf{F}} + \sigma_{\max}\{I - W\}$, *where* $L_{\Lambda\mathbf{F}} = \max_i \left\{ \alpha_i \sqrt{L_i^2 + L_{-i}^2} \right\}$.

Proof. By Assumption 3, we see that $\forall x, y \in \mathbb{R}^n$ such that $x_i, y_i \in \Omega_i$ and $x_{-i}, y_{-i} \in \mathbb{R}^{n-1}$, we have

$$\begin{aligned} &\|\nabla_i J_i(x_i, x_{-i}) - \nabla_i J_i(y_i, y_{-i})\|_2 \\ &= \|\nabla_i J_i(x_i, x_{-i}) - \nabla_i J_i(y_i, x_{-i}) \\ &\quad + \nabla_i J_i(y_i, x_{-i}) - \nabla_i J_i(y_i, y_{-i})\|_2 \\ &\leq (\beta\|\nabla_i J_i(x_i, x_{-i}) - \nabla_i J_i(y_i, x_{-i})\|_2^2 \\ &\quad + \frac{\beta}{\beta - 1}\|\nabla_i J_i(y_i, x_{-i}) - \nabla_i J_i(y_i, y_{-i})\|_2^2)^{\frac{1}{2}} \\ &\quad \text{(for any } \beta > 1) \\ &\leq \left(\beta L_i^2\|x_i - y_i\|_2^2 + \frac{\beta}{\beta - 1}L_{-i}^2\|x_{-i} - y_{-i}\|_2^2 \right)^{\frac{1}{2}} \\ &\quad \text{(choose } \beta = 1 + \frac{L_{-i}^2}{L_i^2}) \\ &= \sqrt{L_i^2 + L_{-i}^2}\|x - y\|_2 \triangleq L_{(i)}\|x - y\|_2, \end{aligned}$$

therefore,

$$\|\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})\|_{\text{Fro}} \leq L_{\mathbf{F}}\|\mathbf{x} - \mathbf{y}\|_{\text{Fro}}, \ \forall \ \mathbf{x}, \ \mathbf{y}, \quad (11)$$

where $L_{\mathbf{F}} = \max_i \left\{ L_{(i)} \right\}$.

Thus, we have

$$\begin{aligned} &\|\Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})) + (I - W)(\mathbf{x} - \mathbf{y})\|_{\text{Fro}} \\ &\quad \leq (L_{\Lambda\mathbf{F}} + \sigma_{\max}\{I - W\}) \|\mathbf{x} - \mathbf{y}\|_{\text{Fro}}, \ \forall \ \mathbf{x}, \ \mathbf{y}, \end{aligned}$$

where $L_{\Lambda\mathbf{F}} = \max_i \left\{ \alpha_i L_{(i)} \right\}$. ∎

Further on, we consider the case, where the diagonal matrix Λ has equal diagonal entries, i.e., $\Lambda = \text{Diag}(\alpha, \ldots, \alpha)$ with $\alpha > 0$. Next, given Assumptions 2–4, let us define $L^m = \max_i L_{-i}$ and

$$\begin{aligned} a_1 = {} & \tilde{\lambda}_{\min}\{I - W\} \\ & - \alpha \max \left\{ \mu_{\mathbf{F}} - L^m\sqrt{n - 1}, \frac{L^m}{2\sqrt{n - 1}} \right\}, \end{aligned} \quad (12a)$$

$$a_2 = \frac{\alpha}{2n} \left[\mu_{\mathbf{F}} - 2L^m\sqrt{n - 1} \right]. \quad (12b)$$

The following assumption formulates the condition on the parameters a_1, a_2 under which we will further prove the strongly monotone property of the augmented game mapping.

Assumption 5. *Assume that* $\min\{a_1, a_2\} > 0$.

With this assumption we can now prove the following result.

Lemma 2. *Under Assumptions 2, 3, 4, and 5, the augmented mapping $\mathbf{F}_a(\mathbf{x})$ of the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ is strongly monotone over Ω_a with the constant $\mu_{\mathbf{F}_a} = \min\{a_1, a_2\}$.*

Proof. Let $\mathbf{x}, \mathbf{y} \in \Omega_a$ be arbitrary but fixed matrices. By the definition of the mapping $\tilde{\mathbf{F}}$ in (5) and our choice of the matrix Λ (i.e., $\Lambda = \alpha I$), we have

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})), \mathbf{x} - \mathbf{y} \rangle$$
$$= \alpha \sum_{i=1}^{n} \langle \nabla_i J_i(x_i, \tilde{x}_{-i}) - \nabla_i J_i(y_i, \tilde{y}_{-i}), x_i - y_i \rangle$$
$$= \alpha \sum_{i=1}^{n} \langle \nabla_i J_i(x_i, \tilde{x}_{-i}) - \nabla_i J_i(y_i, \tilde{x}_{-i}), x_i - y_i \rangle$$
$$+ \alpha \sum_{i=1}^{n} \langle \nabla_i J_i(y_i, \tilde{x}_{-i}) - \nabla_i J_i(y_i, \tilde{y}_{-i}), x_i - y_i \rangle.$$

Using the strong monotonicity and Lipschitz gradient properties of functions J_i (see Assumptions 2 and 3), we obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})), \mathbf{x} - \mathbf{y} \rangle$$
$$\geq \alpha \sum_{i=1}^{n} \mu_{\mathbf{F}}(x_i - y_i)^2 - \alpha \sum_{i=1}^{n} L_{-i} \|\tilde{x}_{-i} - \tilde{y}_{-i}\|_2 \cdot |x_i - y_i|$$
$$\geq \alpha \sum_{i=1}^{n} \mu_{\mathbf{F}}(x_i - y_i)^2$$
$$- \alpha \sum_{i=1}^{n} \left(\frac{L_{-i}\beta_i}{2} \|\tilde{x}_{-i} - \tilde{y}_{-i}\|_2^2 + \frac{L_{-i}}{2\beta_i}(x_i - y_i)^2 \right),$$

where $\beta_i > 0$ are arbitrary for all $i \in [n]$. Let $\mathbf{z} = \mathbf{x} - \mathbf{y}$, and consider the following decomposition of the matrix $\mathbf{z}$: $\mathbf{z} = \mathbf{c} + \mathbf{n}$, where $\mathbf{c} = \frac{1}{n}\mathbf{1}\mathbf{1}^T\mathbf{z}$, $\mathbf{n} = \mathbf{z} - \mathbf{c}$. The matrix $\mathbf{c}$ has identical rows, where each row is equal to the average row entries of $\mathbf{z}$. Moreover, we have $\mathbf{c}'\mathbf{n} = \mathbf{0}$, implying that $\langle \mathbf{c}, \mathbf{n} \rangle = 0$. Therefore,

$$\|\mathbf{z}\|_{\text{Fro}}^2 = \|\mathbf{c}\|_{\text{Fro}}^2 + \|\mathbf{n}\|_{\text{Fro}}^2. \tag{13}$$

Using the same notation c_i, n_i and $\tilde{c}_{-i}, \tilde{n}_{-i}$ for the matrices $\mathbf{c}, \mathbf{n}$ and taking into account that $x_i - y_i = c_i + n_i$ and $\tilde{x}_{-i} - \tilde{y}_{-i} = \tilde{c}_{-i} + \tilde{n}_{-i}$, we further obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})), \mathbf{x} - \mathbf{y} \rangle$$
$$\geq \alpha \left[\mu_{\mathbf{F}} - \max_i \frac{L_{-i}}{2\beta_i} \right] \sum_{i=1}^{n}(c_i + n_i)^2$$
$$- \alpha \sum_{i=1}^{n} \frac{L_{-i}\beta_i}{2} \|\tilde{c}_{-i} + \tilde{n}_{-i}\|_2^2$$
$$\geq \alpha \left[\mu_{\mathbf{F}} - \max_i \frac{L_{-i}}{2\beta_i} \right] \sum_{i=1}^{n}(c_i + n_i)^2$$
$$- \alpha \max_i L_{-i}\beta_i \sum_{i=1}^{n}(\|\tilde{c}_{-i}\|_2^2 + \|\tilde{n}_{-i}\|_2^2),$$

where in the last inequality we use the relation $-\|\tilde{c}_{-i} + \tilde{n}_{-i}\|_2^2 \geq -2(\|\tilde{c}_{-i}\|_2^2 + \|\tilde{n}_{-i}\|_2^2)$. Let $\beta_i = \frac{1}{2\sqrt{n-1}}$ for all i and note that, in view of Assumption 5, it follows that $\mu_{\mathbf{F}} -$

$\max_i \frac{L_{-i}}{2\beta_i} > 0$. Thus, using the relation $(c_i + n_i)^2 \geq \frac{1}{2}c_i^2 - n_i^2$ and the notation $L^m = \max_i L_{-i}$, we obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})), \mathbf{x} - \mathbf{y} \rangle$$
$$\geq \alpha \left[\mu_{\mathbf{F}} - L^m\sqrt{n-1} \right] \frac{1}{2} \sum_{i=1}^{n} c_i^2 - \alpha L^m \frac{1}{2\sqrt{n-1}} \sum_{i=1}^{n} \|\tilde{c}_{-i}\|_2^2$$
$$- \alpha \left[\mu_{\mathbf{F}} - L^m\sqrt{n-1} \right] \sum_{i=1}^{n} n_i^2 - \alpha L^m \frac{1}{2\sqrt{n-1}} \sum_{i=1}^{n} \|\tilde{n}_{-i}\|_2^2,$$

Next, due to the fact that $n\sum_{i=1}^{n} c_i^2 = \|\mathbf{c}\|_{\text{Fro}}^2$ and $\sum_{i=1}^{n} \|\tilde{c}_{-i}\|_2^2 = (n-1)\sum_{i=1}^{n} c_i^2$, we obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{y})), \mathbf{x} - \mathbf{y} \rangle$$
$$\geq \frac{\alpha}{2n} \left[\mu_{\mathbf{F}} - 2L^m\sqrt{n-1} \right] \|\mathbf{c}\|_{\text{Fro}}^2$$
$$- \alpha \max \left\{ \mu_{\mathbf{F}} - L^m\sqrt{n-1}, \frac{L^m}{2\sqrt{n-1}} \right\} \|\mathbf{n}\|_{\text{Fro}}^2.$$

By Assumption 4, we have that $(I - W)\mathbf{c} = \mathbf{0}$, implying that

$$\langle (I - W)\mathbf{z}, \mathbf{z} \rangle \geq \tilde{\lambda}_{\min}\{I - W\}\|\mathbf{n}\|_{\text{Fro}}^2.$$

Thus, it follows that

$$\langle \mathbf{F}_a(\mathbf{x}) - \mathbf{F}_a(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$
$$\geq \frac{\alpha}{2n} \left[\mu_{\mathbf{F}} - 2L^m\sqrt{n-1} \right] \|\mathbf{c}\|_{\text{Fro}}^2$$
$$+ \left(\tilde{\lambda}_{\min}\{I - W\} - \alpha \max \left\{ \mu_{\mathbf{F}} - L^m\sqrt{n-1}, \frac{L^m}{2\sqrt{n-1}} \right\} \right)$$
$$\times \|\mathbf{n}\|_{\text{Fro}}^2$$
$$\geq \min\{a_1, a_2\}(\|\mathbf{c}\|_{\text{Fro}}^2 + \|\mathbf{n}\|_{\text{Fro}}^2) = \min\{a_1, a_2\}\|\mathbf{z}\|_{\text{Fro}}^2$$
$$= \mu_{\mathbf{F}_a}\|\mathbf{x} - \mathbf{y}\|_{\text{Fro}}^2,$$

where the first equality is due to (13). ∎

Remark 4. *Note that if the second part of Assumption 2 does not hold and one only assumes strongly monotone game mapping $\mathbf{F}$ over Ω, then the arguments used in the proof of Lemma 2 would imply that the augmented mapping $\mathbf{F}_a$ is strongly monotone over the matrix set Ω_a', which is defined by matrices whose rows belong to the joint action set Ω.*

Remark 5. *Lemma 2 holds for games with $\mu_{\mathbf{F}} > 2\sqrt{n-1}L_{-i}$ for all i. Intuitively speaking, this means that, the strong monotonicity of the game mapping should be strong enough while for each player i, its local objective's dependency on other player's strategies should be less than quadratically or have small enough quadratic terms. To make this easier to imagine, let us consider a class of game with objectives in the form of $J_i(x_i, x_{-i}) = f_i(x_i) + l_i(x_{-i})x_i$, where f_i's are twice differentiable and strongly convex while $l_i(x_{-i})$'s are linear in x_{-i}. For this special case, the Jacobian of the game mapping $\mathbf{F}(\mathbf{x})$ is always positive definite as long as the strong convexity constant of $f_i(x_i)$'s are large enough because the Jacobian's off-diagonal entries are all bounded. Meanwhile, in this case, the Lipschitz constant L_{-i} is any number that satisfies $|l_i(x_{-i}) - l_i(y_{-i})| \leq L_{-i}\|x_{-i} - y_{-i}\|_2$ for all $x_{-i}, y_{-i} \in \mathbb{R}^{n-1}$. This is fulfilled by any Lipschitz*

continuous function and linear function is obviously one of the members.

To rectify the issue pointed out in Remark 5 and, thus, to broaden the class of games under consideration, let us relax Assumption 2. Namely, instead of the assumption on strongly monotone game mapping $\mathbf{F}$, let us introduce the following assumption of *restricted strongly monotone* game mapping.

Assumption 6. *The game mapping $\mathbf{F}$ is restricted strongly monotone over $\mathbb{R}^n$ in respect to any Nash equilibrium $x^* \in \Omega$ with the constant $\mu_r > 0$, namely for any $x \in \mathbb{R}^n$*

$$\langle \mathbf{F}(x) - \mathbf{F}(x^*), x - x^* \rangle \geq \mu_r \|x - x^*\|_2^2.$$

Next, we show that under Assumption 6 there exist settings such that the resulting augmented mapping $\mathbf{F}_a$ is restricted strongly monotone with respect to the Nash equilibrium matrix $\mathbf{x}^*$. In the case when Assumption 6 holds, to set up the augmented mapping $\mathbf{F}_a(\mathbf{x})$ we use the diagonal matrix Λ such that $\alpha_i = \alpha$ for all $i \in [n]$.

Lemma 3. *Under Assumptions 3, 4, and 6, there exist settings with $\Lambda = \mathrm{Diag}(\alpha, \ldots, \alpha)$, $\alpha > 0$, such that the the resulting augmented mapping $\mathbf{F}_a(\mathbf{x})$ of the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ is restricted strongly monotone over Ω_a with respect to any Nash equilibrium matrix $\mathbf{x}^* = \mathbf{1}(x^*)^{\mathrm{T}}$ with the constant*

$$\mu_{r,\mathbf{F}_a} = \min\{b_1, b_2\},$$
$$b_1 = \alpha \left(\frac{\mu_r}{n} - L_{\mathbf{F}}(\beta^2 + 2\beta) \right),$$
$$b_2 = \frac{\tilde{\lambda}_{\min}\{I - W\}}{1 + \frac{1}{\beta^2}} - \alpha L_{\mathbf{F}},$$

where $L_{\mathbf{F}} = \alpha \max_i \left\{ \sqrt{L_i^2 + L_{-i}^2} \right\}$ and $\beta > 0$ is such that $\min\{b_1, b_2\} > 0$.

Proof. Analogously to the proof of Lemma 2, for any matrix $\mathbf{x} \in \Omega_a$, we consider the following orthogonal decomposition $\mathbf{x} = \mathbf{c} + \mathbf{n}$, where $\mathbf{c} = \frac{1}{n}\mathbf{1}\mathbf{1}^{\mathrm{T}}\mathbf{x}$, and $\mathbf{n} = \mathbf{x} - \mathbf{c}$. Thus, $\langle \mathbf{c}, \mathbf{n} \rangle = 0$. Since $\mathbf{x}^* = \mathbf{1}(x^*)^{\mathrm{T}}$, it can be seen that $\langle \mathbf{c} - \mathbf{x}^*, \mathbf{n} \rangle = 0$. Thus, we get

$$\|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 = \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 + \|\mathbf{n}\|_{\mathrm{Fro}}^2. \quad (14)$$

Let $c = \mathrm{diag}(\mathbf{c}) \in \mathbb{R}^n$ be the vector in rows of the matrix $\mathbf{c}$. Due to Assumption 6 and given $\alpha_i = \alpha$ for all $i \in [n]$, we obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{c}) - \tilde{\mathbf{F}}(\mathbf{x}^*)), \mathbf{c} - \mathbf{x}^* \rangle$$
$$= \sum_{i=1}^n \alpha_i (\nabla_i J_i(c) - \nabla_i J_i(x^*))(c_i - x_i^*)$$
$$\geq \alpha \langle \mathbf{F}(c) - \mathbf{F}(x^*), c - x^* \rangle$$
$$\geq \alpha \mu_r \|c - x^*\|_2^2 = \alpha \frac{\mu_r}{n} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2. \quad (15)$$

Since $\tilde{\mathbf{F}}$ is Lipschitz continuous on Ω_a with the constant $L_{\mathbf{F}} = \max_i \left\{ \alpha \sqrt{L_i^2 + L_{-i}^2} \right\}$ (see equation (11) in the proof

of Lemma 1), we conclude that

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{c}) - \tilde{\mathbf{F}}(\mathbf{x}^*)), \mathbf{x} - \mathbf{c} \rangle \geq -\alpha L_{\mathbf{F}} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}} \|\mathbf{x} - \mathbf{c}\|_{\mathrm{Fro}}$$
$$= -\alpha L_{\mathbf{F}} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}} \|\mathbf{n}\|_{\mathrm{Fro}}, \quad (16a)$$

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{c})), \mathbf{c} - \mathbf{x}^* \rangle \geq -\alpha L_{\mathbf{F}} \|\mathbf{x} - \mathbf{c}\|_{\mathrm{Fro}} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}, \quad (16b)$$

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{c})), \mathbf{x} - \mathbf{c} \rangle \geq -\alpha L_{\mathbf{F}} \|\mathbf{x} - \mathbf{c}\|_{\mathrm{Fro}}^2$$
$$= -\alpha L_{\mathbf{F}} \|\mathbf{n}\|_{\mathrm{Fro}}^2. \quad (16c)$$

Next, taking into account inequalities (15) and (16), we obtain

$$\langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{x}^*)), \mathbf{x} - \mathbf{x}^* \rangle = \langle \Lambda(\tilde{\mathbf{F}}(\mathbf{c}) - \tilde{\mathbf{F}}(\mathbf{x}^*)), \mathbf{c} - \mathbf{x}^* \rangle$$
$$+ \langle \Lambda(\tilde{\mathbf{F}}(\mathbf{c}) - \tilde{\mathbf{F}}(\mathbf{x}^*)), \mathbf{x} - \mathbf{c} \rangle + \langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{c})), \mathbf{c} - \mathbf{x}^* \rangle$$
$$+ \langle \Lambda(\tilde{\mathbf{F}}(\mathbf{x}) - \tilde{\mathbf{F}}(\mathbf{c})), \mathbf{x} - \mathbf{c} \rangle$$
$$\geq \alpha \left(\frac{\mu_r}{n} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 - L_{\mathbf{F}} \|\mathbf{n}\|_{\mathrm{Fro}}(2\|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}} + \|\mathbf{n}\|_{\mathrm{Fro}}) \right). \quad (17)$$

Moreover, due to (14),

$$\langle (I - W)(\mathbf{x} - \mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq \tilde{\lambda}_{\min}\{I - W\} \|\mathbf{n}\|_{\mathrm{Fro}}^2, \quad (18)$$

since $(I - W)(\mathbf{c} - \mathbf{x}^*) = \mathbf{0}$ (see (2c)). Summing inequalities (17) and (18), we conclude that

$$\langle \mathbf{F}_a(\mathbf{x}) - \mathbf{F}_a(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq \tilde{\lambda}_{\min}\{I - W\} \|\mathbf{n}\|_{\mathrm{Fro}}^2$$
$$+ \alpha \left(\frac{\mu_r}{n} \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 - L_{\mathbf{F}} \|\mathbf{n}\|_{\mathrm{Fro}}(2\|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}} + \|\mathbf{n}\|_{\mathrm{Fro}}) \right). \quad (19)$$

Let us fix some positive $\beta > 0$.

(i) If $\|\mathbf{n}\|_{\mathrm{Fro}} \leq \beta \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}$, then, according to (19) and (14),

$$\langle \mathbf{F}_a(\mathbf{x}) - \mathbf{F}_a(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq \tilde{\lambda}_{\min}\{I - W\} \|\mathbf{n}\|_{\mathrm{Fro}}^2$$
$$+ \alpha \left(\frac{\mu_r}{n} - L_{\mathbf{F}}(\beta^2 + 2\beta) \right) \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2$$
$$\geq b_1 \|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2, \quad (20)$$

where

$$b_1 = \min \left\{ \alpha \left(\frac{\mu_r}{n} - L_{\mathbf{F}}(\beta^2 + 2\beta) \right), \tilde{\lambda}_{\min}\{I - W\} \right\}.$$

(ii) If $\|\mathbf{n}\|_{\mathrm{Fro}} \geq \beta \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}$, then, according to (19) and (14) and since $\tilde{\mathbf{F}}$ is Lipschitz continuous with the constant $L_{\mathbf{F}}$, we obtain

$$\langle \mathbf{F}_a(\mathbf{x}) - \mathbf{F}_a(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq \tilde{\lambda}_{\min}\{I - W\} \|\mathbf{n}\|_{\mathrm{Fro}}^2$$
$$- \alpha L_{\mathbf{F}} \|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2$$
$$= -\alpha L_{\mathbf{F}} \|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2$$
$$+ \tilde{\lambda}_{\min}\{I - W\} \left(\frac{\beta^2 \|\mathbf{n}\|_{\mathrm{Fro}}^2}{1 + \beta^2} + \frac{\|\mathbf{n}\|_{\mathrm{Fro}}^2}{1 + \beta^2} \right)$$
$$\geq -\alpha L_{\mathbf{F}} \|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2$$
$$+ \tilde{\lambda}_{\min}\{I - W\} \left(\frac{\beta^2 \|\mathbf{n}\|_{\mathrm{Fro}}^2}{1 + \beta^2} + \frac{\beta^2 \|\mathbf{c} - \mathbf{x}^*\|_{\mathrm{Fro}}^2}{1 + \beta^2} \right)$$
$$= b_2 \|\mathbf{x} - \mathbf{x}^*\|_{\mathrm{Fro}}^2, \quad (21)$$

where $b_2 = \frac{\tilde{\lambda}_{\min}\{I - W\}}{1 + \frac{1}{\beta^2}} - \alpha L_{\mathbf{F}}$.

Thus, combining (20) and (21), we conclude that $\mu_{r,\mathbf{F}_a} = \min\{b_1, b_2\}$. We proceed with showing existence of settings under which $\mu_{r,\mathbf{F}_a} > 0$. Indeed, we can choose β such that $b_1 = \alpha\left(\frac{\mu_r}{n} - L_{\mathbf{F}}(\beta^2 + 2\beta)\right)$ is positive. It holds, for example, if $\beta^2 + 2\beta = \frac{\mu_r}{2nL_{\mathbf{F}}}$. Note that as $\frac{\mu_r}{2nL_{\mathbf{F}}} > 0$, there exists some positive $\beta > 0$ solving the equation above. Further, based on the choice of $\beta > 0$, we can proceed with settings for $\alpha > 0$. With these settings we need to guarantee that $b_2 = \frac{\tilde{\lambda}_{\min}\{I-W\}}{1 + \frac{1}{\beta^2}} - \alpha L_{\mathbf{F}} > 0$. It holds, for example, if $\alpha = \frac{\tilde{\lambda}_{\min}\{I-W\}}{2L_{\mathbf{F}}(1 + \frac{1}{\beta^2})}$. ∎

IV. GRADIENT ALGORITHMS WITH GEOMETRIC RATES

A. Gradient and Accelerated Gradient Algorithms for Strongly Monotone Variational Inequalities

In this subsection we summarize the results presented in [13] that allow us to set up distributed procedures with geometric convergence to the Nash Equilibrium in games with strongly monotone mappings.

Let $Q \subseteq \mathbb{R}^d$ be some convex closed subset of $\mathbb{R}^d$ and $g : Q \to \mathbb{R}^d$ be some operator on Q. The goal is to solve the corresponding variational inequality problem $VI(Q, g)$ (see Definition 3). The paper [13] proposes the following standard iterative gradient-type method to solve $VI(Q, g)$:

$$x^0 = x \in Q,$$
$$x^{k+1} = \mathcal{P}_Q\left\{x^k - \lambda g(x^k)\right\}, \quad k \geq 0, \qquad (22)$$

where λ is some stepsize. Let the operator g be strongly monotone with the constant μ and Lipschitz continuous with the constant L over the set Q. Let $\gamma = \frac{L}{\mu} \geq 1$ be the condition number of the operator g. In this case, given an optimal setting of the stepsize λ, the procedure above converges to the unique solution of $VI(Q, g)$ with a geometric rate.

Theorem 5. ([13]) Let the operator g be continuous, strongly monotone with some constant $\mu > 0$, and Lipschitz continuous with some constant $L > 0$. Let x^* be the unique solution of $VI(Q, g)$, where Q is convex and closed. Then, given the optimal stepsize $\lambda = \frac{\mu}{L^2}$, the following holds for the gradient algorithm (22): $\|x^k - x^*\|^2 \leq \exp\left\{-\frac{k}{\gamma^2}\right\}\|x^0 - x^*\|^2$.

According to Theorem 5, the convergence rate of the procedure (22) depends on the squared condition number γ^2. As $\gamma > 1$, the convergence rate would be faster, if the constant factor γ^2 is replaced by γ. To obtain this better dependence of the rate on the condition number, namely to accelerate the algorithm (22), the paper [13] follows the idea of the Nesterov type acceleration in optimization of strongly convex functions with Lipshitz continuous gradients (see [12], Section 2.2). For this purpose the authors use the sequence of positive weights $\{\lambda^t\}$, $S^k = \sum_{t=0}^k \lambda^t$, and the following functions:

$$\psi_y^\beta(x) = \langle g(y), y - x \rangle - 0.5\beta\|x - y\|^2, \quad (\beta > 0),$$

$$\Psi^k(x) = \sum_{t=0}^k \lambda^t \psi_{y^t}^\mu(x), \qquad (23)$$

where y^t is some sequence generated by the corresponding iterative process. The accelerated algorithm is based on the

definitions above and can be presented by the following iterations:

$$\lambda^0 = 1,\ y^0 \in Q,$$
$$x^k = \arg\max_{x \in Q} \Psi^k(x),$$
$$y^{k+1} = \arg\max_{x \in Q} \psi_{x^k}^L(x),$$
$$\lambda^{k+1} = \frac{1}{\gamma} \cdot S^k, \quad k \geq 0. \qquad (24)$$

The following result holds for the procedure above.

Theorem 6. ([13], Theorem 3) Let the operator g be continuous, strongly monotone with some constant $\mu > 0$, and Lipschitz continuous with some constant $L > 0$. Let x^* be the unique solution of $VI(Q, g)$, where Q is convex and closed. Let the output of the algorithm (24) be $\tilde{y}^k = \frac{1}{S^k}\sum_{t=0}^k \lambda^t y^t$. Then $0.5\mu\|\tilde{y}^k - x^*\|^2 \leq f(y^0) \cdot \gamma^2 \exp\left\{-\frac{k}{\gamma+1}\right\}$, where $f(y^0) = \sup_{y \in Q}\{\langle g(y), y^0 - y \rangle + 0.5\mu\|y^0 - y\|^2\}$.

In the next section we apply the results above to set up and accelerate distributed Nash equilibrium seeking in the case of strongly monotone game mappings.

B. Distributed Algorithms for Nash Equilibrium Seeking

Recall that in Proposition 1 we obtained the following result for the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$:

$$x^* \text{ is a Nash Equilibrium} \Leftrightarrow \langle \mathbf{F}_a(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0,$$
$$\forall \mathbf{x} \in \Omega_a, \qquad (25)$$

where $\mathbf{F}_a$ is defined in (10) and $\mathbf{x}^*$ is the Nash equilibrium matrix. Moreover, it was shown in Lemmas 1 and 2 that under Assumptions 2 and 3 there are settings for the parameters α_i, $i \in [n]$, such that the mapping $\mathbf{F}_a$ is Lipschitz continuous and strongly monotone on Ω_a with the constants $L_{\mathbf{F}_a}$ and $\mu_{\mathbf{F}_a}$ respectively (see also Remark 5). Thus, we can apply the results from [13] discussed in the preceding section to distributed Nash equilibrium seeking in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$.

We start with adapting the process (22) to $VI(\Omega_a, \mathbf{F}_a)$ in (25) with the mapping $\mathbf{F}_a(\mathbf{x}) = (I - W)\mathbf{x} + \Lambda\tilde{\mathbf{F}}(\mathbf{x})$. Analogous to notations in Subsection IV-A, we use the upper index to denote the iteration index. By writing down the kth iteration for each row in the estimation matrix $\mathbf{x}^k$ we get the following **Algorithm 1**.

Algorithm 1: GRANE

Set mixing matrix W; Choose step size $\lambda > 0$;
All players $i \in [n]$ pick arbitrary initial $x_{(i)}^0 \in \mathbb{R}^n$,
where the ith coordinate of $x_{(i)}^0$ is from Ω_i;
for $k = 0, 1, \ldots$, all players $i \in [n]$ **do**
$$x_i^{k+1} = \mathcal{P}_{\Omega_i}\{(1-\lambda)x_i^k + \sum_{j \in \mathcal{N}_i}\lambda w_{ij}\tilde{x}_{(j)i}^k$$
$$-\lambda\alpha_i\nabla_i J_i(x_i^k, \tilde{x}_{-i}^k)\};$$
  for $l = \{1, \cdots, i-1, i+1, \cdots, n\}$
  $$\tilde{x}_{(i)l}^{k+1} = (1-\lambda)\tilde{x}_{(i)l}^k + \sum_{j \in \mathcal{N}_i}\lambda w_{ij}\tilde{x}_{(j)l}^k;$$
end for;
end for.

As we can see, to run GRANE, agents start with arbitrary estimations of the joint action, namely, they do not need any prior information on the action sets of other players in the game. As time runs, agents exchange the information about their current estimations with their neighbors over the communication graph with the mixing matrix W and update their action based on local gradient steps. Thus, GRANE is a distributed gradient algorithm. Moreover, due to the relation (25), Lemmas 1, 2, and Theorems 3, 5, the following corollary can be formulated.

Corollary 1. *Let us consider the game* $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ *for which Assumptions 1-4 hold. Then there exists the unique Nash equilibrium* x^* *in* Γ. *Moreover, given* α_i, $i \in [n]$, *such that* $\mu_{\mathbf{F}_a} > 0$, *and the stepsize* $\lambda = \frac{\mu_{\mathbf{F}_a}}{L_{\mathbf{F}_a}^2}$, *the estimation matrix sequence* $\{\mathbf{x}^k\}$ *generated by the algorithm GRANE satisfies*

$$\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq \exp\left\{-\frac{k}{\gamma_{\mathbf{F}_a}^2}\right\}\|\mathbf{x}^0 - \mathbf{x}^*\|_{\mathrm{Fro}}^2,$$

where $\gamma_{\mathbf{F}_a} = \frac{L_{\mathbf{F}_a}}{\mu_{\mathbf{F}_a}}$ *and* $\mathbf{x}^*$ *is the Nash equilibrium matrix.*

To accelerate GRANE, we apply the updates in (24) to the augmented game mapping $\mathbf{F}_a$. To set up the accelerated procedure, agents need to update not only the matrix $\mathbf{x}^k$, but also an auxiliary estimation matrix $\mathbf{y}^k$. The matrix $\mathbf{y}^0$ is initialized in Ω_a. Note that, according to the definition of the functions ψ and Ψ in (23), the updates of x^k and y^{k+1} in (24) are equivalent to

$$x^k = \mathcal{P}_Q\left\{\frac{1}{S^k}\sum_{i=0}^{k}\left[\lambda^i y^i - \frac{\lambda^i}{\mu}g(y^i)\right]\right\}$$
$$y^{k+1} = \mathcal{P}_Q\left\{x^k - \frac{1}{L}g(x^k)\right\}, \tag{26}$$

where, as before, $\lambda^0 = 1$, $S^k = \sum_{t=0}^{k}\lambda^t$, $\lambda^{k+1} = \frac{S^k}{\gamma}$, μ and L are the strong monotonicity and Lipschitz continuity constants of the mapping g respectively, and $\gamma = \frac{L}{\mu}$. Thus, replacing Q and g by Ω_a and $\mathbf{F}_a$ in (26) respectively, we obtain the distributed accelerated process presented by **Algorithm 2** below[2]. Similarly to GRANE, Acc-GRANE does not require players to have any prior information on the action sets of other players in the game. Taking into account (25), Lemmas 1, 2, and Theorems 3, 6, we obtain the following corollary.

Corollary 2. *Let us consider the game* $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ *for which Assumptions 1-4 hold. Then there exists the unique Nash equilibrium* x^* *in* Γ. *Moreover, given* α_i, $i \in [n]$, *such that* $\mu_{\mathbf{F}_a} > 0$, *the matrix sequence of outputs* $\{\tilde{\mathbf{y}}^k\}$ *generated by the algorithm Acc-GRANE satisfies*

$$0.5\mu_{\mathbf{F}_a}\|\tilde{\mathbf{y}}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq F(\mathbf{y}^0) \cdot \gamma_{\mathbf{F}_a}^2 \exp\left\{-\frac{k}{\gamma_{\mathbf{F}_a}+1}\right\},$$

where $\gamma_{\mathbf{F}_a} = \frac{L_{\mathbf{F}_a}}{\mu_{\mathbf{F}_a}}$, $\mathbf{x}^*$ *is the Nash equilibrium matrix, and* $F(\mathbf{y}^0) = \sup_{\mathbf{y}\in\Omega_a}\{\langle\mathbf{F}_a(\mathbf{y}), \mathbf{y}^0 - \mathbf{y}\rangle + 0.5\mu_{\mathbf{F}_a}\|\mathbf{y}^0 - \mathbf{y}\|_{\mathrm{Fro}}^2\}$.

---

[2]Analogously to (3) and (4), we define vectors $y_{(i)}$ and $\tilde{y}_{-i}$ in respect to the matrix $\mathbf{y} \in \Omega_a$.

Remark 6. *If in Assumption 2 one only assumes strongly monotone game mapping* $\mathbf{F}$ *over* Ω, *then, according to Remark 4, the augmented mapping* $\mathbf{F}_a$ *is strongly monotone over the matrix set* Ω_a'. *In this case, if the initial estimations are such that* $x_{(i)}^0 \in \Omega$ *and* $y_{(i)}^0 \in \Omega$, $i \in [n]$, *in Algorithms 1 and 2 respectively, then Corollaries 1 and 2 hold, since, due to the rules of the procedures, all estimations* $x_{(i)}^k$, $y_{(i)}^k$, $i \in [n]$ *stay in the set* Ω *as time runs.*

Algorithm 2: Acc-GRANE

Set mixing matrix W; $\lambda^0 = 1$;
All players $i \in [n]$ pick arbitrary initial $y_{(i)}^0 \in \mathbb{R}^n$, where the ith coordinate of $y_{(i)}^0$ is from Ω_i;
for $k = 0, 1, \ldots$, all players $i \in [n]$ **do**
$\quad S^k = \sum_{t=0}^{k}\lambda^t$;
$\quad \lambda^{k+1} = \frac{S^k}{\gamma_{\mathbf{F}_a}}$;
$\quad x_i^k = \mathcal{P}_{\Omega_i}\{\frac{1}{S^k}\sum_{t=0}^{k}[\lambda^t(1-\frac{1}{\mu_{\mathbf{F}_a}})y_i^t - \frac{\lambda^t}{\mu_{\mathbf{F}_a}}(\sum_{j\in\mathcal{N}_i}w_{ij}\tilde{y}_{(j)i}^t$
$\qquad\qquad -\alpha_i\nabla_i J_i(y_i^t, \tilde{y}_{-i}^t))]\}$;
$\quad$**for** $l = \{1, \cdots, i-1, i+1, \cdots, n\}$
$\qquad \tilde{x}_{(i)l}^k = \frac{1}{S^k}\sum_{t=0}^{k}[\lambda^t\tilde{y}_{(i)l}^t + \frac{\lambda^t}{\mu_{\mathbf{F}_a}}\sum_{j\in\mathcal{N}_i}\lambda w_{ij}\tilde{y}_{(j)l}^t]$;
$\quad$**end for**;
$\quad y_i^{k+1} = \mathcal{P}_{\Omega_i}\{(1-\frac{1}{L_{\mathbf{F}_a}})x_i^k - \frac{1}{L_{\mathbf{F}_a}}[\sum_{j\in\mathcal{N}_i}w_{ij}\tilde{x}_{(j)i}^k$
$\qquad\qquad -\alpha_i\nabla_i J_i(x_i^k, \tilde{x}_{-i}^k)]\}$;
$\quad$**for** $l = \{1, \cdots, i-1, i+1, \cdots, n\}$
$\qquad y_{(i)l}^{k+1} = \tilde{x}_{(i)l}^k + \frac{1}{L_{\mathbf{F}_a}}\sum_{j\in\mathcal{N}_i}w_{ij}\tilde{x}_{(j)l}^k$;
$\quad$**end for**;
$\quad$Output $\tilde{\mathbf{y}}^k = \frac{1}{S^k}\sum_{t=0}^{k}\lambda^t\mathbf{y}^t$;
end for.

As it has been pointed out in Remark 5, the results of Corollaries 1 and 2 apply only to the class of games with strongly monotone mappings that additionally satisfy the condition $\mu_{\mathbf{F}} > \sqrt{n-1}L_{-i}$ for all i, implying a restrictive structure of cost functions. To set up the GRANE for a broader class of games, we use Lemma 3 to get the following result.

Theorem 7. *Let us consider the game* $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ *for which Assumptions 1, 3, 4, and 6 hold. Let each action set* Ω_i, $i \in [n]$, *be compact. Then, there exists a unique Nash equilibrium* x^* *in* Γ. *Moreover, given* $\alpha_i = \alpha > 0$, $i \in [n]$, *such that* $\mu_{r,\mathbf{F}_a} > 0$ *in Lemma 3, and the stepsize* $\lambda = \frac{\mu_{r,\mathbf{F}_a}}{L_{\mathbf{F}_a}^2}$, *the estimation matrix sequence* $\{\mathbf{x}^k\}$ *generated by the algorithm GRANE satisfies*

$$\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq \exp\left\{-\frac{k}{\gamma_r^2}\right\}\|\mathbf{x}^0 - \mathbf{x}^*\|_{\mathrm{Fro}}^2,$$

where $\gamma_r = \frac{L_{\mathbf{F}_a}}{\mu_{r,\mathbf{F}_a}}$ *and* $\mathbf{x}^*$ *is the Nash equilibrium matrix.*

Proof. According to Theorem 4, there exists a Nash equilibrium x^* in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \mathcal{A})$ with compact action sets and satisfying Assumptions 1. Moreover, it can be shown that under Assumption 6 the Nash equilibrium is unique. Indeed, due to Theorem 1, any Nash equilibrium $x^* \in \Omega$ solves $VI(\Omega, \mathbf{F})$, namely $\langle\mathbf{F}(x^*), x - x^*\rangle \geq 0$ for any $x \in \Omega$. Let us assume there exists another Nash equilibrium $y^* \in \Omega$.

Then, the inequality above implies $\langle \mathbf{F}(x^*), y^* - x^* \rangle \geq 0$ and $\langle \mathbf{F}(y^*), x^* - y^* \rangle \geq 0$. Hence, $\langle \mathbf{F}(x^*) - \mathbf{F}(y^*), x^* - y^* \rangle \leq 0$, which together with Assumption 6 implies $x^* = y^*$.

According to the previous discussion, the distributed setting of the GRANE presented in **Algorithm 1** can be expressed in terms of updates of estimation matrices follows: $\mathbf{x}^{k+1} = \mathcal{P}_{\Omega_a}\{\mathbf{x}^k - \lambda \mathbf{F}_a(\mathbf{x}^k)\}$, $k \geq 0$ for some $\mathbf{x}^0 = \mathbf{x} \in \Omega_a$. Moreover, since the Nash equilibrium matrix $\mathbf{x}^*$ solves $VI(\Omega_a, \mathbf{F}_a)$ (see Proposition 1), we conclude that $\mathbf{x}^* = \mathcal{P}_{\Omega_a}\{\mathbf{x}^* - \lambda \mathbf{F}_a(\mathbf{x}^*)\}$ for any $\lambda > 0$. Thus, we get the following estimation for the distance to $\mathbf{x}^* \in \Omega_a$:

$$\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq \|\mathbf{x}^k - \lambda \mathbf{F}_a(\mathbf{x}^k) - \mathbf{x}^* + \lambda \mathbf{F}_a(\mathbf{x}^*)\|_{\mathrm{Fro}}^2$$
$$= \|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2 + \lambda^2 \|\mathbf{F}_a(\mathbf{x}^k) - \mathbf{F}_a(\mathbf{x}^*)\|_{\mathrm{Fro}}^2$$
$$- 2\lambda\langle \mathbf{F}_a(\mathbf{x}^k) - \mathbf{F}_a(\mathbf{x}^*), \mathbf{x}^k - \mathbf{x}^* \rangle. \quad (27)$$

Next, according to Lemma 3, there exists $\alpha > 0$ such that $\mathbf{F}_a$ is restricted strongly monotone in respect to $\mathbf{x}^*$ with some $\mu_{r,\mathbf{F}_a} > 0$ and, hence, $\lambda\langle \mathbf{F}_a(\mathbf{x}^k) - \mathbf{F}_a(\mathbf{x}^*), \mathbf{x}^k - \mathbf{x}^* \rangle \geq \mu_{r,\mathbf{F}_a}\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2$. Thus, taking into account Lemma 1, we obtain from (27) that $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq (1 + L_{\mathbf{F}}^2\lambda^2 - 2\mu_{r,\mathbf{F}_a}\lambda)\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2$. Hence, under the optimal choice of $\lambda = \frac{\mu_{r,\mathbf{F}_a}}{L_{\mathbf{F}_a}^2}$, we conclude that $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|_{\mathrm{Fro}}^2 \leq \left(1 - \frac{1}{\gamma_r^2}\right)\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2$, and, thus, the result follows. $\blacksquare$

Remark 7. *Theorem 7 holds for games whose mapping is restricted strongly monotone. Every strongly monotone mapping is restricted strongly monotone. Thus, for example, the game of $n = 3$ players with the cost functions $J_1(x_1, x_{-1}) = 0.5x_1^2 - x_1 x_2$, $J_2(x_2, x_{-2}) = 0.5x_2^2 + x_1 x_2 + x_3 x_2$, and $J_3(x_3, x_{-3}) = 0.5x_3^2 + x_2 x_3$ defined on the whole $\mathbb{R}^3$ possesses the strongly monotone mapping with $\mu_{\mathbf{F}} = 1$. Note that Assumption 3 holds as well with $L_{-1} = L_{-3} = 1$ and $L_{-2} \geq 1$. Thus, the result from Theorem 7 applies to this game. However, Theorem 2 does not hold in this case, as $\mu_{\mathbf{F}} < \sqrt{2}L_{-i}$ for all i (see Remark 5).*

C. Discussion

In this section, we study how the functional conditions (Lipschitz constants, strong convexity, etc.), the graph topology (algebraic connectivity), and the ratio parameter α affect the convergence speed of the algorithms.

As illustrated in the above results, to reach ϵ-accuracy, GRANE requires $O(\gamma_r^2 \log(1/\epsilon))$ iterations (Theorem 7) in the case of the restricted strongly monotone augmented game mapping $\mathbf{F}_a$ with the constant $\mu_{r,\mathbf{F}_a}$, while Acc-GRANE is applicable in the case of the strongly monotone augmented game mapping $\mathbf{F}_a$ with the constant $\mu_{\mathbf{F}_a}$ and needs $O(\gamma_{\mathbf{F}_a}[\log(2F(\mathbf{y}^0)) + \log(\gamma_{\mathbf{F}_a}^2/\mu_{\mathbf{F}_a}) + \log(1/\epsilon)])$ (Theorem 2), where we usually consider $\log(2F(\mathbf{y}^0))$ and $\log(\gamma_{\mathbf{F}_a}^2/\mu_{\mathbf{F}_a})$ to be negligible compared to $\log(1/\epsilon)$. Thus, the major factor that impacts the convergence rates/complexities is γ_r and $\gamma_{\mathbf{F}_a}$ equal to $\frac{L_{\mathbf{F}_a}}{\mu_{r,\mathbf{F}_a}}$ and $L_{\mathbf{F}_a}/\mu_{\mathbf{F}_a}$ respectively.

First, let us consider the constant $\gamma_{\mathbf{F}_a}$ defining the convergence rate of Acc-GRANE. We assume that $\Lambda = \alpha I$ and

denote $H = \max_i\{L_{-i}\}$. Then the condition number can be expressed as follows:

$$\gamma_{\mathbf{F}_a} = \max \left\{ \frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I - W\}}{\tilde{\lambda}_{\min}\{I - W\} - 0.5\alpha(\sqrt{\mu_{\mathbf{F}}^2 + H^2} - \mu_{\mathbf{F}})}, \frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I - W\}}{\alpha(\mu_{\mathbf{F}} - H\sqrt{n-1})/n} \right\}.$$

By restricting to a smaller class of problems where $H \leq 0.5\mu_{\mathbf{F}}/\sqrt{n-1}$, we are able to obtain that

$$\gamma_{\mathbf{F}_a} \leq \max \left\{ \frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I - W\}}{\tilde{\lambda}_{\min}\{I - W\} - \alpha\mu_{\mathbf{F}}/(16(n-1))}, \frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I - W\}}{\alpha\mu_{\mathbf{F}}/(2n)} \right\}. \quad (28)$$

Note that in Lemma 2, Remark 5, and Theorem 2, the applicable class of games is restricted to those with $H \leq \mu_{\mathbf{F}}/\sqrt{n-1}$. In the following, we discuss the results under different choices of α. Let us denote $C = \frac{16(n-1)\tilde{\lambda}_{\min}\{I - W\}}{\mu_{\mathbf{F}}}$.

If we choose $\alpha = \frac{C}{9}$, then

$$\gamma_{\mathbf{F}_a} \leq 2n\frac{L_{\mathbf{F}}}{\mu_{\mathbf{F}}} + \frac{9}{8}\frac{\lambda_{\max}\{I - W\}}{\tilde{\lambda}_{\min}\{I - W\}}, \quad (29)$$

where $\gamma_{\mathbf{F}} \triangleq L_{\mathbf{F}}/\mu_{\mathbf{F}}$ can be understood as the condition number of the game mapping $\mathbf{F}$ and $\gamma_{\mathcal{G}} \triangleq \lambda_{\max}\{I - W\}/\tilde{\lambda}_{\min}\{I - W\}$ is strongly dependent on the algebraic connectivity of the graph. When W is chosen in the form of $W = I - t\mathbf{L}$ where $\mathbf{L}$ is the Laplacian matrix of the graph $\mathcal{G}$ and $t \in (0, 2/\lambda_{\max}\{\mathbf{L}\})$, we have $\gamma_{\mathcal{G}} = \lambda_{\max}\{\mathbf{L}\}/\tilde{\lambda}_{\min}\{\mathbf{L}\}$ which coincides with the conventional condition number of the Laplacian. An unpleasant comment for (29) is that the number of nodes has a linear multiplicative effect on the functional condition number $\gamma_{\mathbf{F}}$, which implies either the bound is not tight or the algorithm may suffer slow convergence when the number of nodes becomes large. Nonetheless an inspiring observation on (29) is that the condition number $\gamma_{\mathbf{F}_a}$ is upper bounded by the weighted sum of $\gamma_{\mathcal{G}}$ and $\gamma_{\mathbf{F}}$. This is very different from what many other decentralized algorithms illustrate in the literature. For example, in reference [24] for consensus optimization employing ADMM, the analogous complexity derived under our notation would be $O((\gamma_{\mathcal{G}}^2 + \gamma_{\mathcal{G}}\gamma_{\mathbf{F}})\log(1/\epsilon))$ where a multiplicative coupling of $\gamma_{\mathcal{G}}$ and $\gamma_{\mathbf{F}}$ is noticed. More examples of such multiplicative coupling can be found in Table II of reference [9].

By (28), if we choose $\alpha = rC$ where $r \in \left(0, \frac{1}{9}\right)$, then

$$\gamma_{\mathbf{F}_a} \leq 2n\frac{L_{\mathbf{F}}}{\mu_{\mathbf{F}}} + \frac{1}{8r}\frac{n}{n-1}\frac{\lambda_{\max}\{I - W\}}{\tilde{\lambda}_{\min}\{I - W\}}. \quad (30)$$

If we choose $\alpha = rC$ where $r \in \left(\frac{1}{9}, 1\right)$, then

$$\gamma_{\mathbf{F}_a} \leq \frac{r}{1-r}16n\frac{L_{\mathbf{F}}}{\mu_{\mathbf{F}}} + \frac{1}{1-r}\frac{\lambda_{\max}\{I - W\}}{\tilde{\lambda}_{\min}\{I - W\}}. \quad (31)$$

Note that the right-hand-sides of (30) and (31) are both larger than the right-hand-side of (29). We can see that for any absolute constant $r \in (0, 1)$, we have $\gamma_{\mathbf{F}_a} = O(n\gamma_{\mathbf{F}} + \gamma_{\mathcal{G}})$. Furthermore, too large α or too small α can both lead to poor scalability and slow convergence while a theoretically optimal α is provided as $\alpha = \frac{C}{9}$.

Similarly to the discussion above, we can obtain the following estimation for the constant γ_r from Theorem 7. Indeed, under the choice of the parameters β and α as in Lemma 3,

$\gamma_r \leq \max\left\{\frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I-W\}}{\alpha \mu_{r,\mathbf{F_a}}/2n}, \frac{\alpha L_{\mathbf{F}} + \sigma_{\max}\{I-W\}}{\lambda_{\min}\{I-W\}}\right\}$, which im-
plies that $\gamma_r \leq 2n\frac{L_{\mathbf{F}}}{\mu_{r,\mathbf{F_a}}} + \frac{\lambda_{\max}\{I-W\}}{\lambda_{\min}\{I-W\}}$. Thus, analogously
to the analysis of the constant $\gamma_{\mathbf{F_a}}$, we conclude that $\gamma_r = O(n\gamma_r + \gamma_{\mathcal{G}})$ and, hence, the number of players, condition number of the initial game mapping, and connectivity of the communication graph contribute the most to the convergence rate of GRANE.

V. SIMULATION RESULTS

Let us consider a class of games that we have discussed in Remark 5. An example of such games can be found in applications of efficient electric energy distribution over electricity markets [2], [14], [18].Specifically, we have users $\{1, 2, \ldots, 20\}$. Each user $i \in [20]$ decides on amount of electricity x_i to be purchased and its objective is to minimize the electricity price $J_i(x_i, x_{-i}) = f_i(x_i) + l_i(x)x_i$, where $f_i(x_i) = 0.5a_i x_i^2$ is the cost for demanding x_i units and $l_i(x) = \sum_{j \in [20]} c_{ij} x_j$ is the price defined by the energy tariffs which depend on the congestion in the system. Due to the fact that the cost functions are coupled through the users' decision variables, we deal with a game, where players are represented by the users of the electricity market under consideration. Each player i is imposed with a randomly generated interval $[0, d_i]$ for its own strategy, where d_i is chosen uniformly at random from the interval $[50, 100]$. The local cost function is in general dependent on actions of all players, but the underlying communication graph is a randomly generated connected graph. We randomly select c_{ij} for all possible i and j but manipulate the data so that $c_{ij} \neq c_{ji}$. In particular, the matrix C with elements c_{ij} is such that its diagonal elements are generated from the interval $[10, 20]$ uniformly at random, whereas the other ones are chosen uniformly at random from $[1, 5]$ in such a way that the matrix C is positive definite. Note that the choice of the matrix C guarantees that the game under consideration is not potential. Once we defined this matrix, we choose $\alpha_i = \alpha$ and the coefficients a_i for all i such that Assumption 5 holds. Thus, according to Corollary 2, the accelerated gradient algorithm, namely Acc-GRANE, can be applied to learn the Nash equilibrium in the game under consideration. Convergence curves for GRANE and Acc-GRANE are provided in Fig. 1. The Nash equilibrium x^* is computed using centralized gradient play method with over 12000 iterations. We use this x^* as our reference and define the relative error as $\|\mathbf{x}^k - \mathbf{x}^*\|_{\mathrm{Fro}}^2 / \|\mathbf{x}^k - \mathbf{x}^0\|_{\mathrm{Fro}}^2$. The convergence of GRANE is slow due to the small step-size and the small parameter α which implies a small constant $\mu_{\mathbf{F_a}}$ and, thus, a small step size λ ($\lambda = 0.04$) in Corollary 1.

The small parameter α and specific settings for the coefficients a_i, c_{ij}, $i, j = 1, \ldots, 20$, are chosen to guarantee that $\mu_{\mathbf{F_a}} > 0$ and, thus, Acc-GRANE is applicable. To relax these assumptions on the cost functions and be able to choose some α for any settings of a strongly convex game, we refer to the results in Lemma 3 and Theorem 7. These results allow us to choose α as in Lemma 3 and to apply GRANE to the game with any a_i, c_{ij}, $i, j = 1, \ldots, 20$, such that $\min_i a_i > 0$. The implementation of GRANE for a different choice of a_i, b_i, c_{ij}, and the corresponding parameter α are demonstrated

in Fig. 2. As we can see, in contrast to the run of GRANE applied to the strongly monotone augmented mapping $\mathbf{F_a}$ (see Fig. 1, where the relative error does not change significantly after 500 iterations), the relative error of GRANE applied to the restricted strongly monotone augmented mapping $\mathbf{F_a}$ decreases with time.
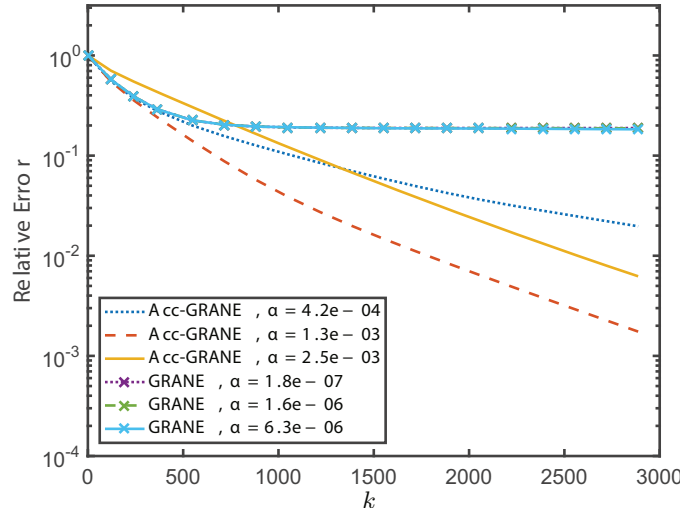

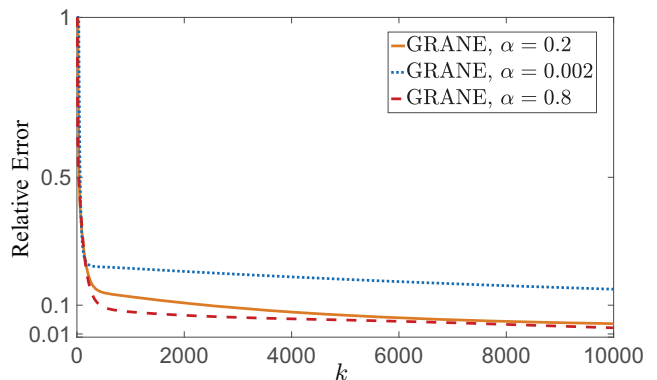Fig. 1. GRANE and Acc-GRANE applied to games with strongly monotone $\mathbf{F_a}$.

Fig. 2. GRANE applied to games with restricted strongly monotone $\mathbf{F_a}$.

VI. CONCLUSION

In this paper, we have studied a class of games with (restricted) strongly monotone mappings. We have considered the case in which players do not have the full information of joint actions but can exchange their estimates with neighbors through a decentralized communication network. To let the networked system reach the Nash equilibrium at a geometric rate, a decentralized gradient play algorithm is introduced based on the investigation of the conditions for a Nash equilibria. Moreover, by leveraging the idea of an accelerated approach for solving variational inequalities, we have obtained a version of the gradient play algorithm that guarates a faster convergence to the Nash equilibrium, due

to a better dependence on the condition number. Future works can be devoted to the investigation of modifications for the accelerated algorithm to make it applicable in games without assumption on strongly monotone augmented game mappings. Another future direction is to adapt the procedures to settings that do not require knowledge of the strong monotonicity constant and the Lipschitz continuity constant of the game mapping.

REFERENCES

[1] T. Alpcan and T. Başar. Distributed Algorithms for Nash Equilibria of Flow Control Games. In *Advances in Dynamic Games*, pages 473–498. Springer, 2005.

[2] I. Atzeni, L. G. Ordóñez, G. Scutari, D. P. Palomar, and J. R. Fonollosa. Noncooperative and cooperative optimization of distributed energy generation and storage in the demand-side of the smart grid. *IEEE Transactions on Signal Processing*, 61(10):2454–2472, May 2013.

[3] A. E. Brouwer and W. H. Haemers. *Spectra of Graphs*. New York, NY, 2012.

[4] Z. Deng and X. Nian. Distributed generalized nash equilibrium seeking algorithm design for aggregative games over weight-balanced digraphs. *IEEE Transactions on Neural Networks and Learning Systems*, 30(3):695–706, March 2019.

[5] P. Frihauf, M. Krstić, and T. Başar. Nash equilibrium seeking in noncooperative games. *Automatic Control, IEEE Transactions on*, 57(5):1192–1207, 2012.

[6] D. Gadjov and L. Pavel. A passivity-based approach to nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, 64(3):1077–1092, March 2019.

[7] J. Koshal, A. Nedić, and U. Shanbhag. A Gossip Algorithm for Aggregative Games on Graphs. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 4840–4845. IEEE, 2012.

[8] N. Li and J. Marden. Designing Games for Distributed Optimization. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):230–242, 2013.

[9] Z. Li, W. Shi, and M. Yan. A decentralized proximal-gradient method with network independent step-sizes and separated convergence rates. *arXiv preprint arXiv:1704.07807*, 2017.

[10] J. R. Marden, G. Arslan, and J. S. Shamma. Cooperative control and potential games. *Trans. Sys. Man Cyber. Part B*, 39(6):1393–1407, December 2009.

[11] J. R. Marden and J. S. Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, 75(2):788 – 808, 2012.

[12] Yu. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87. Springer, 2004.

[13] Yu. Nesterov and L. Scrimali. Solving strongly monotone variational and quasi-variational inequalities. *Discrete and Continuous Dynamical Systems - A*, 31(4):1383–1396, 2006.

[14] D. Paccagnan, M. Kamgarpour, and J. Lygeros. On aggregative and mean field games with applications to electricity markets. In *2016 European Control Conference (ECC)*, pages 196–201, June 2016.

[15] J.-S. Pang and F. Facchinei. *Finite-dimensional variational inequalities and complementarity problems : vol. 1*. Springer series in operations research. Springer, New York, Berlin, Heidelberg, 2003.

[16] L. Pavel. Distributed gne seeking under partial-decision information over networks via a doubly-augmented operator splitting approach. *arXiv preprint arXiv:1808.04465*, 2018.

[17] S. Perkins, P. Mertikopoulos, and D. S. Leslie. Mixed-strategy learning with continuous action sets. *IEEE Transactions on Automatic Control*, 62(1):379–384, Jan 2017.

[18] W. Saad, H. Zhu, H. V. Poor, and T. Başar. Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications. *IEEE Signal Processing Magazine*, 29(5):86–105, 2012.

[19] F. Salehisadaghiani and L. Pavel. Distributed Nash equilibrium seeking: A gossip-based algorithm. *Automatica*, 72:209–216, 2016.

[20] F. Salehisadaghiani, W. Shi, and L. Pavel. Distributed nash equilibrium seeking under partial-decision information via the alternating direction method of multipliers. *Automatica*, 103:27 – 35, 2019.

[21] G. Scutari, S. Barbarossa, and D. P. Palomar. Potential games: A framework for vector power control problems with coupled constraints. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 4, pages 241–244, May 2006.

[22] G. Scutari, F. Facchinei, J. Pang, and D. P. Palomar. Real and complex monotone communication games. *IEEE Transactions on Information Theory*, 60(7):4197–4231, July 2014.

[23] W. Shi, Q. Ling, G. Wu, and W. Yin. EXTRA: An Exact First-Order Algorithm for Decentralized Consensus Optimization. *SIAM Journal on Optimization*, 25(2):944–966, 2015.

[24] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin. On the Linear Convergence of the ADMM in Decentralized Consensus Optimization. *IEEE Transactions on Signal Processing*, 62(7):1750–1761, 2014.

[25] M. Stanković, K. Johansson, and D. Stipanović. Distributed Seeking of Nash Equilibria with Applications to Mobile Sensor Networks. *IEEE Transactions on Automatic Control*, 57(4):904–919, 2012.

[26] B. Swenson, S. Kar, and J. Xavier. Empirical centroid fictitious play: An approach for distributed learning in multi-agent games. *IEEE Transactions on Signal Processing*, 63(15):3888–3901, Aug 2015.

[27] T. Tatarenko. Proving convergence of log-linear learning in potential games. In *American Control Conference (ACC), 2014*, pages 972–977, June 2014.

[28] T. Tatarenko. Independent log-linear learning in potential games with continuous actions. *IEEE Transactions on Control of Network Systems*, PP(99):1–1, 2017.

[29] T. Tatarenko and M. Kamgarpour. Learning generalized nash equilibria in a class of convex games. *IEEE Transactions on Automatic Control*, pages 1–1, 2018.

[30] T. Tatarenko, W. Shi, and A. Nedić. Geometric convergence of gradient play algorithms for distributed nash equilibrium seeking. In *2018 IEEE 57th Annual Conference on Decision and Control (CDC)*. IEEE, accepted.

[31] M. Ye and G. Hu. Distributed nash equilibrium seeking by a consensus based approach. *IEEE Transactions on Automatic Control*, 62(9):4811–4818, Sep. 2017.

[32] M. Zhu and S. Martínez. Distributed coverage games for energy-aware mobile sensor networks. *SIAM J. Control and Optimization*, 51(1):1–27, 2013.
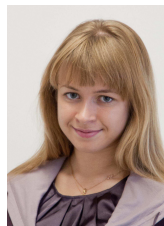
Tatiana Tatarenko is currently Postdoctoral Researcher with the Control Methods and Robotics Laboratory at TU Darmstadt, Germany. She received her Dipl.-Math. degree in Mathematics with focus on Statistics and Stochastic Processes from Lomonosov Moscow State University, Russia, in 2011, and her Ph.D. from TU Darmstadt, Germany, in 2017. She was a postdoc at Arizona State University. Her main research interests include distributed optimization, game theoretic learning, and stochastic processes in networked multi-agent systems.

Wei (Wilbur) Shi is currently Postdoctoral Researcher at the Electrical and Computer Engineering Department of Princeton University. He was a postdoc at Arizona State University, Boston University, and University of Illinois at Urbana-Champaign (Coordinated Science Laboratory). His current research interests include distributed methods in optimization, learning, and control, and applications in cyberphysical systems. He was a recipient of the Young Author Best Paper Award in IEEE Signal Processing Society.

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