# Distributed No-Regret Learning for Stochastic Aggregative Games over Networks

Jinlong Lei[1,2,3], Peng Yi [1,2,3], Li Li[1,3]

1. The Department of Control Science and Engineering, Tongji University, Shanghai, 201804, China;

2. The Institute for Advanced Study, Tongji University, Shanghai, 201804, China;

3. Shanghai Research Institute for Intelligent Autonomous Systems

E-mail: leijinlong@tongji.edu.cn, yipeng@tongji.edu.cn, lili@tongji.edu.cn

**Abstract:** In this work, we propose a distributed no-regret learning for stochastic aggregative games over time-varying networks. We consider a finite set of players repeatedly playing the network aggregative game in stochastic regimes, where each player has an expectation-valued objective function depending on its own strategy and the aggregate of all player strategies, and the players can estimate gradients of their objective functions up to a zero-mean error with bounded variance. We consider the scenario in which players cannot directly obtain the aggregate value, but they are able to share their estimates of the aggregate with their neighbors without disclosing their own strategies. We design a distributed learning algorithm based on the mirror descent and dynamical averaging tracking. We then provide analysis for both the variational regret and the cost regret for aggregative games with player-specific problem being convex, and show that the expected regret bounds can be $O(\sqrt{K})$ for specific step-sizes. These analyses indicate the key correlations between the regret bounds, the network connectivity, and game structures, etc. In addition, we validate the almost sure convergence to the Nash equilibrium for the class of strictly monotone games. Finally, we present preliminary numerics by applying the proposed scheme to the Nash-Cournot competition problem.

**Key Words:** Network Aggregative Games, Distributed Learning, No-regret Learning, Mirror Descent

## 1 Introduction

In recent years, the game-theoretic models and designing tools have been extensively used for controlling or prescribing behaviors in distributed engineered systems since the decision making of individuals is inherently distributed [1]. In non-cooperative settings, they enable a flexible control paradigm for individuals to autonomously optimize their selfish objectives, e.g., congestion control [2] and resource allocation [3]. While in cooperative settings, they provide potentially tractable decentralized algorithms, e.g., distributed optimization [4] and cooperative control [5]. In particular, [6] introduces the game-based control system and investigates its controllability, [7] addresses the intrinsic formation control by an infinite-horizon non-cooperative differential game, [8] uses the zero-sum game to tackle the containment control problem with conflicting leaders, [9] adopts an aggregate game to model and analyze the energy consumption control in smart grid, etc.

In noncooperative games, there is a set of players aiming to optimize its private objective depending on its rival strategies. The solution is captured by Nash equilibrium (NE) [10], at which no player can improve its payoff by unilaterally deviating from the equilibrium strategy. On one hand, players may not have full information of the game and thus cannot compute a Nash equilibrium in an introspective manner. On the other hand, computing NE based on equilibrium analysis (such as fixed point of the best-response and gradient-response mappings) is impractical in network regime due to the high computational complexity [11]. Thus, learning is indispensable for searching Nash equilibria. The learning process can also be regarded as players taking actions in repeated games, and it can account for how each player adjusts its strategy in response to the other players over time to search for strategies resulting in higher payoffs. From this perspective, NE can be interpreted as the steady state of the learning process.

In practice, it is difficult or even impossible for players to acquire perfect global information for networked games. For example, in the bandit feedback learning, the information at the players' disposal merely includes their own payoffs received at each stage [12]. In large-scale networks, each player knows its own objective function but not that of the rest of players. But the players may observe or receive information from its neighbors through the network connecting them, and such distributed NE seeking with partially information has been recently studied in [13–15].

The most widely used class of no-regret policies for repeated games is the online mirror descent methods [12, 16]. Mirror descent method that can be traced back to [17], is a primal-dual method for constrained convex optimization. Subsequently, the method was used in distributed optimization [18, 19] and games [12, 16, 20–22]. In particular, [12] examines the long-run behavior of learning in non-cooperative concave games with bandit feedback, while [16] studies the convergence of no-regret learning in continuous games with stochastic gradient observations of the payoff function. Mirror descent method shows good averaging properties in the presence of noise [23] and achieves the optimal rate in online optimization with bandit feedback [24].

The class of aggregative games where each player's payoff is a function of its own strategy and the aggregate of all players', an example is the Cournot model where it is the aggregate supply of rivals that matters rather than their individual strategies [25, 26]. Within recent years such aggregative games have been widely used in network decision and

This work is supported by the National Natural Science Foundation of China under Grant 62003239 and 62003240, Shanghai Sailing Program No. 20YF1452800 and No. 20YF1453000, Shanghai Municipal Science and Technology Major Project No. 2021SHZDZX0100, and the Fundamental Research Funds for the Central Universities, China.

control problems (see e.g., [9, 27–29]). Though players may not directly obtain the aggregate, they can observe its neighbors' information. Motivated by the consensus-based protocols for distributed optimization [30], [14] developed distributed gradient schemes and [31] proposed best-response algorithms with sufficient conditions for guaranteeing convergence to NE. Aggregative game with coupling constraints was investigated by [32], which however requires a central node for updating the Lagrange multiplier associated with the coupling constraints. Distributed variable sample-size gradient scheme was proposed in [33] for stochastic aggregative games, which uses an increasing number of sampling gradients and communication rounds. In addition, [34–36] have also proposed continuous-time distributed algorithms via consensus-based approaches. However, most of the aforementioned works concentrated on proving the convergence to Nash equilibrium with some monotonicity conditions, while have not discussed the regret bounds for general convex aggregative games.

We consider a finite set of players repeatedly playing the network aggregative game in stochastic regimes. As far as we know, this work is the first to study the distributed no-regret learning for stochastic aggregative games. Our major contributions are summarized as follows.

1) We propose a distributed stochastic learning scheme via mirror descent and dynamical averaging tracking. At each stage, every player takes a weighted average of its neighbors' estimates of the aggregate, observes an unbiased estimate of its payoff gradients with bounded variance, takes small steps along the estimated gradient, and then mirrors it back to the strategy set.
2) For the aggregative game where each player's subproblem is convex, we provide a detailed analysis of the variational regret defined by the gap function associated with the variational inequality, and each player's cost regret being the difference between the total cost over $K$ steps and the cost of best possible strategy in the hindsight. Specially, we obtain the expected regret bounds $\mathcal{O}(\sqrt{K})$ for a specific selection of steplenghts, which implies the proposed distributed learning method is no-regret.
3) Finally, we show that for a special class of strictly monotone games, the sequence converges almost surely to the Nash equilibrium. The analysis relies on tools and techniques from stochastic approximation, martingale limit theory, and convex analysis.

The rest of paper is organized as follows. We state the problem formulation along with some basic assumptions in Section 2. In Section 3, we propose a distributed no-regret learning scheme and establish its almost sure convergence, for which the proofs are given in Section 4. We provide some preliminary numerics on Nash-Cournot games in Section 5 and conclude the paper in Section 6.

**Notations:** When referring to a vector $x$, it is assumed to be a column vector while $x^T$ denotes its transpose and $[x]_j$ denotes the $j$-th entry. $\|x\|$ denotes the Euclidean vector norm, i.e., $\|x\| = \sqrt{x^T x}$. $\langle x, y\rangle = x^T y$ denotes the inner product of vectors $x, y$. A nonnegative square matrix $A$ is called doubly stochastic if $A\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T A = \mathbf{1}^T$, where $\mathbf{1}$ denotes the vector with each entry equal 1. $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ denotes the identity matrix. Let $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$ be a directed graph with $\mathcal{N} = \{1, \cdots, N\}$ denoting the set of players and $\mathcal{E}$ denoting the set of directed edges between players, where $(j, i) \in \mathcal{E}$ if player $i$ can obtain information from player $j$. Denote by $N_i \triangleq \{j \in \mathcal{N} : (j, i) \in \mathcal{E}\}$ the set of neighboring players of player $i$. The graph $\mathcal{G}$ is called strongly connected if for any $i, j \in \mathcal{N}$ there exists a directed path from $i$ to $j$, i.e., there exists a sequence of edges $(i, i_1), (i_1, i_2), \cdots, (i_{p-1}, j)$ in the digraph with distinct nodes $i_m \in \mathcal{N}$, $\forall m : 1 \le m \le p-1$.

## 2 Problem Formulation

In this section, we define a class of aggregative games in stochastic regimes and introduce some basic assumptions.

### 2.1 Problem Statement

We now formally define the class of aggregative game $\langle \mathcal{N}, (X_i)_{i=1}^N, (f_i)_{i=1}^N\rangle$, where $\mathcal{N} \triangleq \{1, \cdots, N\}$ is a finite set of players. Each player $i \in \mathcal{N}$ selects its strategy $x_i \in \mathbb{R}^{m_i}$ from a strategy set $X_i \subset \mathbb{R}^{m_i}$. Denote by $x \triangleq (x_1^T, \cdots, x_N^T)^T \in \mathbb{R}^{\sum_{i=1}^N m_i}$ and $x_{-i} \triangleq \{x_j\}_{j\ne i}$ the strategy profile and the rival strategies, respectively. For each $i \in \mathcal{N}$, player $i$ has a cost function $f_i(x_i, \sigma(x))$ depending on its own strategy $x_i$ and an aggregate function of all players' strategies defined as $\sigma(x) \triangleq \sum_{j=1}^N h_j(x_j)$, where $h_j : X_j \to \mathbb{R}^m$. Given $\sigma(x_{-i}) \triangleq \sum_{j=1, j\ne i}^N h_j(x_j)$, the $i$th player solves the following parameterized stochastic optimization problem:

$$\min_{x_i \in X_i} f_i(x_i, h_i(x_i) + \sigma(x_{-i})), \tag{1}$$

where $f_i(x_i, \sigma(x)) \triangleq \mathbb{E}[\psi_i(x_i, \sigma(x); \xi_i(\omega))]$, the random variable $\xi_i : \Omega \to \mathbb{R}^{d_i}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_i)$, $\psi_i : \mathbb{R}^{m_i} \times \mathbb{R}^m \times \mathbb{R}^{d_i} \to \mathbb{R}$ is a scalar-valued function, and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the probability measure $\mathbb{P}_i$.

A Nash equilibrium (NE) of the aggregative game (1) is a tuple $x^* = \{x_i^*\}_{i=1}^N$ such that for each $i \in \mathcal{N}$,

$$f_i(x_i^*, \sigma(x^*)) \le f_i(x_i, h_i(x_i) + \sigma(x_{-i}^*)), \quad \forall x_i \in X_i.$$

In other words, $x^*$ is a NE if no player can gain more by unilaterally deviating from its equilibrium strategy $x_i^*$.

We consider the setting that the agents repeatedly play the game $\langle \mathcal{N}, (X_i)_{i=1}^N, (f_i)_{i=1}^N\rangle$. Suppose that each player $i \in \mathcal{N}$ knows the structure of its private function $f_i$ and $h_i$, but cannot directly observe the aggregate $\sigma(x)$. Nevertheless, players at time instant $k$ may exchange information with their neighbors over a digraph $\mathcal{G}_k = \{\mathcal{N}, \mathcal{E}_k\}$. Define $W_k = [\omega_{ij,k}]_{i,j=1}^N$ as the adjacency matrix, where $\omega_{ij,k} > 0$ if and only if $(j, i) \in \mathcal{E}_k$, and $\omega_{ij,k} = 0$, otherwise. Denote by $N_{i,k} \triangleq \{j \in \mathcal{N} : (j, i) \in \mathcal{E}_k\}$ the set of neighboring players of player $i$ at time $k$.

### 2.2 Assumptions

We impose the following conditions on the time-varying communication graphs $\mathcal{G}_k = \{\mathcal{N}, \mathcal{E}_k\}$.

**Assumption 1.** *(a) $W_k$ is doubly stochastic for any $k \ge 0$; (b) There exists a constant $0 < \mu < 1$ such that*

$$\omega_{ij,k} \ge \mu, \quad \forall j \in \mathcal{N}_{i,k}, \ \forall i \in \mathcal{N}, \ \forall k \ge 0;$$

*(c) There exists a positive integer $B$ such that the union graph $\left\{\mathcal{N}, \bigcup_{l=1}^{B} \mathcal{E}_{k+l}\right\}$ is strongly connected for all $k \geq 0$.*

We define a transition matrix $\Phi(k,s) = W_k W_{k-1} \cdots W_s$ for any $k \geq s \geq 0$ with $\Phi(k, k+1) = \mathbf{I}_N$, and state a result that will be used in the sequel.

**Lemma 1.** *[30, Proposition 1] Let Assumption 1 hold. Then there exist $\theta = (1 - \mu/(4N^2))^{-2} > 0$ and $\beta = (1 - \mu/(4N^2))^{1/B} \in (0,1)$ such that for any $k \geq s \geq 0$,*

$$\left|[\Phi(k,s)]_{ij} - 1/N\right| \leq \theta \beta^{k-s}, \quad \forall i,j \in \mathcal{N}. \tag{2}$$

We require the player-specific problem to be convex and continuously differentiable.

**Assumption 2.** *For each player $i \in \mathcal{N}$,*
*(a) the strategy set $X_i$ is closed, convex and compact;*
*(b) the cost function $f_i(x_i, \sigma(x))$ is convex in $x_i \in X_i$ for every fixed $x_{-i} \in X_{-i} = \prod_{j \neq i} X_j$;*
*(c) $f_i(x_i, \sigma)$ is continuously differentiable in $(x_i, \sigma) \in X_i \times \mathbb{R}^m$, and $h_i(x_i)$ is continuously differentiable in $x_i \in X_i$.*

For any $x \in X \triangleq \prod_{i=1}^N X_i$ and $z \in \mathbb{R}^m$, define

$$\begin{aligned} &F_i(x_i, z) \triangleq \left(\nabla_{x_i} f_i(\cdot, \sigma) + \nabla h_i(x_i) \nabla_\sigma f_i(x_i, \cdot)\right)|_{\sigma = z}, \\ &\phi_i(x) \triangleq \nabla_{x_i} f_i(x_i, \sigma(x)) = F_i(x_i, \sigma(x)), \text{ and} \\ &\phi(x) \triangleq (\phi_i(x))_{i=1}^N . \end{aligned} \tag{3}$$

By (3) and using Assumption 2(c), we see that the pseudogradient $\phi(x)$ is continuous.

**Remark 1.** *Since each player-specific optimization problem is convex, by [37, Proposition 1.4.2], $x^*$ is a NE of (1) if and only if $x^*$ is a solution to a variational inequality problem VI$(X, \phi)$, i.e., finding $x^* \in X$ such that*

$$(\tilde{x} - x^*)^\top \phi(x^*) \geq 0, \quad \forall \tilde{x} \in X. \tag{4}$$

*Since $\phi$ is continuous, $X$ is convex and compact, the existence of NE follows immediately by [37, Corollary 2.2.5].*

The introduction of $h_i : X_i \to \mathbb{R}^m$ allows each player's strategy to have distinct dimensions. We impose the following Lipschitz conditions on $h_i$ and $F_i(x_i, z)$.

**Assumption 3.** *For each player $i \in \mathcal{N}$,*
*(a) $h_i(\cdot)$ is $L_{hi}$-Lipschitz continuous in $x_i \in X_i$, i.e.,*

$$\|h_i(x_i) - h_i(x_i')\| \leq L_{hi} \|x_i - x_i'\|, \quad \forall x_i, x_i' \in X_i;$$

*(b) for any $x_i \in X_i$, $F_i(x_i, z)$ is Lipschitz continuous in $z$ over any compact set, i.e., for any positive constant $c_z$, there exists a constant $L_i$ possibly depending on $c_z$ such that for all $z_1, z_2 \in \mathbb{R}^d$ with $\|z_1\| \leq c_z$ and $\|z_2\| \leq c_z$:*

$$\|F_i(x_i, z_1) - F_i(x_i, z_2)\| \leq L_i \|z_1 - z_2\|.$$

In addition, we require each $\psi_i(x_i, \sigma; \xi_i)$ to be differentiable and assume there exists a stochastic oracle that returns unbiased gradient estimates with bounded variance.

**Assumption 4.** *For each player $i \in \mathcal{N}$ and any $\xi_i \in \mathbb{R}^{d_i}$,*
*(a) $\psi_i(x_i, \sigma; \xi_i)$ is differentiable in $x_i \in X_i$ and $\sigma \in \mathbb{R}^m$;*
*(b) for any $x_i \in X_i$ and $z \in \mathbb{R}^m$, $q_i(x_i, z; \xi_i) = (\nabla_{x_i} \psi_i(x_i, \sigma; \xi_i) + \nabla h_i(x_i) \nabla_\sigma \psi_i(x_i, \sigma; \xi_i))\ |_{\sigma = z}$ satisfies $\mathbb{E}\left[q_i(x_i, z_i; \xi_i) | x_i\right] = F_i(x_i, z_i)$ and $\mathbb{E}[\|q_i(x_i, z; \xi_i) - F_i(x_i, z)\|^2 | x_i] \leq \nu_i^2$ for some constant $\nu_i > 0$.*

**Example 1.** *(Nash Cournot competition problem). Suppose there is a collection of $N$ firms competing over $m$ markets denoted by $\mathcal{M} \triangleq \{1, \cdots, m\}$. Each firm $i$ supplies $m_i$ markets with $x_i = (x_{ij})_{j=1}^{m_i} \in \mathbb{R}^{m_i}$ amount of products. Matrix $A_i \in \mathbb{R}^m \times \mathbb{R}^{m_i}$ is used to specify the participation of firm $i$ in the markets, where $[A_i]_{lj} = 1$ if and only if firm $i$ sells $x_{ij}$ amount of products to market $l \in \mathcal{M}$, and $[A_i]_{lj} = 0$, otherwise. Denote by $A = [A_1, \cdots, A_N]$, $Ax = \sum_{i=1}^N A_i x_i \in \mathbb{R}^m$, and by $S_l = [Ax]_l$ the aggregated products of all connected firms delivered to market $l \in \mathcal{M}$. By the law of supply and demand, we assume that the price $p_l$ of products sold in market $l \in \mathcal{M}$ is determined by a linear inverse demand function $p_l(S_l; \zeta_l) = d_l + \zeta_l - b_l S_l$ corrupted by noise, where $d_l > 0$ indicates the price when the amount of products is zero, $b_l > 0$ represents the slope of the inverse demand function, and the random disturbance $\zeta_l$ is zero-mean. Let the production cost function of firm $i$ be given by $c_i(x_i; \xi_i) = (c_i + \xi_i)^T x_i$, where $c_i > 0$ is the pricing parameter and $\xi_i$ is a random variable with zero-mean. Then firm $i \in \mathcal{N}$ has a stochastic cost $\psi_i(x; \xi_i, \zeta_i) = c_i(x_i; \xi_i) - \Big(d + \zeta - BAx\Big)^T A_i x_i$ with $d = (d_1, \cdots, d_m)^T$, $B = \operatorname{diag}\{b_1, \cdots, b_m\}$, and $\zeta = col\{\zeta_1, \cdots, \zeta_m\}$. Firm $i$ aims to minimize its expected cost while satisfying a finite capacity constraint $X_i$, i.e., $\min_{x_i \in X_i} f_i(x) \triangleq \mathbb{E}[\psi_i(x; \xi_i, \zeta_i)]$.*

Note that Example 1 fits well into the aggregative game formulation (1) with $h_i(x_i) = A_i x_i$, $\sigma(x) = \sum_{i=1}^N A_i x_i$, and $f_i(x, \sigma) = c_i^T x_i - (d - B\sigma)^T A_i x_i$. Then by (3), $F_i(x_i, z) = c_i - A_i^T d + A_i^T B(z + A_i x_i)$ and $\phi_i(x) = c_i - A_i^T d + A_i^T B\big(\sum_{i=1}^N A_i x_i + A_i x_i\big)$. We can validate that Assumptions 2, 3, and 4 hold for Example 1 when the random variables $\xi_i, \zeta_l, i \in \mathcal{N}, l \in \mathcal{M}$ are zero mean with bounded variance.

Suppose that players in the network repeatedly play the stochastic aggregative games (1). At each round $k$, each player $i$ chooses a strategy $x_{i,k} \in X_i$, receives the information from its neighbors, and forms an estimate for the average aggregate denoted by $\hat{v}_{i,k+1}$. The cumulative cost regret of player $i$ up to time $K$ is defined as

$$\begin{aligned} R_{i,K} = &\sum_{k=1}^K f_i(x_{i,k}, N\hat{v}_{i,k+1}) \\ &- \min_{x_i \in X_i} \sum_{k=1}^K f_i\big(x_i, h_i(x_i) + N\hat{v}_{i,k+1} - h_i(x_{i,k})\big), \end{aligned} \tag{5}$$

indicating how much player $i$ would have gained by always taking the best decision in the hindsight given the history of utilities, strategies, and the estimated average aggregate observed up to iteration $K$. An algorithm is called no-regret in expectation if $\mathbb{E}[R_{i,K}]/K \xrightarrow[K \to \infty]{} 0$ for each player $i$.

Let $x_k = (x_{i,k})_{i=1}^N$ be the strategy profile at stage $k$ of the learning process. We can also quantify the network's total regret by a variational regret constructed from the gap function $G(x) \triangleq \sup_{y \in X} (x - y)^T \phi(x)$. It is shown by [38, Theorem 3.1] that $G(x) \geq 0$ for any $x \in X$, and $G(x) = 0$ if and only if $x$ solves VI$(X, \phi)$. We define the **variational regret** as $\mathrm{VR}_K \triangleq \sum_{k=0}^K G(x_k)$. If $\mathrm{VR}_K$ grows sublinearly in $K$, then the learning process is called "no regret".

It is desirable for players to adopt a no-regret learning algorithm since no player wants to realize that the actions it employed is strictly inferior to some fixed action in the hindsight. This paper aims to design a distributed no-regret learning algorithm using its neighboring and local information that converges almost surely to a Nash equilibrium.

## 3 Algorithm Design and Main Results

In this section, we design a distributed learning algorithm, show that the algorithm is no-regret measured by both the cost regret and the variational regret, and prove its almost sure convergence to the Nash equilibrium.

### 3.1 Mirror Descent

We assign a continuously differentiable $\sigma_i$-strongly convex function $r_i(x_i) : X_i \to \mathbb{R}$ to each player $i \in \mathcal{N}$, i.e.,

$$r_i(x_i') \geq r_i(x_i) + \langle \nabla r_i(x_i), x_i' - x_i \rangle + \frac{\sigma_i}{2} \|x_i - x_i'\|^2$$

for any $x_i, x_i' \in X_i$. The Bregman divergence associated with the function $r_i$ for any $x_i, y_i \in X_i$ is defined by

$$D_i(x_i, y_i) \triangleq r_i(x_i) - r_i(y_i) - \langle \nabla r_i(y_i), x_i - y_i \rangle. \quad (6)$$

Then $D_i(x_i, y_i) \geq \frac{\sigma_i}{2} \|x_i - y_i\|^2$ for any $x_i, y_i \in X_i$.

We have $D_i(x_i, y_i) = 0$ if and only if $x_i = y_i$. Thus, the convergence of a sequence $\{x_{i,k}\}$ to $x_i^*$ can be checked by $D_i(x_{i,k}, x_i^*) \to 0$. For technical analysis, we assume the inverse that $D_i(x_{i,k}, x_i^*) \to 0$ when $x_{i,k} \to x_i^*$. Such a "Bregman reciprocity" condition (see e.g., [12, 39]) is the blanket assumption of this work. In particular, this condition trivially holds for the Euclidean norm and the entropy regularizer, see [12, Examples 3.1 and 3.2].

### 3.2 Distributed Learning Algorithm

Each player $i$ at stage $k$ selects a strategy $x_{i,k} \in X_i$, and holds an estimate $v_{i,k}$ for the average aggregate. At stage $k+1$, player $i$ observes its neighbors' past information $v_{j,k}, j \in \mathcal{N}_{i,k}$ and updates an intermediate estimate by (7), then it computes its partial gradient and updates its strategy $x_{i,k+1}$ by a mirror descent scheme (8), and, finally, updates the average aggregate with the renewed strategy $x_{i,k+1}$ by (9). The procedures are summarized in Algorithm 1.

---

**Algorithm 1** Distributed Learning via Mirror Descent

*Initialize:* Set $x_{i,0} \in X_i$ and $v_{i,0} = h_i(x_{i,0})$ for each $i \in \mathcal{N}$.
*Iterate until convergence*

**Consensus.** Each player computes an intermediate estimate by

$$\hat{v}_{i,k+1} = \sum_{j \in \mathcal{N}_{i,k}} w_{ij,k} v_{j,k}. \quad (7)$$

**Strategy Update.** Each player $i \in \mathcal{N}$ updates its equilibrium strategy and the average aggregate by

$$x_{i,k+1} = \operatorname*{argmin}_{x_i \in X_i} \big(\langle -\alpha_k q_i(x_{i,k}, N\hat{v}_{i,k+1}; \xi_{i,k}), x_{i,k} - x_i \rangle + D_i(x_i, x_{i,k})\big), \quad (8)$$

$$v_{i,k+1} = \hat{v}_{i,k+1} + h_i(x_{i,k+1}) - h_i(x_{i,k}), \quad (9)$$

where $\alpha_k > 0$ is the steplength, and $\xi_{i,k}$ denotes a random realization of $\xi_i$ at time $k$.

---

Define the gradient noise $\zeta_{i,k} \triangleq q_i(x_{i,k}, N\hat{v}_{i,k+1}; \xi_{i,k}) - F_i(x_{i,k}, N\hat{v}_{i,k+1})$ and a proximal mapping as follows,

$$P_{i,x_i}(y_i) = \operatorname*{argmin}_{x_i' \in X_i} \big(\langle y_i, x_i - x_i' \rangle + D_i(x_i', x_i)\big). \quad (10)$$

Then (8) can be rewritten as

$$x_{i,k+1} = P_{i,x_{i,k}}\big(-\alpha_k \left(F_i(x_{i,k}, N\hat{v}_{i,k+1}) + \zeta_{i,k}\right)\big). \quad (11)$$

Define $\mathcal{F}_k \triangleq \{x_0, \zeta_{i,l}, i \in \mathcal{V}, l = 0, 1, \cdots, k-1\}$. Then by Algorithm 1 it is seen that $x_{i,k}$ and $\hat{v}_{i,k+1}$ are adapted to $\mathcal{F}_k$. From Assumption 4 it follows that for each $i \in \mathcal{V}$ :

$$\mathbb{E}[\zeta_{i,k} | \mathcal{F}_k] = 0 \text{ and } \mathbb{E}[\|\zeta_{i,k}\|^2 | \mathcal{F}_k] \leq \nu_i^2. \quad (12)$$

### 3.3 Main Results

We first provide a bound on the expected regrets $\mathbb{E}[\mathrm{VR}_k]$ and $\mathbb{E}[R_{i,K}]$, where $\mathrm{VR}_K = \sum_{k=0}^{K} G(x_k)$ is the variational regret, and $R_{i,K}$ is the cost regret defined by (5). Define

$$M_i \triangleq \max_{x_i \in X_i} \|x_i\|, \; M_H \triangleq \sum_{j=1}^{N} \max_{x_j \in X_j} \|h_j(x_j)\|, \quad (13)$$

$$\tilde{C} \triangleq \theta M_H + 2\theta \sum_{j=1}^{N} L_{hj} M_j / (1 - \beta), \quad (14)$$

$$C_i \triangleq N \tilde{C} L_{fi} + \max_{x \in X} \|\phi_i(x)\|, \; i \in \mathcal{N}, \quad (15)$$

where the boundedness of $\|\phi_i(x)\|$ follows by the compactness of $X$ and the continuity of $\phi_i(x)$.

**Theorem 1.** *Suppose Assumptions 1, 2, 3, 4 hold, and that $\{\alpha_k\}$ is a monotonically nonincreasing sequence. Let $\{x_k\}$ be generated by Algorithm 1. Define $R_i \triangleq \max_{x_i, y_i \in X_i} D_i(x_i, y_i)$ and $R \triangleq \sum_{i=1}^{N} R_i$. Then*

$$\mathbb{E}[\mathrm{VR}_K] \leq F_1 + \frac{R}{2\alpha_K} + F_2 \sum_{k=0}^{K} \alpha_k \quad (16)$$

$$\text{with } F_1 \triangleq \tfrac{M_H \theta N}{1-\beta} \sum_{i=1}^{N} L_{fi} M_i, \text{ and}$$

$$F_2 \triangleq \tfrac{\theta N \sum_{i=1}^{N} L_{fi} M_i}{\beta(1-\beta)} \sum_{j=1}^{N} L_{hj} \sigma_j^{-1} (C_j + \nu_j) + \sum_{i=1}^{N} \tfrac{C_i + \nu_i^2}{4\sigma_i}.$$

*In addition, the following holds for each $i \in \mathcal{N}$.*

$$\mathbb{E}[R_{i,K}] \leq \tfrac{C_i(C_i + \nu_i)}{\sigma_i} \sum_{k=0}^{T} \alpha_k + \frac{R_i}{\alpha_T}. \quad (17)$$

The proof of Theorem 1 is given in Section 4.2. The following corollary shows how a specific selection of learning rate $\alpha_k$ result in practical bounds on the expected variational regret $\mathbb{E}[\mathrm{VR}_K]$ and the expected cost regret $\mathbb{E}[R_{i,K}]$.

**Corollary 1.** *Suppose Assumptions 1, 2, 3, and 4 hold. Let $\{x_k\}$ be generated by Algorithm 1 with $\alpha_k = \frac{1}{\sqrt{k+1}}$. Then*

$$\mathbb{E}[\mathrm{VR}_K] \leq F_1 + \frac{F_2}{2} + \frac{R + F_2}{2} \sqrt{K+1}$$

*and for each $i \in \mathcal{N}$,*

$$\mathbb{E}[R_i(K)] \leq \tfrac{C_i(C_i + \nu_i)}{2\sigma_i} + \left(\tfrac{C_i(C_i + \nu_i)}{2\sigma_i} + R_i\right) \sqrt{K+1}.$$

The result follows immediately from Theorem 1 and

$$\begin{aligned}&\textstyle\sum_{k=0}^{K}\frac{1}{\sqrt{k+1}}\le 1+\int_1^{K+1}x^{-1/2}dx\\&=1+\frac{x^{1/2}}{2}|_1^{K+1}\sqrt{K+1}=\frac{\sqrt{K+1}+1}{2}.\end{aligned}$$

Since both $\mathbb{E}[\mathrm{VR}_K]/K\to 0$ and $\mathbb{E}[R_i(K)]/K\to 0$, i.e., $\mathbb{E}[\mathrm{VR}_K]$ and $\mathbb{E}[R_i(K)]$ are sublinear in $K$, Algorithm 1 is called a distributed no-regret learning scheme.

To further establish the almost sure convergence, we require the pseudogradient mapping $\phi(x)$ to be strictly monotone similarly to [14].

**Assumption 5.** $(\phi(x)-\phi(x'))^T(x-x')>0$ *for any* $x,x'\in X$ *and* $x\neq x'$.

We impose the following condition on steplengths, which is required to be square-summable but not summable. Such a steplength condition is widely used in stochastic approximation algorithms, see e.g., [40].

**Assumption 6.** $\{\alpha_k\}$ *is a monotonically nonincreasing sequence,* $\sum_{k=0}^{\infty}\alpha_k^2<\infty$, *and* $\sum_{k=0}^{\infty}\alpha_k=\infty$.

The following theorem shows that for the case where the pseudogradient of the aggregative game is strictly monotone, the sequence $\{x_k\}$ generated by Algorithm 1 converges almost surely to the unique Nash equilibrium.

**Theorem 2.** *Suppose Assumptions 1, 2, 5, 3, 4, and 6 hold. Let* $\{x_k\}$ *be generated by Algorithm 1. Then*

$$\lim_{k\to\infty}x_k=x^*,\quad a.s.$$

The proof of Theorem 2 is given in Section 4.3.

## 4 Proof of Main Results

### 4.1 Preliminary Results

We now establish a bound on the consensus error of the aggregate, measured by $\|\sigma(x_k)-N\hat{v}_{i,k+1}\|$.

**Proposition 1.** *Consider Algorithm 1. Let Assumptions 1, 2, and 3 hold. Then for each player* $i\in\mathcal{N}$,

$$\begin{aligned}&\|\sigma(x_k)-N\hat{v}_{i,k+1}\|\le\theta M_H N\beta^k\\&+\theta N\sum_{s=1}^{k}\beta^{k-s}\alpha_{s-1}\sum_{j=1}^{N}L_{hj}\sigma_j^{-1}(C_j+\|\zeta_{j,s-1}\|).\end{aligned}\tag{18}$$

**Proof.** Since $v_{i,0}=h_i(x_{i,0})$ and $W(k)$ is doubly stochastic, similarly to [14, Lemma 2], we can show by induction that

$$\sum_{i=1}^{N}v_{i,k}=\sum_{i=1}^{N}h_i(x_{i,k})=\sigma(x_k),\quad\forall k\ge 0.\tag{19}$$

Akin to [14, Eqn. (16)], we give an upper bound on $\left\|\frac{\sigma(x_k)}{N}-\hat{v}_{i,k+1}\right\|$. By combining (9) with (7), we have

$$\begin{aligned}&v_{i,k+1}=\textstyle\sum_{j=1}^{N}w_{ij,k}v_{j,k}+h_i(x_{i,k+1})-h_i(x_{i,k})\\&=\textstyle\sum_{j=1}^{N}[\Phi(k,0)]_{ij}v_{j,0}+h_i(x_{i,k+1})-h_i(x_{i,k})\\&+\textstyle\sum_{s=1}^{k}\sum_{j=1}^{N}[\Phi(k,s)]_{ij}(h_i(x_{i,s})-h_i(x_{i,s-1})).\end{aligned}$$

Then by (9), we have $\hat{v}_{i,k+1}=\sum_{j=1}^{N}[\Phi(k,0)]_{ij}v_{j,0}+\sum_{s=1}^{k}\sum_{j=1}^{N}[\Phi(k,s)]_{ij}(h_j(x_{j,s})-h_j(x_{j,s-1}))$. By using (19), we have that

$$\frac{\sigma(x_k)}{N}=\frac{\sum_{j=1}^{N}v_{j,0}}{N}+\sum_{s=1}^{k}\sum_{j=1}^{N}\frac{1}{N}(h_j(x_{j,s})-h_j(x_{j,s-1})).$$

Therefore, we obtain the following bound.

$$\begin{aligned}&\left\|\tfrac{\sigma(x_k)}{N}-\hat{v}_{i,k+1}\right\|\le\textstyle\sum_{j=1}^{n}\left|\frac{1}{N}-[\Phi(k,0)]_{ij}\right|\|v_{j,0}\|\\&+\textstyle\sum_{s=1}^{k}\sum_{j=1}^{N}\left|\frac{1}{N}-[\Phi(k,s)]_{ij}\right|\left\|h_j(x_{j,s})-h_j(x_{j,s-1})\right\|.\end{aligned}$$

Then by using (2), $v_{i,0}=h_i(x_{i,0})$, and Assumption 3(a), we obtain that

$$\begin{aligned}&\left\|\tfrac{\sigma(x_k)}{N}-\hat{v}_{i,k+1}\right\|\le\theta\beta^k\textstyle\sum_{j=1}^{N}\|h_j(x_{j,0})\|\\&+\theta\textstyle\sum_{s=1}^{k}\beta^{k-s}\sum_{j=1}^{N}L_{hj}\left\|x_{j,s}-x_{j,s-1}\right\|.\end{aligned}\tag{20}$$

This combined with (13) proves

$$\begin{aligned}&\left\|\frac{\sigma(x_k)}{N}-\hat{v}_{i,k+1}\right\|\le\theta\beta^k M_H+2\theta\sum_{s=1}^{k}\beta^{k-s}\sum_{j=1}^{N}L_{hj}M_j\\&\le\theta M_H+2\theta\sum_{j=1}^{N}L_{hj}M_j/(1-\beta)\overset{(14)}{=}\tilde{C}.\end{aligned}\tag{21}$$

By (13) and (19), we have that $\|\sigma(x_k)\|\le M_H$ for any $k\ge 0$. Thus by (21), we obtain that for each $i\in\mathcal{N}$

$$\|N\hat{v}_{i,k+1}\|\le N\tilde{C}+M_H,\quad\forall k\ge 0.$$

Then by using Assumption 3(b) and (3), we obtain that for each $j\in\mathcal{N}$ and any $s\ge 0$ :

$$\begin{aligned}&\|F_j(x_{j,s},N\hat{v}_{j,s})\|\\&\le\|F_j(x_{j,s},N\hat{v}_{j,s})-F_j(x_{j,s},\sigma(x_s)\|+\|F_j(x_{j,s},\sigma(x_s)\|\\&\le L_{fj}\|N\hat{v}_{j,s}-\sigma(x_s)\|+\|\phi_j(x_s)\|\\&\overset{(21)}{\le}N\tilde{C}L_{fi}+\max_{x\in X}\|\phi_i(x)\|\overset{(15)}{=}C_i.\end{aligned}\tag{22}$$

By applying the optimality condition to (8), and using the definition (6), we have that

$$\begin{aligned}&(x_i-x_{i,k+1})^T\big(\alpha_k(\zeta_{i,k}+F_i(x_{i,k},N\hat{v}_{i,k+1}))\\&+\nabla r_i(x_{i,k+1})-\nabla r_i(x_{i,k})\big)\ge 0,\quad\forall x_i\in X_i.\end{aligned}\tag{23}$$

By setting $x_i=x_{i,k}$ in (23), rearranging the terms, and using the $\sigma_i$-strongly convexity of $r_i$, we have that

$$\begin{aligned}&(x_{i,k}-x_{i,k+1})^T\alpha_k(\zeta_{i,k}+F_i(x_{i,k},N\hat{v}_{i,k+1}))\\&\ge(x_{i,k}-x_{i,k+1})^T\big(\nabla r_i(x_{i,k})-\nabla r_i(x_{i,k+1})\big)\\&\ge\sigma_i\|x_{i,k}-x_{i,k+1}\|^2.\end{aligned}$$

Then from (22) it follows that

$$\begin{aligned}\|x_{i,k}-x_{i,k+1}\|&\le\frac{\alpha_k}{\sigma_i}(\|\zeta_{i,k}\|+\|F_i(x_{i,k},N\hat{v}_{i,k+1})\|)\\&\overset{(22)}{\le}\frac{\alpha_k}{\sigma_i}\left(C_i+\|\zeta_{i,k}\|\right).\end{aligned}\tag{24}$$

This together with (13) and (20) produces (18). □

We now state a property from [12, Proposition B.3] regarding the Bregman divergence and proximal mapping defined in (6) and (10), respectively.

**Lemma 2.** *Let $r_i$ be a smooth and $\sigma_i$-strongly regularizer over the convex set $X_i$. Then for all $x_i, y_i, z_i \in X_i$, $D_i(x_i, y_i) \geq \frac{\sigma_i}{2}\|x_i - y_i\|^2$,*

$$\begin{aligned} &D_i(y_i, x_i) - D_i(y_i, z_i) - D_i(z_i, x_i) \\ &= \langle \nabla r_i(z_i) - \nabla r_i(x_i), y_i - z_i \rangle, \end{aligned} \tag{25}$$

*and*

$$D_i(z_i, P_{i,x_i}(y_i)) \leq D_i(z_i, x_i) + \langle y_i, x_i - z_i \rangle + \frac{1}{2\sigma_i}\|y_i\|^2. \tag{26}$$

The following proposition gives a recursion for the error $D(x, x_{k+1}) = \sum_{i=1}^N D_i(x_i, x_{i,k+1})$ measured by the Bregman distance. Define

$$\varepsilon_k \triangleq \alpha_k \sum_{i=1}^N L_{fi} M_i \|N\hat{v}_{i,k+1} - \sigma(x_k)\|. \tag{27}$$

**Proposition 2.** *Suppose Assumptions 1, 2, 3, and 4 hold. Let $\{x_k\}$ be generated by Algorithm 1. Then the following holds for any $x \in X$ with $C_i$ is defined by* (15).

$$\begin{aligned} &\mathbb{E}[D(x, x_{k+1})|\mathcal{F}_k] \leq D(x, x_k) + 2\varepsilon_k \\ &- 2\alpha_k(x_k - x)^T\phi(x_k) + \alpha_k^2 \sum_{i=1}^N \frac{C_i + \nu_i^2}{2\sigma_i},\ a.s. \end{aligned} \tag{28}$$

**Proof.** By using (11) and (26), we have

$$\begin{aligned} &D_i(x_i, x_{i,k+1}) \leq D_i(x_i, x_{i,k}) \\ &-\langle \alpha_k \left(F_i(x_{i,k}, N\hat{v}_{i,k+1}) + \zeta_{i,k}\right), x_{i,k} - x_i \rangle \\ &+\tfrac{1}{2\sigma_i}\|\alpha_k \left(F_i(x_{i,k}, N\hat{v}_{i,k+1}) + \zeta_{i,k}\right)\|^2 \\ &\leq D_i(x_i, x_{i,k}) - \alpha_k\langle \zeta_{i,k}, x_{i,k} - x_i \rangle \\ &-\alpha_k\langle F_i(x_{i,k}, N\hat{v}_{i,k+1}), x_{i,k} - x_i \rangle \\ &+\tfrac{\alpha_k^2}{2\sigma_i}\left(\|F_i(x_{i,k}, N\hat{v}_{i,k+1})\|^2 + \|\zeta_{i,k}\|^2\right) \\ &+\tfrac{\alpha_k^2}{\sigma_i}\langle F_i(x_{i,k}, N\hat{v}_{i,k+1}), \zeta_{i,k} \rangle. \end{aligned}$$

Since $x_{i,k}$ and $\hat{v}_{i,k+1}$ are adapted to $\mathcal{F}_k$, by taking conditional expectation on $\mathcal{F}_k$, using (12) and (22), we obtain

$$\begin{aligned} &\mathbb{E}[D_i(x_i, x_{i,k+1})|\mathcal{F}_k] \leq D_i(x_i, x_{i,k}) \\ &- \alpha_k\langle F_i(x_{i,k}, N\hat{v}_{i,k+1}), x_{i,k} - x_i^* \rangle + \frac{C_i + \nu_i^2}{2\sigma_i}\alpha_k^2. \end{aligned} \tag{29}$$

Note by Assumption 3(b), (3), and (13) that

$$\begin{aligned} &\langle -F_i(x_{i,k}, N\hat{v}_{i,k+1}), x_{i,k} - x_i \rangle \\ &= \langle -F_i(x_{i,k}, \sigma(x_k)), x_{i,k} - x_i \rangle \\ &+\langle F_i(x_{i,k}, \sigma(x_k)) - F_i(x_{i,k}, N\hat{v}_{i,k+1}), x_{i,k} - x_i \rangle \\ &\leq -\langle \phi_i(x_k), x_{i,k} - x_i \rangle \\ &+\|F_i(x_{i,k}, \sigma(x_k)) - F_i(x_{i,k}, N\hat{v}_{i,k+1})\|\|x_{i,k} - x_i\| \\ &\leq -\langle \phi_i(x_k), x_{i,k} - x_i \rangle + 2L_{fi}M_i\|\sigma(x_k) - N\hat{v}_{i,k+1})\|. \end{aligned}$$

This together with (29) implies that

$$\begin{aligned} &\mathbb{E}[D_i(x_i, x_{i,k+1})|\mathcal{F}_k] \leq D_i(x_i, x_{i,k}) \\ &- \alpha_k\langle \phi_i(x_k), x_{i,k} - x_i \rangle + \tfrac{\alpha_k^2}{2\sigma_i}\left(C_i + \nu_i^2\right) \\ &+2L_{fi}M_i\alpha_k\|N\hat{v}_{i,k+1} - \sigma(x_k)\|, \quad a.s. \end{aligned}$$

Summing up the above inequality from $i = 1$ to $N$, we prove (28). $\square$

We state a supermartingale convergence result that will be employed in the proof, see e.g., [41, Lemma 11, p. 50].

**Lemma 3.** *Let $v_k, e_k, \zeta_k, \gamma_k$ be nonnegative random variables adapted to some $\sigma$-algebra $\mathcal{F}_k$. If almost surely, $\sum_{k=0}^\infty e_k < \infty$, $\sum_{k=0}^\infty \gamma_k < \infty$, and*

$$\mathbb{E}[v_{k+1}|\mathcal{F}_k] \leq (1 + \gamma_k)v_k + \varepsilon_k - \zeta_k,$$

*Then $v_k$ converges almost surely and $\sum_{k=0}^\infty \zeta_k < \infty,\ a.s.$.*

### 4.2 Proof of Theorem 1

By taking unconditional expectations on both sides of (28), and rearranging the terms, we obtain that for any $x \in X$,

$$\begin{aligned} \mathbb{E}[(x_k - x)^T\phi(x_k)] \leq &\tfrac{\mathbb{E}[D(x,x_k)] - \mathbb{E}[D(x,x_{k+1})]}{2\alpha_k} \\ &+ \tfrac{\mathbb{E}[\varepsilon_k]}{\alpha_k} + \alpha_k \sum_{i=1}^N \tfrac{C_i + \nu_i^2}{4\sigma_i}. \end{aligned}$$

Summing the above inequality from $k = 0, \ldots, K$ produces

$$\begin{aligned} &\sum_{k=0}^K \mathbb{E}[(x_k - x)^T\phi(x_k)] \leq \frac{\mathbb{E}[D(x, x_0)]}{2\alpha_0} - \frac{\mathbb{E}[D(x, x_{K+1})]}{2\alpha_K} \\ &+ \sum_{k=1}^K \left(\frac{1}{2\alpha_k} - \frac{1}{2\alpha_{k-1}}\right)\mathbb{E}[D(x, x_k)] \\ &+ \sum_{k=0}^K \frac{\mathbb{E}[\varepsilon_k]}{\alpha_k} + \sum_{i=1}^N \frac{C_i + \nu_i^2}{4\sigma_i}\sum_{k=0}^K \alpha_k \\ &\leq \frac{R}{2\alpha_K} + \sum_{k=0}^K \frac{\mathbb{E}[\varepsilon_k]}{\alpha_k} + \sum_{i=1}^N \frac{C_i + \nu_i^2}{4\sigma_i}\sum_{k=0}^K \alpha_k, \ \ \forall x \in X. \end{aligned} \tag{30}$$

where the last inequality follows from $D(x, x_{K+1}) \geq 0$ and the definition of $R$.

By (12) and the Jensen's inequality, we have that $\mathbb{E}[\|\zeta_{i,s}\|] \leq \sqrt{\mathbb{E}[\|\zeta_{i,s}\|^2]} \leq \nu_i$ for each $i \in \mathcal{N}$ and any $s \geq 0$. This combined with (18) and (27) proves that

$$\mathbb{E}[\varepsilon_k/\alpha_k] = \theta N \textstyle\sum_{i=1}^N L_{fi}M_i\left(M_H\beta^k + e_k\right), \tag{31}$$

where $e_k \triangleq \sum_{s=1}^k \beta^{k-s}\alpha_{s-1}\sum_{j=1}^N L_{hj}\sigma_j^{-1}(C_j + \nu_j)$. Note that $\sum_{k=0}^K \beta^k \leq \sum_{k=0}^\infty \beta^k \leq \frac{1}{1-\beta}$, and $\sum_{k=0}^K\sum_{s=1}^k \beta^{k-s}\alpha_{s-1} \leq \sum_{s=0}^{K-1}(\sum_{k=0}^K \beta^{k-1})\alpha_s \leq \frac{\sum_{s=0}^{K-1}\alpha_s}{\beta(1-\beta)}$. This together with (30) and (31) implies $\sum_{k=1}^K \mathbb{E}[(x_k - x)^T\phi(x_k)] \leq \frac{R}{2\alpha_K} + \frac{\theta N M_H}{1-\beta}\sum_{i=1}^N L_{fi}M_i + F_2\sum_{k=0}^K \alpha_k$. By maximizing this inequality over $x \in X$ and using the Jensen's inequality to interchange the $\max$ and $\mathbb{E}$ operations, we prove (16).

(ii) By Assumption 2, using the definition of $F_i(x_i, z)$ in (3), (22), and (24), we obtain that

$$\begin{aligned} &f_i(x_{i,k}, N\hat{v}_{i,k+1}) - f_i(x_i, h_i(x_i) + N\hat{v}_{i,k+1} - h_i(x_{i,k})) \\ &\leq (x_{i,k} - x_i)^T F_i(x_{i,k}, N\hat{v}_{i,k+1}) \\ &\leq (x_{i,k} - x_{i,k+1})^T F_i(x_{i,k}, N\hat{v}_{i,k+1}) \\ &+ (x_{i,k+1} - x_i)^T F_i(x_{i,k}, N\hat{v}_{i,k+1}) \\ &\leq \frac{\alpha_k C_i}{\sigma_i}\left(C_i + \|\zeta_{i,k}\|\right) \\ &+ (x_{i,k+1} - x_i)^T F_i(x_{i,k}, N\hat{v}_{i,k+1}) \end{aligned} \tag{32}$$

Note by (23) that

$$\begin{aligned} &(x_{i,k+1} - x_i)^T\left(\alpha_k(\zeta_{i,k} + F_i(x_{i,k}, N\hat{v}_{i,k+1})\right) \\ &\leq (x_{i,k+1} - x_i)^T\left(\nabla r_i(x_{i,k}) - \nabla r_i(x_{i,k+1})\right). \end{aligned} \tag{33}$$

By using (25), we have that

$$\begin{aligned}&(x_{i,k+1}-x_i)^T\big(\nabla r_i(x_{i,k})-\nabla r_i(x_{i,k+1})\big)\\&=D_i(x_i,x_{i,k})-D_i(x_i,x_{i,k+1})-D_i(x_{i,k+1},x_{i,k})\\&\le D_i(x_i,x_{i,k})-D_i(x_i,x_{i,k+1}).\end{aligned}$$

This combined with (33) implies that

$$\begin{aligned}&(x_{i,k+1}-x_i)^T F_i(x_{i,k},N\hat{v}_{i,k+1})\\&\le \tfrac{D_i(x_i,x_{i,k})-D_i(x_i,x_{i,k+1})}{\alpha_k}+(x_i-x_{i,k+1})^T\zeta_{i,k}.\end{aligned}$$

Since $\{\alpha_k\}$ is monotonically non-increasing, by the definition of $R_i$, we have that

$$\begin{aligned}&\textstyle\sum_{k=0}^K \frac{D_i(x_i,x_{i,k})-D_i(x_i,x_{i,k+1})}{\alpha_k}\\&\le \tfrac{D_i(x_i,x_{i,0})}{\alpha_0}+\textstyle\sum_{k=1}^K\left(\frac{1}{\alpha_k}-\frac{1}{\alpha_{k-1}}\right)D_i(x_i,x_{i,k})\\&\le \tfrac{R_i}{\alpha_0}+\textstyle\sum_{k=1}^K\left(\frac{1}{\alpha_k}-\frac{1}{\alpha_{k-1}}\right)R_i=\frac{R_i}{\alpha_K}.\end{aligned}$$

Then by summing up (32) from $k=0$ to $K$, we obtain that

$$\begin{aligned}R_{i,K}\le&\textstyle\sum_{k=0}^K\frac{\alpha_k C_i}{\sigma_i}(C_i+\|\zeta_{i,k}\|)\\&+\tfrac{R_i}{\alpha_K}+\textstyle\sum_{k=0}^K(x_i-x_{i,k+1})^T\zeta_{i,k}.\end{aligned}$$

Then by taking expectations on both sides of the above equation, and using $\mathbb{E}[\|\zeta_{i,k}\|]\le\nu_i$, we prove (17). □

### 4.3 Proof of Theorem 2

Since $\alpha_k\le\alpha_s$ for all $k\ge s$, we have

$$\begin{aligned}&\textstyle\sum_{k=0}^K\alpha_k\sum_{s=1}^k\beta^{k-s}\alpha_{s-1}\le\sum_{k=0}^K\sum_{s=1}^k\beta^{k-s}\alpha_{s-1}^2\\&=\textstyle\sum_{k=0}^K\sum_{s=1}^k\beta^{k-1-(s-1)}\alpha_{s-1}^2=\sum_{k=0}^K\sum_{t=0}^{k-1}\beta^{k-1-t}\alpha_t^2\\&\le\textstyle\sum_{t=0}^{K-1}\left(\sum_{t=0}^K\beta^{t-1}\right)\alpha_t^2\le\frac{1}{\beta(1-\beta)}\sum_{t=0}^{K-1}\alpha_t^2.\end{aligned}$$

This together with (31) produces

$$\textstyle\sum_{k=0}^K\alpha_k e_k\le\frac{\sum_{j=1}^N L_{hj}\sigma_j^{-1}(C_j+\nu_j)}{\beta(1-\beta)}\sum_{t=0}^{K-1}\alpha_t^2.$$

By recalling that $\alpha_k\le\alpha_0$ for all $k\ge0$, we have $\sum_{k=0}^K\alpha_k\beta^k\le\alpha_0\sum_{k=0}^\infty\beta^k\le\frac{\alpha_0}{1-\beta}$. Then by (31), we have

$$\begin{aligned}\textstyle\sum_{k=0}^K\mathbb{E}[\varepsilon_k]\le\theta N\sum_{i=1}^N L_{fi}M_i\Big(&\tfrac{\alpha_0 M_H}{1-\beta}\\&+\tfrac{\sum_{j=1}^N L_{hj}\sigma_j^{-1}(C_j+\nu_j)}{\beta(1-\beta)}\textstyle\sum_{t=0}^{K-1}\alpha_t^2\Big).\end{aligned}\tag{34}$$

Thus, $\sum_{k=0}^\infty\mathbb{E}[\varepsilon_k]<\infty$ and $\sum_{k=0}^\infty\varepsilon_k<\infty,\ a.s.$

Note by Assumption 5 and (4) that

$$(x_k-x^*)^T\phi(x_k)\ge(x_k-x^*)^T\phi(x^*)\ge0,\ \forall k\ge0.\tag{35}$$

By setting $x=x^*$ in (28), applying Lemma 3, using $\sum_{k=0}^\infty\alpha_k^2<\infty$ and $\sum_{k=0}^\infty\varepsilon_k<\infty, a.s.$, we conclude that $D(x^*,x_k)$ converge almost surely, and $\sum_{k=0}^\infty\alpha_k(x_k-x^*)^T\phi(x_k)<\infty,\ a.s.$ The requirement $\sum_{t=0}^\infty\alpha(t)=\infty$ implies that $\liminf_{t\to\infty}(x_k-x^*)^T\phi(x_k)=0$. So, there exists a subsequence $\{t_r\}$ such that $\lim\limits_{r\to\infty}(x_{t_r}-x^*)^T\phi(x_{t_r})=0$. Let $\tilde{x}$ be a limit point of the bounded sequence $\{x(t_r)\}$. Then $(\tilde{x}-x^*)^T\phi(\tilde{x})=0$ by the continuity of $\phi(x)$. Hence $\tilde{x}=x^*$ by (4) and the strict monotonicity of $\phi(x)$. Then by the Bregman reciprocity condition, $D(x^*,x_{t_r})$ converges almost surely to zero. By recalling that $D(x^*,x_k)$ converges almost surely, we obtain $D(x^*,x_k)\xrightarrow[k\to\infty]{a.s.}0$ and $x_k\xrightarrow[k\to\infty]{a.s.}x^*$ by Lemma 2. □

## 5 Numerical Simulations

In this section, we empirically validate the performance of Algorithm 1 on the Nash Cournot competition problem described in Example 1.

Suppose that the capacity constraint of firm $i\in\mathcal{N}$ is a simplex $X_i\triangleq\{x_i\ge0:\mathbf{1}^Tx_i=1\}$. We consider the entropic regularizer $r_i(x_i)=\sum_{j=1}^{m_i}x_{ij}\log(x_{ij})$. Then the pseudo-distance defined by (6) becomes $D_i(x_i,y_i)=\sum_{j=1}^{m_i}x_{ij}\log(x_{ij}/y_{ij})$, which is called the Kullback–Leibler divergence. Then the update (8) becomes

$$[x_{i,k+1}]_j=\frac{\left([x_{i,k}]_j\exp(-\alpha_k[q_i(x_{i,k},N\hat{v}_{i,k+1};\xi_{i,k},\zeta_k)]_j)\right)_{j=1}^{m_i}}{\sum_{j=1}^{m_i}[x_{i,k}]_j\exp(-\alpha_k[q_i(x_{i,k},N\hat{v}_{i,k+1};\xi_{i,k},\zeta_k)]_j)}.$$

This is known as the exponentiated gradient algorithm and has been extensively studied in online learning [12, 16, 20].

Set $N=20, m=10, m_i=3$, and let the interactions among the firms be described by an undirected connected Erdős–Rényi graph $\mathcal{G}=\{\mathcal{N},\mathcal{E}\}$, where each edge connecting two firms is included in the graph with probability 0.2 independent from every other edge. Set the adjacency matrix $W=[w_{ij}]$, where $w_{ij}=\frac{1}{\max\{|\mathcal{N}_i|,|\mathcal{N}_j|\}}$ for any $i\neq j$ with $(i,j)\in\mathcal{E}$, $w_{ii}=1-\sum_{j\neq i}w_{ij}$, and $W_{ij}=0$, otherwise. Suppose that for each firm $i\in\mathcal{N}$, each entry of $c_i$ satisfies the uniform distribution $U[3,4]$. The pricing parameters $d_l,b_l$ of market $l\in\mathcal{M}$ are drawn from uniform distributions $U[4,5]$ and $U[1,1.1]$, respectively. Let the random variables $\xi_i, i\in\mathcal{N}$ and $\zeta_l, l\in\mathcal{M}$ be drawn from uniform distributions $U[-c_i/8,c_i/8]$ and $U[-d_l/8,d_l/8]$, respectively.

We implement Algorithm 1 with $\alpha_k=\frac{1}{\sqrt{k}}$ and display the empirical results in Figs. 1 and 2. Fig. 1 demonstrates that the rescaled regret $\mathrm{VR}_K/\sqrt{K}$ is bounded, indicating that the variational regret $\mathrm{VR}_K=\mathcal{O}(\sqrt{K})$ matching the bound established in Corollary 1. In addition, Fig. 2 shows that the squared error $\max_{i\in\mathcal{N}}\|x_{i,k}-x_i^*\|$ asymptotically decreases to zero, implying that the iterate $\{x_k\}$ asymptotically converges to the Nash equilibrium.

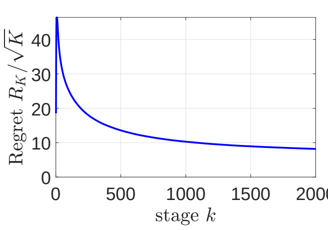


Fig. 1: Bound on the Variational Regret

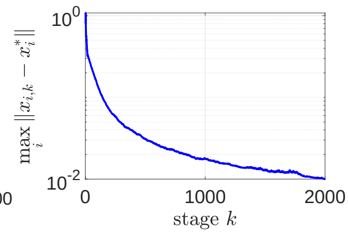


Fig. 2: Asymptotic Convergence to Nash Equilibrium

## 6 Conclusions

This paper proposes a distributed no-regret learning for stochastic aggregative game based on mirror descent. For the class of strictly monotone games, the iterate is shown to converge almost surely to the unique Nash equilibrium with suitably selected diminishing steplengths. It is of interest to extend the no-regret learning methods to the other classes of network games in distributed and stochastic settings.

## References


[1] J. R. Marden and J. S. Shamma, "Game theory and distributed control," in *Handbook of game theory with economic applications*. Elsevier, 2015, vol. 4, pp. 861–899.

[2] T. Alpcan and T. Basar, "A game-theoretic framework for congestion control in general topology networks," in *Proceedings of the 41st IEEE Conference on Decision and Control, 2002.*, vol. 2. IEEE, 2002, pp. 1218–1224.

[3] J. R. Marden and A. Wierman, "Distributed welfare games," *Operations Research*, vol. 61, no. 1, pp. 155–168, 2013.

[4] N. Li and J. R. Marden, "Designing games for distributed optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 230–242, 2013.

[5] J. R. Marden, G. Arslan, and J. S. Shamma, "Cooperative control and potential games," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 6, pp. 1393–1407, 2009.

[6] R.-R. Zhang and L. Guo, "Controllability of Nash equilibrium in game-based control systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 4180–4187, 2019.

[7] Y. Li and X. Hu, "Intrinsic formation of regular polyhedra: A differential game approach," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4748–4753.

[8] J. Ma, Y. Zheng, B. Wu, and L. Wang, "Equilibrium topology of multi-agent systems with two leaders: A zero-sum game perspective," *Automatica*, vol. 73, pp. 200–206, 2016.

[9] M. Ye and G. Hu, "Game design and analysis for price-based demand response: An aggregate game approach," *IEEE transactions on cybernetics*, vol. 47, no. 3, pp. 720–730, 2016.

[10] J. F. Nash, Jr., "Equilibrium points in $n$-person games," *Proc. Nat. Acad. Sci. U. S. A.*, vol. 36, pp. 48–49, 1950.

[11] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a Nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.

[12] M. Bravo, D. Leslie, and P. Mertikopoulos, "Bandit learning in concave n-person games," *Advances in Neural Information Processing Systems*, vol. 31, pp. 5661–5671, 2018.

[13] S. Talebi, S. Alemzadeh, L. J. Ratliff, and M. Mesbahi, "Distributed learning in network games: a dual averaging approach," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5544–5549.

[14] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, no. 3, pp. 680–704, 2016.

[15] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.

[16] P. Mertikopoulos and Z. Zhou, "Learning in games with continuous action sets and unknown payoff functions," *Mathematical Programming*, vol. 173, no. 1, pp. 465–507, 2019.

[17] A. S. Nemirovskij and D. B. Yudin, "Problem complexity and method efficiency in optimization," 1983.

[18] T. T. Doan, S. Bose, D. H. Nguyen, and C. L. Beck, "Convergence of the iterates in mirror descent methods," *IEEE control systems letters*, vol. 3, no. 1, pp. 114–119, 2018.

[19] D. Yuan, Y. Hong, D. W. Ho, and S. Xu, "Distributed mirror descent for online composite optimization," *IEEE Transactions on Automatic Control*, 2020.

[20] Z. Zhou, P. Mertikopoulos, A. L. Moustakas, N. Bambos, and P. Glynn, "Mirror descent learning in continuous games," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 5776–5783.

[21] W. Krichene, S. Krichene, and A. Bayen, "Convergence of mirror descent dynamics in the routing game," in *2015 European Control Conference (ECC)*. IEEE, 2015, pp. 569–574.

[22] B. Gao and L. Pavel, "Continuous-time discounted mirror-descent dynamics in monotone concave games," *IEEE Transactions on Automatic Control*, 2020.

[23] N. Flammarion and F. Bach, "Stochastic composite least-squares regression with convergence rate $o(1/n)$," in *Conference on Learning Theory*. PMLR, 2017, pp. 831–875.

[24] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge university press, 2006.

[25] N. S. Kukushkin, "Best response dynamics in finite games with additive aggregation," *Games and Economic Behavior*, vol. 48, no. 1, pp. 94–110, 2004.

[26] M. K. Jensen, "Aggregative games and best-reply potentials," *Economic theory*, vol. 43, no. 1, pp. 45–66, 2010.

[27] Z. Ma, D. S. Callaway, and I. A. Hiskens, "Decentralized charging control of large populations of plug-in electric vehicles," *IEEE Transactions on control systems technology*, vol. 21, no. 1, pp. 67–78, 2011.

[28] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to Nash equilibria in constrained deterministic mean field control," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3315–3329, 2015.

[29] H. Kebriaei, S. J. Sadati-Savadkoohi, M. Shokri, and S. Grammatico, "Multi-population aggregative games: Equilibrium seeking via mean-field control and consensus," *IEEE Transactions on Automatic Control*, 2021.

[30] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[31] F. Parise, S. Grammatico, B. Gentile, and J. Lygeros, "Distributed convergence to Nash equilibria in network and average aggregative games," *Automatica*, vol. 117, p. 108959, 2020.

[32] D. Gadjov and L. Pavel, "Single-timescale distributed gne seeking for aggregative games over networks via forward-backward operator splitting," *IEEE Transactions on Automatic Control*, 2020.

[33] J. Lei and U. V. Shanbhag, "Linearly convergent variable sample-size schemes for stochastic Nash games: Best-response schemes and distributed gradient-response schemes," in *2018 58th IEEE Conference on Decision and Control*. IEEE, 2018, pp. 3547–3552.

[34] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.

[35] M. Ye and G. Hu, "Distributed Nash equilibrium seeking by a consensus based approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4811–4818, 2017.

[36] Y. Zhu, W. Yu, G. Wen, and G. Chen, "Distributed nash equilibrium seeking in an aggregative game on a directed graph," *IEEE Transactions on Automatic Control*, 2020.

[37] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. I*, ser. Springer Series in Operations Research. New York: Springer-Verlag, 2003.

[38] T. Larsson and M. Patriksson, "A class of gap functions for variational inequalities," *Mathematical Programming*, vol. 64, no. 1-3, pp. 53–79, 1994.

[39] G. Chen and M. Teboulle, "Convergence analysis of a proximal-like minimization algorithm using bregman functions," *SIAM Journal on Optimization*, vol. 3, no. 3, pp. 538–543, 1993.

[40] H.-F. Chen, *Stochastic approximation and its applications*. Springer Science & Business Media, 2006, vol. 64.

[41] B. Polyak, *Introduction to optimization*. New York: Optimization Software, Inc., 1987.