2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Fully Distributed Nash Equilibrium Seeking in n-Cluster Games with Zero-Order Information

Tatiana Tatarenko

Abstract— **This paper deals with distributed Nash equilibrium seeking in n-cluster games with zero-order information. In such games agents have an access only to the values of their own cost functions and can communicate with their neighbors in the same cluster. The agents within each cluster are cooperative and intend to minimize the overall cluster's cost. This cost depends, however, on actions of agents from other clusters. Thus, there is a game between the clusters. We present a fully distributed gradient play algorithm to solve this game. The algorithm does not require agents to have any knowledge about action sets, actions or cost functions of others and is based on the zero-order information in the system. We prove convergence of the proposed procedure and estimate its convergence rate which turns out to be optimal up to a logarithmic term in the class of problems under consideration.**

I. Introduction

In our technical world with its increasing complexity there are a lot of systems consisting of many individuals, which can be considered independently operating subsystems. Such so-called multi-agent systems are, for example, robot swarms, wireless networks, transportation networks, and energy systems [9], [14], [15]. Formally, any multi-agent system consists of several agents with definite objectives to be reached by appropriately chosen local actions. These objectives can be expressed either by a common goal to be achieved in the system or by individual cost minimization. To achieve the objectives cooperative or game-theoretic optimization need to be solved respectively.

On the other hand, cooperative and individual goals may coexist in many practical situations, such as cloud computing, hierarchical optimization in Smart Grid, and adversarial networks [6], [11]. Recently, such hybrid problems have gained attention as they can be analyzed in terms of a single model called *n-cluster game*. In such an n-cluster game, each cluster corresponds to a player whose goal is to minimize its own cost function. However, the clusters in this game have some structure and consist of agents who are decision-makers. Each agent in every cluster possesses its local cost function and have no access to the cost functions of others. The objective of agents in each cluster is to minimize the sum of their local cost functions. However, the local costs and, thus, the clusters' cost functions, depend on the joint action of all agents in the system. Therefore, in such models, agents aim to achieve a Nash equilibrium in the resulting game between clusters.

Some solution approaches based on discrete-time optimization algorithms can be found in the works [4], [8], [20], [23]. The papers [4], [8] prove linear convergence in games with strongly monotone mappings under the first-order information, where agents can calculate gradients of their cost functions at any point and use this information to update their states. However, in many practical situations the agents do not know the functional form of their objectives and, thus, cannot access to their gradients. The only piece of information available to each agent is the current value of her objective function at some query point. Such situations arise, for example, in electricity markets with unknown price functions [21]. In such cases, the information structure is referred to as *zero-order one*. Due to relevancy and non-restrictiveness of the zero-order information, the further discussion is focused on the corresponding information setting. The works [20], [23] deal with n-cluster games endowed with the zero-order information. The gradient estimations in [23] are constructed in such a way that only convergence to a neighborhood of the equilibrium can be guaranteed. The paper [20] rectifies this issue and presents a procedure which converges to a Nash equilibrium. However, the algorithm proposed by [20] is not fully distributed as agents are assumed to exchange their estimations of the joint action within their clusters and, thus, need to know action sets of all other agents from the cluster they belong to in order to update the estimations.

In contrast to the mentioned works, we present a fully distributed algorithm requiring each agent to use information regarding her cost value and local action set. We assume agents to communicate with their direct neighbors within the corresponding cluster over some graph. However, we assume gents to have no access to the analytical form of their cost functions and gradients. Our contribution is as follows.

- Unlike the approach in [20], agents exchange not their estimations of joint actions but merely the experienced values of their cost functions with their neighbors in the local cluster. This information exchange takes place at the inner-loop of the optimization procedure. It allows for an efficient estimation of the cluster's gradient. Thus, each agent can perform her local update without any access to action sets of other agents. Moreover, agents exchange the scalar values of actual costs and not the vectors of joint action estimations, which reduces communication costs for problems of large dimensions. Finally, we note that it is reasonable to assume agents within each cluster to share the experienced costs with their neighbors, since a cooperative task in the cluster is to minimize its overall cost.

The author is with the Intelligent Systems and Robotics Lab at the TU Darmstadt, Germany.

- We prove convergence of the proposed algorithm and investigate its convergence rate in the case of strongly monotone games. We use two-point estimations for gradients, which significantly speed up the procedure in [20] based on a one-point estimation approach. We demonstrate that the iteration complexity of the proposed two-time-scale algorithm required to achieve the target accuracy $\varepsilon > 0$ is $O\left(\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}\right)$. This rate is optimal up to the logarithmic factor $\ln\frac{1}{\varepsilon}$, given a class of strongly convex stochastic optimization problems (see Theorem 2 in [1]).

Notations. The set $\{1,\dots,n\}$ is denoted by $[n]$. 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 $x \in \mathbb{R}^n$. We consider real normed space E, which is the space of real vectors, i.e. $E = \mathbb{R}^n$. We use (u,v) to denote the inner product in $\mathbb{R}^n$. We use $\|\cdot\|$ to denote the Euclidean norm induced by the standard dot product in $\mathbb{R}^n$. Any mapping $g: \mathbb{R}^n \to \mathbb{R}^n$ is said to be *strongly monotone on* $Q \subseteq \mathbb{R}^n$, if $(g(u)-g(v), u-v) > \nu\|u-v\|$ for some $\nu > 0$ any $u, v \in Q$, where $u \neq v$. For $x \in \mathbb{R}^d$ and a convex closed set $Y \subset \mathbb{R}^d$, $\mathcal{P}_Y\{x\}$ denotes the projection of x onto Y. The set of points within ρ distance of the boundary is denoted by $(1-\rho)Y := \{\mathbf{x} \in Y : \mathrm{dist}(\mathbf{x}, \partial Y) \geq \rho\}$ with $0 < \rho < 1$. The mathematical expectation of a random value ξ is denoted by $\mathbb{E}\{\xi\}$. We use the big-O notation, that is, the function $f(x): \mathbb{R} \to \mathbb{R}$ is $O(g(x))$ as $x \to a$, $f(x) = O(g(x))$ as $x \to a$, if $\lim_{x\to a}\frac{|f(x)|}{|g(x)|} \leq K$ for some positive constant K.

II. Nash Equilibrium Seeking Problem

We consider a non-cooperative game between n clusters. Each cluster $i \in [n]$ itself consists of n_i agents. Let J_i^j and $\Omega_i^j \subseteq \mathbb{R}$[1] denote respectively the cost function and the feasible action set of the agent j in the cluster i. We denote the joint action set of the agents in the cluster i by $\Omega_i = \Omega_i^1 \times \ldots \times \Omega_i^{n_i}$. Each function $J_i^j(x_i, x_{-i})$, $i \in [n]$, depends on $x_i = (x_i^1, \dots, x_i^{n_i}) \in \Omega_i$, which represents the joint action of the agents within the cluster i, and on $x_{-i} \in \Omega_{-i} = \Omega_1 \times \ldots \times \Omega_{i-1} \times \Omega_{i+1} \times \Omega_n$, denoting the joint action of the agents from all clusters except for the cluster i. The cooperative cost function in the cluster $i \in [n]$ is, thus, $J_i(x_i, x_{-i}) = \frac{1}{n_i}\sum_{j=1}^{n_i} J_i^j(x_i, x_{-i})$.
We assume agents within each cluster to interact over a communication graph $\mathcal{G}_i([n_i], \mathcal{A}_i)$. Here the set of nodes is equal to the set of agents $[n_i]$, whereas $\mathcal{A}_i$ is the set of arcs over which agents can exchange their local peaces of information. In our approach this information corresponds to the experienced values of the local cost functions. We associate a non-negative mixing matrix $W_i = [w_i^{jl}] \in \mathbb{R}^{n_i \times n_i}$ which defines the weights on the arcs such that $w_i^{kj} > 0$ if and only if $(j,l) \in \mathcal{A}_i$ and, thus, the agent j can receive a message from the agent l weighted by the value w_i^{jl}. The matrix W_i is called "mixing" as it allows for an efficient information exchange leading agents to a required consensus over their estimations of the cooperative cost function J_i given some technical assumption (see Assumption 5 below). However, there is *no explicit communication between the clusters*. Instead, we consider the following *zero-order* information structure in the system: No agent has access to the analytical form of any cost function and gradient, including its own. Each agent can only observe the value of its local cost function given any point from the joint action set. Formally, given a query point $\boldsymbol{x} \in \Omega$, where $\Omega = \Omega_1 \times \ldots \times \Omega_n \subseteq \mathbb{R}^N$ with $N = \sum_{i=1}^n n_i$, each agent $j \in [n_i]$, $i \in [n]$ gets the value $J_i^j(\boldsymbol{x})$ from a so called zero-order oracle, which she can communicate with her neighbors over the communication graph $\mathcal{G}_i$ in her cluster i.

Let us denote the game between the clusters introduced above by $\Gamma(n, \{J_i\}, \{\Omega_i\}, \{\mathcal{G}_i\})$, where the goal is to find a solution to the following coupled optimization problem:

$$\text{Find } \boldsymbol{x}^* = (x_1^*, \dots, x_n^*) \text{ such that:}$$
$$J_i(x_i^*, x_{-i}^*) = \min_{x_i \in \Omega_i} J_i(x_i^*, x_{-i}^*) \text{ for all } i \in [n].$$

The solution $\boldsymbol{x}^*$ to the optimization problem above is called a Nash equilibrium of the game $\Gamma = \Gamma(n, \{J_i\}, \{\Omega_i\}, \{\mathcal{G}_i\})$. We aim to solve this Nash equilibrium optimization problem under the following assumptions regarding the game Γ:

Assumption 1. *The n-cluster game- Γ under consideration is* strongly convex. *Namely, for all $i \in [n]$, the set Ω_i is convex, the cost function $J_i(x_i, x_{-i})$ is continuously differentiable in x_i for each fixed x_{-i}. Moreover, the game pseudo-gradient, which is defined as*[2]

$$\mathbf{F}(\boldsymbol{x}) \triangleq \left[\nabla_1 J_1(x_1, x_{-1}), \dots, \nabla_n J_n(x_n, x_{-n})\right]^T, \qquad (1)$$

is strongly monotone *on Ω with some constant ν.*

Assumption 2. *For each cluster $i \in [n]$, the action sets Ω_i^j, $j \in [n_i]$, are compact.*

Assumption 3. *Each gradient function $\nabla_i J_i(x_i, x_{-i})$ is Lipschitz continuous on Ω.*

Assumption 4. *Each cost function $J_i(\mathbf{x}) = O(\exp\{\|\mathbf{x}\|^\alpha\})$ as $\|\mathbf{x}\| \to \infty$, where $\alpha < 2$.*

Assumptions 1-2 above are standard in the literature on both game-theoretic and zero-order optimization [2]. Note that under Assumption 2 and given differentiable functions $\nabla_i J_i^j(x_i, x_{-i})$, Assumption 3 holds. Since in our approach we use gradient estimations based on sampling from the Gaussian distribution with unbounded support, we need Assumption 4 on the cost functions' behavior at infinity. We notice that Assumption 1 guarantees existence and uniqueness of the Nash equilibrium $\boldsymbol{x}^*$ in the game Γ which also solves the following variational inequality (see [12]):

$$\langle \mathbf{F}(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq 0 \text{ for any } \boldsymbol{x} \in \Omega. \qquad (2)$$

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

[2] $\nabla_i J_i(x_i, x_{-i}) = \frac{\partial J_i(\boldsymbol{x})}{\partial x_i} = \left(\frac{\partial J_i(\boldsymbol{x})}{\partial x_i^1}, \dots, \frac{\partial J_i(\boldsymbol{x})}{\partial x_{n_i}^1}\right) \in \mathbb{R}^{n_i}$, see **Notations**.

Finally, we make the following assumption on the communication graph, which guarantees sufficient and fast information "mixing" in the network within each cluster. To formulate the assumption, let us assume the agent j in the cluster i to have an initial estimation $u_i^j(0) \in \mathbb{R}$ of some system's parameter. Under the communication protocol defined by the graph $\mathcal{G}_i([n_i], \mathcal{A}_i)$ and the associated matrix W_i, these estimations evolve as follows: $u_i^j(t+1) = \sum_{l=1}^{n_i} w_i^{jl} u_i^l(t)$, $t = 0, 1, 2, \ldots$. Thus, during the updates the agent j receives the local estimations from its neighbors in the graph $\mathcal{G}_i$ weighted by a positive value w_i^{jl} for each neighbor l. We say that the *averaging consensus is achieved asymptotically with a geometric rate*, if there exists $\beta \in (0,1)$ such that

$$\left| u_i^j(t) - \frac{1}{n_i} \sum_{k=0}^{n_i} u_i^k(0) \right| = O(\beta^t) \text{ as } t \to \infty. \quad (3)$$

This rate we denote by $O(\beta^t)$.

Assumption 5. *The underlying communication graph $\mathcal{G}_i([n_i], \mathcal{A}_i)$ and the associated non-negative mixing matrix W_i represent such a communication protocol under which* the averaging consensus is achieved asymptotically with a geometric rate $O(\beta^t)$, $\beta \in (0,1)$.

Remark 1. *Examples of such communication protocols include undirected graphs with double stochastic mixing matrices (see [10]) as well as push-sum protocols with directed graphs and column stochastic mixing matrices [7], [22].*

Thus, in this work, we are interested in *distributed seeking of a Nash equilibrium* in the game $\Gamma(n, \{J_i\}, \{\Omega_i\}, \{\mathcal{G}_i\})$ which is endowed with the information structure described above and for which Assumptions 1-5 hold.

III. Online Zero-Order Gradient Play between Clusters

In this section we present a two-time-scale algorithm solving the problem formulated above. The main idea of the algorithm consists in introduction of an inner-loop intending to achieve a consensus on the cooperative cost value estimation at each cluster i. The outer-loop in its turn searches for the Nash equilibrium in the n-cluster game by means of efficient states' updates using the appropriate gradient estimations.

A. Algorithm Discussion

As it has been mentioned above, in the outer-loop, agents adapt the procedure to achieve a Nash equilibrium in the n-cluster game. Since this work considers strongly monotone games satisfying Assumption 1, we focus on the standard gradient play algorithm. To set up this algorithm, we introduce an auxiliary variable $\mu_i^j \in \mathbb{R}$ for each agent $j \in [n_i]$, $i \in [n]$, to be updated. We refer to the variable $\mu_i^j(t)$ as to the state (in contrast to the action $x_i^j(t) \in \mathbb{R}$) of the jth agent from the cluster i at time t. The *outer-loop* updates these variables according to the gradient play iterations:

$$\boldsymbol{\mu}_i^j(t+1) = \mathcal{P}_{(1-\rho_t)\Omega_i^j} \left\{ \boldsymbol{\mu}_i^j(t) - \alpha_t \boldsymbol{d}_i^j(t) \right\}, \quad (4)$$

where ρ_t is a shrinking set parameter, whose role will be clarified below (see **Notations.** for definition of the set $(1-\rho_t)\Omega_i^j$), α_t is a time-step parameter, and $\boldsymbol{d}_i^j(t)$ is an estimation for the local gradient $\nabla_{i,j} J_i(\boldsymbol{\mu}(t)) = \frac{\partial J_i(\boldsymbol{\mu}(t))}{\partial x_i^j}$ of the cooperative cost function at the point of the joint state $\boldsymbol{\mu}(t) = (\boldsymbol{\mu}_1(t), \ldots, \boldsymbol{\mu}_n(t)) \in \mathbb{R}^N$ with $\boldsymbol{\mu}_i(t) = (\mu_i^1(t), \ldots, \mu_i^{n_i}) \in \mathbb{R}^{n_i}$ being the joint state over the cluster i. Thus, the vector-form update within each cluster i is as follows:

$$\boldsymbol{\mu}_i(t+1) = \mathcal{P}_{(1-\rho_t)\Omega_i} \{ \boldsymbol{\mu}_i(t) - \alpha_t \boldsymbol{d}_i(t) \}, \quad (5)$$

where $\boldsymbol{d}_i(t) = (\boldsymbol{d}_i^1(t), \ldots, \boldsymbol{d}_i^{n_i}(t))$ is an estimation of the gradient $\nabla_i J_i(\boldsymbol{\mu}(t))$. We notice that, if $\boldsymbol{d}_i(t) = \nabla_i J_i(\boldsymbol{\mu}(t))$ and $\rho_t = 0$, the iterations (5) correspond to the standard gradient play between the players $i \in [n]$ proven to converge to a unique Nash equilibrium in any strongly monotone game [16]. Thus, our goal is to construct appropriate gradient estimations $\boldsymbol{d}_i(t)$ for each $i \in [n]$.

Since each agent has a direct access only to experienced values of her own cost function, the *inner-loop* is introduced to the algorithm to construct a sufficient estimation of the cooperative cost function values in the clusters. Thus, within the *inner-loop*, agents exchange values of their local costs with their neighbors over the communication graph in the cluster. To guarantee a constant variance of the estimations, we aim to use two-point queries of the gradients as follows. At each step t of the outer-loop every agent j from the cluster i plays two feasible actions: $\mu_i^j(t)$ and $x_i^j(t) = \mathcal{P}_{\Omega_i^j} \left\{ \xi_i^j(t) \right\}$, where $\xi_i^j(t)$ is sampled from the Gaussian distribution with the mean $\mu_i^j(t)$ and the variance $\sigma_t > 0$ (independently on the other agents' samplings), i.e. $\xi_i^j(t) \sim \mathcal{N}(\mu_i^j(t), \sigma_t)$. Once these actions are played the agent calculates the difference $U_i^j(t)$ of the experienced local costs:

$$U_i^j(t) = J_i^j(\boldsymbol{x}(t)) - J_i^j(\boldsymbol{\mu}(t)) \quad (6)$$

and communicates it with the neighbors over the graph $\mathcal{G}_i$ within the inner-loop $k = 0, \ldots, m_t - 1$:

$$U_i^j(t+k+1) = \sum_{l=1}^{n_i} w_{jl}^i U_i^l(t+k). \quad (7)$$

After m_t iterations of this inner-loop, the gradient estimation $d_i^j(t)$ is constructed:

$$d_i^j(t) = U_i^j(t+m_t) \frac{\xi_i^j(t) - \mu_i^j(t)}{\sigma_t^2}. \quad (8)$$

Next, the agent turns back to the outer-loop and performs the iteration (4) to update $\mu_i^j(t)$.

B. Gradient Estimations

Given the iteration t of the outer-loop, the inner-loop of the algorithm presented above performs m_t communication steps to construct the gradient estimation $d_i^j(t)$. Under Assumption 5 these steps guarantee an approach of each difference $U_i^j(t+m_t)$, $j \in [n_i]$, to the average $\frac{1}{n_i}\sum_{j=1}^{n_i} U_i^j(t)$, implying that $d_i^j(t) = U_i^j(t+m_t)\frac{\xi_i^j(t)-\mu_i^j(t)}{\sigma_t^2}$ approaches $\hat{d}_i^j(t)$, where

$$\hat{d}_i^j(t) = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} U_i^j(t)\right)\frac{\xi_i^j(t)-\mu_i^j(t)}{\sigma_t^2} = (J_i(\boldsymbol{x}(t)) - J_i(\boldsymbol{\mu}(t)))\frac{\xi_i^j(t)-\mu_i^j(t)}{\sigma_t^2}. \quad (9)$$

In this section we focus on the properties $\hat{d}_i^j(t)$ for each j and i. These properties have been previously stated in the works [18], [19]. Here we just summarize the results from these papers. First, we rewrite $\hat{d}_i^j(t)$ as follows:

$$\hat{d}_i^j = \underbrace{(J_i(\boldsymbol{\xi}(t)) - J_i(\boldsymbol{\mu}(t)))\frac{\xi_i^j(t)-\mu_i^j(t)}{\sigma_t^2}}_{Q_i^j(t)} + \underbrace{(J_i(\boldsymbol{x}(t)) - J_i(\boldsymbol{\xi}(t)))\frac{\xi_i^j(t)-\mu_i^j(t)}{\sigma_t^2}}_{P_i^j(t)}, \quad (10)$$

where $\boldsymbol{\xi}(t) = (\boldsymbol{\xi}_1(t),\dots,\boldsymbol{\xi}_n(t))$ with $\boldsymbol{\xi}_i(t) = (\xi_i^1(t),\dots,\xi_i^{n_i}(t))$ is the joint sampling point and, thus, $\boldsymbol{x}(t) = \mathcal{P}_\Omega\{\boldsymbol{\xi}(t)\}$. We denote the first term on the right hand side in (10) by $Q_i^j(t)$, whereas the second one is denoted by $P_i^j(t)$ and can be interpreted as an error caused by taking projection of the sampling point $\boldsymbol{\xi}(t)$ which guarantees feasibility of the corresponding query points.

Let $\mathcal{F}_t$ the σ-algebra generated by the random variables $\{\boldsymbol{\mu}(k),\boldsymbol{\xi}(k)\}_{k\le t}$. In the following discussion we assume fulfillment of Assumptions 1-4. We use the result of Proposition 1 in [19] to conclude that

$$\mathbb{E}\left\{Q_i^j(t)|\mathcal{F}_t\right\} = \frac{\partial J_i(\boldsymbol{\mu}(t))}{\partial x_i^j} + O(\sigma_t). \quad (11)$$

Let $\boldsymbol{Q}_i(t) = (Q_i^1(t),\dots,Q_i^{n_i}(t))$. Thus, $\boldsymbol{Q}_i$ is a biased estimation of the cluster ith local gradient $\nabla_i J_i(\boldsymbol{\mu}(t))$ at the point $\boldsymbol{\mu}(t)$. The bias is defined by the variance σ_t of the Gaussian distribution used for the estimation construction. Next, we use the result of Lemma 2 in [18] to estimate the second moment of the random variable $Q_i^j(t)$ as follows:

$$\mathbb{E}\left\{|Q_i^j(t)|^2|\mathcal{F}_t\right\} = O(1). \quad (12)$$

The relation above upper bounds the second moment by a constant value independent on t. This result is achieved due to the proposed estimation approach using two query points to sample the cost, namely $J_i^j(\boldsymbol{x}(t))$ and $J_i^j(\boldsymbol{\mu}(t))$. Finally, we estimate the projection-related term $P_i^j(t)$ as follows (see again Proposition 1 in [19] for the proof):

$$\mathbb{E}\left\{|P_i^j(t)||\mathcal{F}_t\right\} = O(\sigma_t), \quad (13)$$

as $\lim_{t\to\infty}\frac{\rho_t}{\sigma_t} = \infty$. The last relation demonstrates necessity of taking projection to the shrinked set $(1-\rho_t)\Omega_i^j$ with $\rho_t > 0$ in the updates of the states μ_i^j. Without such projection one cannot upper bound the term $|P_i^j(t)|$ in a sufficient way.

Combining the relations in (9)-(13), we obtain the following lemma.

Lemma 1. *Let Assumptions 1 - 4 hold and* $\lim_{t\to\infty}\sigma_t = 0$, $\lim_{t\to\infty}\frac{\rho_t}{\sigma_t} = \infty$. *Then* $\hat{\boldsymbol{d}}_i(t) = (\hat{d}_i^1(t),\dots,\hat{d}_i^{n_i}(t)) = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} U_i^j(t)\right)\frac{\boldsymbol{\xi}_i(t)-\boldsymbol{\mu}_i(t)}{\sigma_t^2}$ *is a biased estimation of the local gradient* $\nabla_i J_i(\boldsymbol{\mu}(t))$ *at the point* $\boldsymbol{\mu}(t)$ *satisfying the following relations:*

$$\mathbb{E}\left\{\hat{\boldsymbol{d}}_i(t)|\mathcal{F}_t\right\} = \nabla_i J_i(\boldsymbol{\mu}(t)) + O(\sigma_t),$$
$$\mathbb{E}\left\{\|\hat{\boldsymbol{d}}_i(t)\|^2|\mathcal{F}_t\right\} = O(1).$$

The iterate (4) of the outer-loop uses the gradient estimations $\boldsymbol{d}_i(t)$. As it has been mentioned at the beginning of this subsection, within the inner-loop of the procedure, $\boldsymbol{d}_i(t)$ approaches the estimation $\hat{\boldsymbol{d}}_i(t)$. The properties of $\hat{\boldsymbol{d}}_i(t)$ stated by the lemma above will be used in the proof of the main result presented in the next subsection.

Remark 2. *We emphasize that at each time t, the gradient estimations in (8) are based on the difference between the costs sampled at two points $\boldsymbol{x}(t)$ and $\boldsymbol{\mu}(t)$ (see (6)). The reason for it is that such two-point estimations allow for a uniformly bounded second moment of the term $\|d_i^j(t)\|$ (see Lemma 1), which speeds up the rate of the algorithm in comparison with the one-point estimations discussed in [20]. To avoid the requirement to estimate the cost at the joint state $\boldsymbol{\mu}(t)$, one can adapt an approach presented, for example, in [3] and use the experienced value of the cost from the previous iteration, namely $J_i^j(\boldsymbol{x}(t-1))$, instead of the value $J_i^j(\boldsymbol{\mu}(t))$ to construct the difference $U_i^j(t)$. A rigorous proof of the same convergence rate guarantee under this approach is one of the directions for future research.*

C. Main Result

We are equipped to provide the main result of this work consisting in the convergence proof of the fully distributed procedure presented in Section III-A.

Theorem 1. *Let the states $\mu_i^j(t)$, $j\in[n_i]$ $i\in[n]$, evolve according to the outer-loop iterates (4) with the m_t inner-loop iterates (7) . Let Assumptions 1–5 hold. Moreover, let the parameters in the procedure be chosen as follows: $\alpha_t = \frac{c}{t}$ with $c > \frac{1}{2\nu}$, $\sigma_t = O\left(\frac{1}{t^s}\right)$, $\rho_t = O\left(\frac{1}{t^r}\right)$, $2\le r < s$, and $m_t = \frac{(1+s)\log t}{\log(1/\beta)}$. Here ν is the strong monotonicity constant (see Assumption 1) and β is the constant related to the communication protocol (see Assumption 5).*

Then the sequence of the outer-loop iterations $\{\boldsymbol{\mu}(t)\}_t$ converges almost surely to the unique Nash equilibrium $\boldsymbol{x}^*$ and satisfies $\mathbb{E}\|\boldsymbol{\mu}(t)-\boldsymbol{x}^*\|^2 = O\left(\frac{1}{t}\right)$. Moreover, the overall number of $O\left(\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}\right)$ iterations are required to achieve $\boldsymbol{\mu}(t)$ such that $\mathbb{E}\|\boldsymbol{\mu}(t)-\boldsymbol{x}^*\|^2 < \varepsilon$, where $\varepsilon > 0$ is the target accuracy.

Proof. Let us consider the outer-loop t. At this time step, the inner-loop consists of m_t iterations. Taking into account Assumption 5 (see equation (3)), we conclude that, as $m_t \to$

∞,

$$\left|U_i^j(t+m_t) - \frac{1}{n_i}\sum_{k=1}^{n_i} U_i^j(t)\right| = O(\beta^{m_t}). \quad (14)$$

We proceed with estimating the squared distance between the joint state $\boldsymbol{\mu}_i(t+1)$ within the cluster i evolving according to (5) and the unique Nash equilibrium action $\boldsymbol{x}_i^*$ of this cluster:

$$\begin{aligned}
&\|\boldsymbol{\mu}_i(t+1) - \boldsymbol{x}_i^*\|^2 \\
&= \|\boldsymbol{\mu}_i(t+1) - \mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\} + \mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\} - \boldsymbol{x}_i^*\|^2 \\
&\leq \|\boldsymbol{\mu}_i(t+1) - \mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\}\|^2 + \|\mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\} - \boldsymbol{x}_i^*\|^2 \\
&\quad + 2\|\boldsymbol{\mu}_i(t+1) - \mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\}\|\|\mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\} - \boldsymbol{x}_i^*\| \\
&= \|\boldsymbol{\mu}_i(t+1) - \mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\}\|^2 + O(\rho_t) \\
&\leq \|\boldsymbol{\mu}_i(t) - \alpha_t \boldsymbol{d}_i(t) - \boldsymbol{x}_i^*\|^2 + O(\rho_t), \quad (15)
\end{aligned}$$

where the inequality is due to the Cauchy-Schwarz one, the second equality is due to the fact that $\|\mathcal{P}_{(1-\rho_t)\Omega_i}\{\boldsymbol{x}_i^*\} - \boldsymbol{x}_i^*\| = O(\rho_t)$ (see Lemma 7 in Appendix D in [17] for an analogous proof), and the last inequality uses non-expansion of the projection operator. Next,

$$\begin{aligned}
\|\boldsymbol{\mu}_i(t) - \alpha_t \boldsymbol{d}_i(t) - \boldsymbol{x}_i^*\|^2 &= \|\boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\|^2 \\
&- 2\alpha_t\langle \boldsymbol{d}_i(t), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle + \alpha_t^2\|\boldsymbol{d}_i(t)\|^2. \quad (16)
\end{aligned}$$

Taking into account the relations $\boldsymbol{d}_i(t) = \boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t) + \hat{\boldsymbol{d}}_i(t)$ and $\|\boldsymbol{d}_i(t)\|^2 \leq 2\|\boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t)\|^2 + 2\|\hat{\boldsymbol{d}}_i(t)\|^2$, we conclude that

$$\begin{aligned}
&\|\boldsymbol{\mu}_i(t) - \alpha_t \boldsymbol{d}_i(t) - \boldsymbol{x}_i^*\|^2 \leq \|\boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\|^2 \\
&- 2\alpha_t\langle \hat{\boldsymbol{d}}_i(t), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle \\
&- 2\alpha_t\langle \boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle \\
&+ 2\alpha_t^2\|\boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t)\|^2 + 2\alpha_t^2\|\hat{\boldsymbol{d}}_i(t)\|^2 \\
&\leq \|\boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\|^2 - 2\alpha_t\langle \hat{\boldsymbol{d}}_i(t), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle \\
&+ 2\alpha_t\|\boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t)\|\|\boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\| \\
&+ 2\alpha_t^2\|\boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t)\|^2 + 2\alpha_t^2\|\hat{\boldsymbol{d}}_i(t)\|^2. \quad (17)
\end{aligned}$$

According to the definition of $\boldsymbol{d}_i(t) = (d_i^j(t), \dots, d_i^{n_i}(t))$ (see (8)), $\hat{\boldsymbol{d}}_i(t) = \left(\frac{1}{n_i}\sum_{j=1}^{n_i} U_i^j(t)\right)\frac{\boldsymbol{\xi}_i(t) - \boldsymbol{\mu}_i(t)}{\sigma_t^2}$, and (14), the following relations hold:

$$\mathbb{E}\{\|\boldsymbol{d}_i(t) - \hat{\boldsymbol{d}}_i(t)\|^2|\mathcal{F}_t\} = O\left(\frac{1}{\sigma_t^2}\beta^{2m_t}\right). \quad (18)$$

We take the conditional expectation in respect to $\mathcal{F}_t$ of both sides in the inequality (15) combined with (17). By taking this expectation, we use the relations $\mathbb{E}\{\langle\hat{\boldsymbol{d}}_i(t), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle|\mathcal{F}_t\} = \langle\nabla_i J_i(\boldsymbol{\mu}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle$ and $\mathbb{E}\{\|\hat{\boldsymbol{d}}_i(t)\|^2\mathcal{F}_t\} = O(1)$ (see Lemma 1) as well as (18) and, thus, obtain

$$\begin{aligned}
&\mathbb{E}\{\|\boldsymbol{\mu}_i(t+1) - \boldsymbol{x}_i^*\|^2|\mathcal{F}_t\} \\
&\leq \|\boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\|^2 - 2\alpha_t\langle\nabla_i J_i(\boldsymbol{\mu}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{x}_i^*\rangle \\
&\quad + O\left(\alpha_t^2 + \frac{\alpha_t}{\sigma_t}\beta^{m_t} + \frac{\alpha_t^2}{\sigma_t^2}\beta^{2m_t} + \rho_t\right),
\end{aligned}$$

where we also used compactness of Ω (Assumption 2). Summing up the inequalities above over $i = 1, \dots, n$, we conclude that

$$\begin{aligned}
&\mathbb{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{x}^*\|^2|\mathcal{F}_t\} \leq \|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2 \\
&\quad - 2\alpha_t\langle \boldsymbol{F}(\boldsymbol{\mu}), \boldsymbol{\mu}(t) - \boldsymbol{x}^*\rangle \\
&\quad + O\left(\alpha_t^2 + \frac{\alpha_t}{\sigma_t}\beta^{m_t} + \frac{\alpha_t^2}{\sigma_t^2}\beta^{2m_t} + \rho_t\right) \\
&\leq \|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2 - 2\alpha_t\{\langle \mathbf{F}(\boldsymbol{\mu}(t)) - \mathbf{F}(\boldsymbol{x}^*), \boldsymbol{\mu}(t) - \boldsymbol{x}^*\rangle\} \\
&\quad + O\left(\frac{\alpha_t}{\sigma_t}\beta^{m_t} + \frac{\alpha_t^2}{\sigma_t^2}\beta^{2m_t} + \alpha_t^2 + \rho_t\right) \\
&\leq (1 - 2\alpha_t\nu)\|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2 \\
&\quad + O\left(\frac{\alpha_t}{\sigma_t}\beta^{m_t} + \frac{\alpha_t^2}{\sigma_t^2}\beta^{2m_t} + \alpha_t^2 + \rho_t\right), \quad (19)
\end{aligned}$$

where the second inequality is due to the solution property (2) and the last one is due to strong monotonicity of $\mathbf{F}$ (see Assumption 1). Hence, the settings for the parameters α_t, m_t, σ_t, and ρ_t imply that

$$\begin{aligned}
&\mathbb{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{x}^*\|^2|\mathcal{F}_t\} \\
&\leq \left(1 - \frac{c'}{t}\right)\|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2 + O\left(\frac{1}{t^2}\right). \quad (20)
\end{aligned}$$

Thus, $\boldsymbol{\mu}(t)$ converges almost surely to $\boldsymbol{x}^*$ (see Lemma 10 in Chapter 2.2. [13]). Taking the full expectation of the both sides in (20), we conclude that

$$\begin{aligned}
&\mathbb{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{x}^*\|^2\} \\
&\leq \left(1 - \frac{c'}{t}\right)\mathbb{E}\{\|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2\} + O\left(\frac{1}{t^2}\right), \quad (21)
\end{aligned}$$

with $c' > 1$. Finally, applying the Chung's lemma (see Lemma 4 in Chapter 2.2. [13]), we conclude that for the sequence of the outer-loop $\{\boldsymbol{\mu}(t)\}_t$ the following relation holds: $\mathbb{E}\{\|\boldsymbol{\mu}(t) - \boldsymbol{x}^*\|^2\} = O\left(\frac{1}{t}\right)$. Moreover, taking into account the length of the inner-loop, i.e. $m_t = \frac{(1+s)\log t}{\log(1/\beta)}$, we obtain the iteration complexity of the order $O\left(\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}\right)$ required to achieve the target accuracy $\varepsilon > 0$. ■

Remark 3. *The obtained iteration complexity $O\left(\frac{1}{\varepsilon}\ln\frac{1}{\varepsilon}\right)$ is optimal up to the term $\ln\frac{1}{\varepsilon}$ in the class of strongly convex optimization problems [1]. The logarithmic term appears due to existence of the inner-loop of the corresponding length. This rate is achieved as we apply a two feasible query point approach for the gradient estimations, which guarantees bounded variance of the estimations (see Lemma 1). We notice that this rate is faster than the rate of the algorithm in [20], as the latter procedure is based on a one-point gradient estimation technique which provides an unbounded variance and leads to the iteration complexity lower bounded by $O\left(\frac{1}{\sqrt{\varepsilon}}\right)$ [5]. Moreover, while the procedure presented in [20] requires each agent to have access to the joint action set of her local cluster, agents following Algorithm (4) do not use this information in their updates. We compare Algorithm (4) with the procedure from [20] numerically in the next section.*

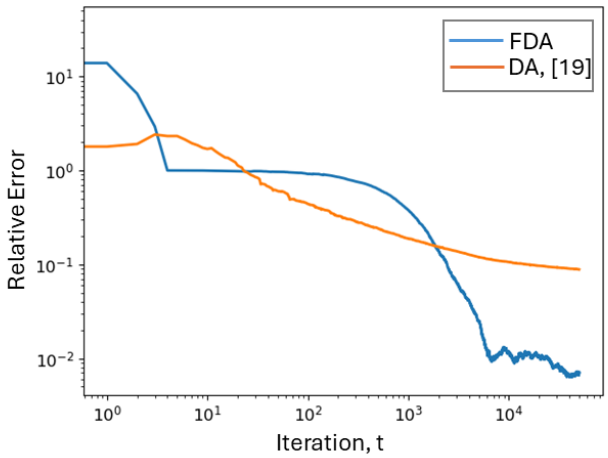

Fig. 1. Comparison of the presented fully distributed Algorithm 1 (FDA) with the distributed method (DA) presented in [20].

IV. Simulations

We consider an n-cluster games with 3 clusters ($n = 3$) and 5 agents in each of them ($n_i = 5$, $i = 1, 2, 3$). Each agent j's in the cluster i aims to minimize the cost function $J_i^j(x_i, x_{-i}) = f_i^j(x_i) + l_i(x_{-i})x_i^j$, where $f_i^j(x_i) = 0.5a_i^j\|x_i\|^2 + \langle b_i, x_i\rangle$ and $l_i(x_{-i}) = \sum_{k\neq i}\langle c_{ik}, x_k\rangle$. Thus, the local cost function of each player (cluster) i in the game is $J_i(\boldsymbol{x}) = \sum_{j=1}^{5} J_i^j(x_i, x_{-i})$. The action set of each agent j is chosen as a random closed interval on $\mathbb{R}$. We assume the communication protocol in each cluster is represented by a strongly connected graph (a randomly generated tree graph) with a double stochastic weight matrix W_i. We randomly select $a_i^j > 0$, $b_i \in \mathbb{R}^5$, and $c_{ik} \in \mathbb{R}^5$ for all possible i, k, and j to guarantee strong monotonicity of the pseudo-gradient.

For this setting, we simulate the proposed fully distributed two-time-scale algorithm and the procedure presented in [20][3]. Figure 1 contains the simulation results and demonstrates dependence of the relative error, namely $\frac{\|t\text{th iteration output} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^*\|}$, on time t Note that the blue graphics corresponds to the outer-loop iteration As it has been mentioned in Remark 3, Algorithm (4) outperforms the procedure presented in [20] in terms of the convergence rate.

V. Conclusion

In this paper we presented the fully distributed two-time-scale gradient play algorithm solving n-cluster games under a zero-order information setting. We prove the almost sure convergence of this procedure to the unique Nash equilibria and estimate the convergence rate, given a strongly monotone pseudo-gradient. The future work will be devoted to investigation of possible modification of Algorithm (4) to a procedure with a single time-scale. Moreover, scalability of the proposed algorithm and, thus, its application to systems with a large number of agents should be investigated.

[3]All parameters for these procedures are set up according to the theoretic results guaranteeing convergence to the unique Nash equilibrium.

References

[1] A. Agarwal, O. Dekel, and L. Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, June 2010.

[2] M. Bravo, D. Leslie, and P. Mertikopoulos. Bandit learning in concave n-person games. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 5666–5676, Red Hook, NY, USA, 2018. Curran Associates Inc.

[3] Y. Huang and J. Hu. A bandit learning method for continuous games under feedback delays with residual pseudo-gradient estimate, 2023.

[4] V. Willert J. Adamy J. Zimmermann, T. Tatarenko. Gradient-tracking over directed graphs for solving leaderless multi-cluster games. *arXiv preprint:2102.09406*, 2021.

[5] K. G. Jamieson, Robert N., and B. Recht. Query complexity of derivative-free optimization. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[6] M. Jarrah, M. Jaradat, Y. Jararweh, M. Al-Ayyoub, and A. Bousselham. A hierarchical optimization model for energy data flow in smart grid power systems. *Information Systems*, 53:190–200, 2015.

[7] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 482–491, 2003.

[8] M. Meng and X. Li. On the linear convergence of distributed nash equilibrium seeking for multi-cluster games under partial-decision information. *arXiv preprint:2005.06923*, 2020.

[9] A. Nedić and J. Liu. Distributed optimization for control. *Annual Review of Control, Robotics, and Autonomous Systems*, 1(1):77–103, 2018.

[10] A. Nedić, A. Ozdaglar, and P. A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

[11] D. Niyato, A. V. Vasilakos, and Z. Kun. Resource and revenue sharing with coalition formation of cloud providers: Game theoretic approach. In *2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, pages 215–224, 2011.

[12] J.-S. Pang and F. Facchinei. *Finite-dimensional variational inequalities and complementarity problems : vol. 2*. Springer series in operations research. Springer, New York, Berlin, Heidelberg, 2003.

[13] B. T. Poljak. *Introduction to optimization*. Optimization Software, 1987.

[14] 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.

[15] 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.

[16] G. Scutari, D. Palomar, F. Facchinei, and J. S. Pang. Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine*, 27:35–49, 2010.

[17] T. Tatarenko and M. Kamgarpour. Bandit learning in convex non-strictly monotone games. *arXiv preprint:2009.04258*, 2023.

[18] T. Tatarenko and M. Kamgarpour. Convergence rate of learning a strongly variationally stable equilibrium. *arXiv, eprint=2304.02355*, 2024.

[19] T. Tatarenko and M. Kamgarpour. Payoff-based learning of nash equilibria in merely monotone games. *IEEE Transactions on Control of Network Systems*, Early Access:1–12, 2024.

[20] T. Tatarenko, J. Zimmermann, and J. Adamy. Gradient play in n-cluster games with zero-order information. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3104–3109, 2021.

[21] A. C. Tellidou and A. G. Bakirtzis. Agent-based analysis of capacity withholding and tacit collusion in electricity markets. *IEEE Transactions on Power Systems*, 22(4):1735–1742, Nov 2007.

[22] K. Tsianos, S. Lawlor, and M. Rabbat. Push-Sum Distributed Dual Averaging for Convex Optimization. In *Proceedings of the 51st IEEE Annual Conference on Decision and Control*, pages 5453–5458, 2012.

[23] G. Hu Y. Pang. Nash equilibrium seeking in n-coalition games via a gradient-free method. *arXiv preprint:2008.12909*, 2020.