Semidecentralized Zeroth-Order Algorithms for Stochastic Generalized Nash Equilibrium Seeking

Journal Article

Author(s):
Zou, Suli; Lygeros, John

Publication date:
2023-02

Permanent link:
https://doi.org/10.3929/ethz-b-000600540

Originally published in:
IEEE Transactions on Automatic Control 68(2), https://doi.org/10.1109/TAC.2022.3151225

Funding acknowledgement:
787845 - Optimal control at large (EC)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3151225, IEEE Transactions on Automatic Control

Semi-decentralized Zeroth-order Algorithms for Stochastic Generalized Nash Equilibrium Seeking

Suli Zou, *Member, IEEE*, John Lygeros, *Fellow, IEEE*

***Abstract*—In this paper, we address the problem of stochastic generalized Nash equilibrium (SGNE) seeking, where a group of noncooperative heterogeneous players aim at minimizing their expected cost under some unknown stochastic effects. Each player's strategy is constrained to a convex and compact set and should satisfy some global affine constraints. In order to decouple players' strategies under the global constraints, an extra player is introduced aiming at minimizing the violation of the coupling constraints, which transforms the original SGNE problems to extended stochastic Nash equilibrium problems. Due to the unknown stochastic effects in the objective, the gradient of the objective function is infeasible and only noisy objective values are observable. Instead of gradient-based methods, a semi-decentralized zeroth-order method is developed to achieve the SGNE under a two-point gradient estimation. The convergence proof is provided for strongly monotone stochastic generalized games. We demonstrate the proposed algorithm through the Cournot model for resource allocation problems.**

***Index Terms*—Stochastic generalized Nash equilibrium, unknown stochastic effects, semi-decentralized zeroth-order algorithm, gradient estimation, convergence.**

I. Introduction

Generalized Nash Equilibrium (GNE) problems have received extensive attention due to their applications in engineering fields, for instance, energy management, intelligent transportation, telecommunication, machine learning and cloud computing, to name a few [1]–[3]. This has motivated research on Nash equilibrium (NE) seeking algorithms, usually associated with the solutions of variational inequality (VI) problems [4]. Several methods to solve GNE problems have been developed, such as the projection method [1], penalty-based method [5], augmented Lagrangian [6], ADMM [7] and other operator-splitting methods [8], [9]. All of these methods are (sub)gradient-based approaches which are commonly used in distributed optimization problems.

In many practical scenarios, stochastic effects affect players' payoffs. As a result, players usually aim at optimizing the payoffs in the sense of expectation [10], [11], leading to stochastic games. However, the form of individual objective functions or their derivative/gradient may not be known; this is the case, for example, if the stochastic effects are unknown, or computation of gradients is computationally demanding. Moreover, when the stochastic variable changes greatly in practice, there will be a large deviation between the expected solution and the real solution. One strategy players can adopt in such cases may be to take decisions based on the rough estimation of the payoff performance at some played strategies [12]. The chosen alternative is then played, its payoff is observed and used to update the estimation for that particular strategy. This procedure is repeated, generating a discrete time stochastic process which is called the learning process. Although players observe only their own payoffs, these values are affected by the choices of other players, revealing information on the game as a whole. The challenge is to determine when a learning procedure based on payoff observations can induce convergence to an equilibrium.

Current works focus on developing learning procedures for stochastic NE seeking, notably using reinforcement learning (RL) [13]. Although solving game theory problems by RL is not a new idea [14], the available results focus mostly on problems with finite action sets. Some works proposed RL algorithms to solve problems with continuous action sets [15], [16] wherein continuous action learning automata (CALA) presented in [15] could be extended to solve NE problems. The results, however, rely on stochastic approximation [17], [18] and require a boundedness assumption on the second moment of the stochastic gradient. Recent works have studied zeroth-order (ZO) algorithms [19], closely related to CALA. An added benefit of ZO algorithms is that they can be directly applied to non-smooth problems, as they do not require the computation of gradients; for this reason they are also known as "gradient-free" methods.

The performance of ZO algorithms for convex problems have been well studied in the centralized case [19], [20]. The work in [21] integrated the idea of zeroth-order optimization with online ADMM and compared the convergence rate with first-order online algorithms. In [22], [23], the distributed problems treated are essentially consensus problems, where the payoff of the players is not directly affected by decisions of other players. In [24], a ZO algorithm has been adapted to solve stochastic NE problems and prove convergence for strongly monotone games and generalized potential games. The results were subsequently extended to monotone games in [25], wherein the player's cost function is constrained not to grow faster than linear functions.

Here, we consider a general class of stochastic generalized Nash equilibrium (SGNE) problems with unknown stochastic objectives and affine coupling constraints. We first deal with the coupling constraints by transforming the SGNE problem to an extended SNE problem by introducing an extra dual player whose goal is to minimize the violation of the coupling constraints. We propose a semi-decentralized zeroth-

Suli Zou is with the School of Automation, Beijing Institute of Technology, Beijing 100081, China. John Lygeros is with the Automatic Control Laboratory, ETH Zürich, Physikstrasse 3, 8092 Zürich, Switzerland. Emails: sulizou@bit.edu.cn, lygeros@control.ee.ethz.ch.

order (DZO) algorithm to achieve the SGNE of a class of strongly monotone games, regarding the dual player as a coordinator which is similar to the structure in [1]. Unlike the existing literature, we adopt two-point sampling to estimate the gradient of the objective function, and introduce an averaging process to reduce the variance of the stochastic effects. To ensure the convergence to the neighborhood of SGNE for strongly monotone stochastic generalized games, we introduce a regularization term to the dual update based on the idea of Tikhonov regularization in [26], [27].

In section II, we give a standard formulation of the stochastic generalized game and SGNE, and introduce an extended game for SGNE for tackling coupling constraints and enforcing distributed optimization. Then, a semi-decentralized zeroth-order algorithm is presented to solve the extended game in Section III. We give in Section IV application scenarios of resource allocation to demonstrate our results. Finally, Section V draws conclusions and future work.

II. Formulation of Game Model

In this paper, we use $\mathbb{R}^d$ to represent the space of d dimensional vectors with $\langle\cdot,\cdot\rangle : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ denoting the standard inner product and $\|\cdot\|$ the corresponding norm. We distinguish vectors by boldface from scalars ($d = 1$). $\mathbb{R}^d_{\geq 0}$ represents the set of d dimensional vectors with non-negative coordinates. If $\mathcal{N} \triangleq \{1, \cdots, N\}$ is a set of indexes, given N vectors $\boldsymbol{x}_i \in \mathbb{R}^d$ for all $i \in \mathcal{N}$, we denote $\boldsymbol{x} \triangleq [\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_N^\top]^\top \in \mathbb{R}^{Nd}$ and $\boldsymbol{x}_{-i} \triangleq [\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_{i-1}^\top, \boldsymbol{x}_{i+1}^\top, \cdots, \boldsymbol{x}_N^\top]^\top \in \mathbb{R}^{(N-1)d}$. The operator $\Pi_{\mathcal{X}}\|\boldsymbol{x}\| : \mathbb{R}^d \to \mathbb{R}^d$ denotes the projection of the vector $\boldsymbol{x}$ into the closed set $\mathcal{X}$.

A. Stochastic Generalized Game

Consider N players indexed by $\mathcal{N}$. Suppose that the players choose strategies $\boldsymbol{u}_i \in \mathbb{R}^d, i \in \mathcal{N}$ to minimize individual costs, which depend on the strategies of all the players. Denote $\mathcal{U}_i \subset \mathbb{R}^d$ the local constraint set of player i; we require that $\boldsymbol{u}_i \in \mathcal{U}_i$. In addition, assume that the strategies of all the players are constrained by an affine coupling constraint

$$\boldsymbol{u} \in \mathcal{C} \triangleq \{\boldsymbol{u} \in \mathbb{R}^{Nd} \,|\, A\boldsymbol{u} \leq b\}, \tag{1}$$

where $A \in \mathbb{R}^{m\times Nd}$ and $b \in \mathbb{R}^m$. Hence, the strategy set of player i, denoted by $\mathcal{Q}_i$, is

$$\mathcal{Q}_i(\boldsymbol{u}_{-i}) = \{\boldsymbol{u}_i \in \mathcal{U}_i \,|\, (\boldsymbol{u}_i, \boldsymbol{u}_{-i}) \in \mathcal{Q}\}, \tag{2}$$

where $\mathcal{Q} \triangleq \mathcal{U} \cap \mathcal{C}$ with $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_N$.

The individual cost of player i is assumed to be affected by uncertainty ξ_i, and is denoted $F_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}; \xi_i)$. In a deterministic game, ξ_i is not present and each player aims at minimizing their real costs. In a stochastic game on the other hand, we assume that, players try to optimize their expected cost over the realizations of ξ_i. Hence, the objective of each player is to find an optimal strategy to minimize the expected cost with respect to other players' strategies,

$$\boldsymbol{u}_i^* = \underset{\boldsymbol{u}_i \in \mathcal{Q}_i(\boldsymbol{u}_{-i})}{\arg\min}\ f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}) \text{ where } f_i \triangleq \mathbb{E}_{\xi_i}[F_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}; \xi_i)].$$

The following assumption is imposed:

Assumption 1. *The set $\mathcal{C}$ satisfies Slater's constraint qualification. Moreover, for each player $i \in \mathcal{N}$,*

- *The set $\mathcal{U}_i$ is compact and convex, and satisfies Slater's constraint qualification.*
- *The expected cost $f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i})$ is convex w.r.t. $\boldsymbol{u}_i$, for each $\boldsymbol{u}_{-i}$, and continuously differentiable in $\boldsymbol{u}$.*

We are interested in characterising the SGNE of the resulting stochastic generalized game

$$\mathcal{G} = \langle \mathcal{N}, \{\boldsymbol{u}_i\}_{i\in\mathcal{N}}, \{\mathcal{Q}_i\}_{i\in\mathcal{N}}, \{f_i\}_{i\in\mathcal{N}} \rangle .$$

Definition 1 (Stochastic Generalized Nash Equilibrium)**.** *A strategy profile $\boldsymbol{u}^*$ is called a SGNE of the game $\mathcal{G}$ if*

$$f_i(\boldsymbol{u}_i^*, \boldsymbol{u}_{-i}^*) \leq f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}^*), \quad \forall \boldsymbol{u}_i \in \mathcal{Q}_i(\boldsymbol{u}_{-i}^*), \forall i \in \mathcal{N}.$$

We denote the set of SGNEs for game $\mathcal{G}$ by SGNE($\mathcal{G}$).

The solutions of SGNE problems are closely related to solutions of variational inequality (VI) problems. Under Assumption 1, the VI problem is defined for game $\mathcal{G}$, denoted by VI($\mathcal{Q}, M$), where the operator $M(\boldsymbol{u}) = [M_1(\boldsymbol{u}); \cdots ; M_N(\boldsymbol{u})] : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$ maps a strategy profile $\boldsymbol{u}$ to the stacked gradients of the cost functions of each player with respect to their own strategies,

$$M_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}) = \frac{\partial f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i})}{\partial \boldsymbol{u}_i}. \tag{3}$$

We impose an assumption on the operator M.

Assumption 2. *The operator M defined by* (3) *is τ-strongly monotone and L_1-Lipschitz continuous in $\mathcal{Q}$.*

If we define the set of solutions to VI($\mathcal{Q}, M$) as SOL($\mathcal{Q}, M$), then SOL($\mathcal{Q}, M$) $\subset$ SGNE($\mathcal{G}$), i.e. a solution to VI($\mathcal{Q}, M$) is also a SGNE of $\mathcal{G}$, under Assumption 1 [4][Theorem 5]. Here, we are only interested in equilibria in SOL($\mathcal{Q}, M$), known as variational SGNE (VSGNE).

Lemma II.1 (Existence & Uniqueness of VSGNE)**.** *Under Assumptions 1 and 2, VI($\mathcal{Q}, M$) has a unique solution [26][Theorem 2.3.3].*

B. Equivalent Extended Game

In the literature, an extended game has been proposed to solve the GNE problems by introducing an extra player to enforce the coupling constraint. By doing this, the original GNE problem is transformed to a NE problem and the resulting NE is equivalent to the original GNE, [24]. Such a technique can be naturally extended to SGNE problems.

Consider an additional player 0, referred to as the dual player, with strategy λ, whose objective is related to the coupling constraint specified in (1)

$$\boldsymbol{\lambda}^* = \underset{\boldsymbol{\lambda}\in\mathbb{R}^m_{\geq 0}}{\arg\min}\ f_0^0(\boldsymbol{\lambda}, \boldsymbol{u}) \triangleq -\boldsymbol{\lambda}^\top(A\boldsymbol{u} - b). \tag{4}$$

To ensure that the coupling constraint is satisfied, it should be incorporated into the cost functions of the original N players, referred to as primal players. That is, for each $i \in \mathcal{N}$,

$$\boldsymbol{u}_i^* = \underset{\boldsymbol{u}_i\in\mathcal{U}_i}{\arg\min}\ f_i^0(\boldsymbol{u}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda}) \triangleq f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}) - f_0^0(\boldsymbol{\lambda}, \boldsymbol{u}). \tag{5}$$

This leads to the extended game

$$\mathcal{G}^0 = \left\langle \mathcal{N}^0, \{\boldsymbol{u}_i\}_{i\in\mathcal{N}} \cup \{\boldsymbol{\lambda}\}, \{\mathcal{Q}_i\}_{i\in\mathcal{N}} \cup \mathbb{R}^m_{\geq 0}, \{f_i^0\}_{i\in\mathcal{N}^0} \right\rangle,$$

where $\mathcal{N}^0 \triangleq \mathcal{N} \cup \{0\}$. Compared with the game $\mathcal{G}$, each of the cost functions of the primal players has an additional term that depends on the strategy of the dual player, which indirectly encodes the coupling constraints. Note, however, that the only constraints remaining in $\mathcal{G}^0$ are the local constraints of each player.

In the interest of brevity we use $\boldsymbol{\eta} = (\boldsymbol{u}; \boldsymbol{\lambda}) \in \mathbb{R}^{Nd+m}$ to denote the strategy profile of the extended game. Then define an operator $M^0(\boldsymbol{\eta}) = [M_1^0(\boldsymbol{\eta}); \cdots; M_N^0(\boldsymbol{\eta}); M_0^0(\boldsymbol{\eta})] : \mathbb{R}^{Nd+m} \to \mathbb{R}^{Nd+m}$, a stacked gradient mapping for the extended game $\mathcal{G}^0$,

$$\begin{aligned} M_i^0(\boldsymbol{\eta}) &= \frac{\partial f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i})}{\partial \boldsymbol{u}_i} + A_i^\top \boldsymbol{\lambda}, \quad \forall i \in \mathcal{N} \\ M_0^0(\boldsymbol{\eta}) &= -(A\boldsymbol{u} - b), \end{aligned} \tag{6}$$

where $A_i \in \mathbb{R}^{m\times d}$ is the i-th component block of A. Under this definition, we derive that for any two strategies of all the players $\boldsymbol{\eta}^1, \boldsymbol{\eta}^2 \in \mathbb{R}^{Nd+m}$,

$$\begin{aligned} &\left\langle M^0(\boldsymbol{\eta}^1) - M^0(\boldsymbol{\eta}^2), \boldsymbol{\eta}^1 - \boldsymbol{\eta}^2 \right\rangle \\ &= \left\langle M(\boldsymbol{u}^1) - M(\boldsymbol{u}^2), \boldsymbol{u}^1 - \boldsymbol{u}^2 \right\rangle. \end{aligned} \tag{7}$$

Lemma II.2. *Under Assumptions 1 and 2, $M^0(\boldsymbol{u}, \boldsymbol{\lambda})$ defined in* (6) *is a monotone operator.*

Proof. By the monotonicity of M and (7), we have the conclusion. □

The following lemma gives some results of the extended game $\mathcal{G}^0$ and connect them with the original game $\mathcal{G}$.

Lemma II.3 ([4])**.** *Under Assumptions 1 and 2 for game $\mathcal{G}$, we have*

- *$\boldsymbol{\eta}^*$ is a stochastic Nash equilibrium (SNE) of $\mathcal{G}^0$ if and only if $\boldsymbol{\eta}^* \in SOL(\mathcal{U} \times \mathbb{R}^m_{\geq 0}, M^0)$;*
- *if $\boldsymbol{\eta}^*$ is a SNE of $\mathcal{G}^0$, then $\boldsymbol{u}^*$ is a SGNE of $\mathcal{G}$;*

Therefore we can achieve the SGNE of game $\mathcal{G}$ by computing the SNE of game $\mathcal{G}^0$, which could in principle be determined gradient-based, e.g. [1], [8]. In many practical situations, however, the forms of the objective functions may be unknown even to the players themselves, since the probability distribution of ξ_i (or, depending on F_i, its moments) is difficult to be aware of. This motivates a gradient-free algorithm wherein players make decisions based on observations of objective values at a given strategy. In the following, we develop a semi-decentralized zeroth-order algorithm to solve such problems.

III. Semi-decentralized Zeroth-order Algorithm

A. Background: DZO algorithms

Before giving the proposed algorithm, we provide some basic knowledge about ZO algorithms applied to the extended game $\mathcal{G}^0$ analyzed above.

The centralized ZO algorithm is analyzed via the idea of Gaussian approximation [19], which can be easily generalized to the distributed cases by introducing other players' strategies. Suppose that the index k denotes the iteration of the ZO algorithm and vectors with subscript k represent the corresponding values of the vectors at iteration k. Given the following Gaussian distribution,

$$\boldsymbol{x}_{i,k} = \boldsymbol{u}_{i,k} + \sigma_k \cdot \boldsymbol{v}_{i,k}, \text{ with } \boldsymbol{v}_{i,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \tag{8}$$

the Gaussian approximation of $f_i^0(\cdot)$ with parameter σ_k is defined as

$$\begin{aligned} &f^0_{i,\sigma_k}(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) \\ &\triangleq \int_{\mathbb{R}^d} f_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)\, p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2 \mathbf{I}_d) d\boldsymbol{x}_{i,k}, \end{aligned} \tag{9}$$

where $p(\cdot)$ denotes the probability density function of an $\mathbb{R}^d$ valued random variable with mean $\boldsymbol{u}_{i,k}$ and covariance matrix $\sigma_k^2 \mathbf{I}_d$.

Taking the partial derivative of $f^0_{i,\sigma_k}(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)$ with respect to $\boldsymbol{u}_{i,k}$, we derive the stochastic gradient of the Gaussian approximation of the original objective function, denoted by $\tilde{M}_i^0(\boldsymbol{\eta}_k)$, i.e.,

$$\begin{aligned} \tilde{M}_i^0(\boldsymbol{\eta}_k) &\triangleq \frac{\partial f^0_{i,\sigma_k}(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)}{\partial \boldsymbol{u}_{i,k}} \\ &= \int_{\mathbb{R}^d} f_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) \cdot \frac{\boldsymbol{x}_{i,k} - \boldsymbol{u}_{i,k}}{\sigma_k^2} \\ &\qquad\qquad p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2 \mathbf{I}_d) d\boldsymbol{x}_{i,k} \\ &= \int_{\mathbb{R}^d} \Big(f_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) - f_i^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) \Big) \\ &\qquad \frac{\boldsymbol{x}_{i,k} - \boldsymbol{u}_{i,k}}{\sigma_k^2} \cdot p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2 \mathbf{I}_d) d\boldsymbol{x}_{i,k}. \end{aligned} \tag{10}$$

The last equality holds since

$$\int_{\mathbb{R}^d} \frac{\boldsymbol{x}_{i,k} - \boldsymbol{u}_{i,k}}{\sigma_k^2} \cdot p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2 \mathbf{I}_d) d\boldsymbol{x}_{i,k} = 0, \tag{11}$$

which is $1/\sigma_k$ of the expectation of $\boldsymbol{v}_{i,k}$.

Assumption 3. *The expected cost function $f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}), i \in \mathcal{N}$ grows no faster than a finite-degree polynomial function of $\boldsymbol{u}_i$ as $\|\boldsymbol{u}_i\| \to \infty$.*

Lemma III.1. *Under Assumptions 1-3, the following holds for each $i \in \mathcal{N}$*

$$\tilde{M}_i^0(\boldsymbol{\eta}_k) = \int_{\mathbb{R}^d} M_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2 \mathbf{I}_d) d\boldsymbol{x}_{i,k} \tag{12}$$

Proof. The proof is provided in Appendix A. □

B. Primal updates: Averaging process

Compared to this standard DZO algorithm, our game includes additional unknown stochastic effects. In this case, we can only obtain noisy observations of the stochastic cost for each primal player

$$F_i^0(\boldsymbol{u}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda}; \xi_i) = F_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}; \xi_i) + \boldsymbol{\lambda}^\top (A\boldsymbol{u} - b),$$

for some realization ξ_i of the uncertainty.

The algorithm proposed in [24] used the current noisy observation, as mentioned in [19], as an estimate of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3151225, IEEE Transactions on Automatic Control

gradient; here we develop an alternative approximation based on averaging. To this end, the following assumption on the second moment of stochastic effects.

Assumption 4. *The second moment of the stochastic effects on the function value is finite, i.e., there exists a constant D_1, such that*

$$\mathrm{Var}[F_i^0(\cdot;\xi)] = \mathbb{E}_\xi[(F_i^0(\cdot;\xi) - f_i^0(\cdot))^2] \le D_1, \forall i \in \mathcal{N}. \tag{13}$$

We compute an approximation g_i^0 of the pseudo-gradient

$$g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k}) = \frac{\tilde{f}_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) - \hat{f}_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)}{\sigma_k} \times \boldsymbol{v}_{i,k}, \tag{14}$$

where $\tilde{f}_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)$, $\hat{f}_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)$ are estimations of $f_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)$ and $f_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)$ respectively, based, respectively, on T_1 and T_2 noisy observations,

$$\tilde{f}_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) = \frac{1}{T_1}\sum_{t_1=1}^{T_1} F_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k; \xi_{i,k,t_1}), \tag{15}$$

$$\hat{f}_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) = \frac{1}{T_2}\sum_{t_2=1}^{T_2} F_i^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k; \xi_{i,k,T_1+t_2}). \tag{16}$$

We then update the strategy of the primal players by

$$\boldsymbol{u}_{i,k+1} = \Pi_{\mathcal{U}_i}\|\boldsymbol{u}_{i,k} - h_k \cdot g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k})\|, \quad \forall i \in \mathcal{N}, \tag{17}$$

where h_k is a step-size. (10) can be estimated as

$$\tilde{M}_i^0(\boldsymbol{\eta}_k) \approx \int_{\mathbb{R}^d} g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k}) \cdot p(\boldsymbol{x}_{i,k}|\boldsymbol{u}_{i,k}, \sigma_k^2\mathbf{I}_d)d\boldsymbol{x}_{i,k}. \tag{18}$$

C. Dual updates: Regularizing term

Even though the operator M for the game $\mathcal{G}$ is assumed to be strongly monotone, the operator M^0 for the extended game $\mathcal{G}^0$ is merely monotone due to the introduction of the dual player. A convergent algorithm can nonetheless be developed by the so-called Tikhonov regularization [26], [27], where a strongly monotone operator is constructed that tends to the original monotone one as the iteration goes to infinity. As the loss of strong monotonicity comes from the dual variable $\boldsymbol{\lambda}$, we only add regularization to the dual update,

$$\boldsymbol{\lambda}_{k+1} = \Pi_{\mathbb{R}^m_{\ge 0}}\|\boldsymbol{\lambda}_k + h_k \cdot (A\boldsymbol{u}_k - b - r_k\boldsymbol{\lambda}_k)\|, \tag{19}$$

where $r_k > 0$ is the regularizing parameter that diminishes as the iteration k increases.

Therefore, instead of solving a VI problem of the monotone operator M^0, at each iteration k we solve a VI problem with the new mapping $M_{reg,k}^0$ defined as

$$M_{reg,k}^0(\boldsymbol{\eta}_k) = M^0(\boldsymbol{\eta}_k) + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ r_k \cdot \boldsymbol{\lambda}_k \end{bmatrix}, \tag{20}$$

which is strongly monotone with parameter $\min\{\tau, r_k\}$, based on Assumption 2 and (7). Note that as $r_k \downarrow 0$, $M_{reg,k}^0$ converges to $M^0(\boldsymbol{\eta})$.

The solution sequence $\{\boldsymbol{y}_k\}$ of the variational inequality $\mathrm{VI}(\mathcal{U} \times \mathbb{R}^m_{\ge 0}, M_{reg,k}^0)$, say $\boldsymbol{y}_k \in \mathrm{SOL}(\mathcal{U} \times \mathbb{R}^m_{\ge 0}, M_{reg,k}^0(y))$, is called the Tikhonov sequence.

Lemma III.2 (Theorem 12.2.3 in [26])**.** *Under Assumptions 1 and 2, $\{\boldsymbol{y}_k\}$ exists and is unique for each k. Moreover, for $r_k \downarrow 0$, $\{\boldsymbol{y}_k\}$ is uniformly bounded by a constant M_y and converges to the least norm solution to VI($\mathcal{U} \times \mathbb{R}^m_{\ge 0}, M^0$).*

Lemma III.3. *Under Assumptions 1 and 2, if the sequence r_k is decreasing then*

$$\|\boldsymbol{y}_k - \boldsymbol{y}_{k-1}\| \le \frac{|r_{k-1} - r_k|}{\kappa r_k} M_{\boldsymbol{y},\boldsymbol{\lambda}}, \quad \forall k \ge 1. \tag{21}$$

where $M_{\boldsymbol{y},\boldsymbol{\lambda}}$ is a norm bound on the dual term in the Tikhonov sequence and $\kappa = \min\{\frac{\tau}{r_0}, 1\}$.

Proof. The proof is similar to the proof of [28][Lemma 3] based on (7) and Lemma II.2, but somewhat tighter in our case. □

Note that we do not define a Gaussian approximation for player 0 because the dual player can obtain exact gradient information and does not need the gradient estimation.

D. ZO Algorithm for Stochastic Game

To summarize our algorithm for the stochastic game $\mathcal{G}^0$, the pseudocode for our algorithm is written as follows.

Algorithm 1 Regularized DZO (RDZO) Algorithm

Require:
  Initialize $k \leftarrow 0$, $\boldsymbol{u}_{i,k}, \forall i \in \mathcal{N}$ and $\boldsymbol{\lambda}_k$;
  The number of iterations K;
while $k < K$ **do**
  $\boldsymbol{\eta}_k = [\boldsymbol{u}_{1,k}; \cdots; \boldsymbol{u}_{N,k}; \boldsymbol{\lambda}_k]$;
  $h_k = \frac{1}{(k+1)^a}$, $\sigma_k = \frac{1}{(k+1)^b}$, $r_k = \frac{1}{(k+1)^c}$;
  for $i \in \mathcal{N}$ **do**
    Generate

$$\boldsymbol{v}_{i,k} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I}_d), \ \boldsymbol{x}_{i,k} = \boldsymbol{u}_{i,k} + \sigma_k \cdot \boldsymbol{v}_{i,k}$$
$$F_i^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k; \xi_{i,k,t_1}), \forall t_1 \in \{1, \cdots, T_1\},$$
$$F_i^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k; \xi_{i,k,T_1+t_2}), \forall t_2 \in \{1, \cdots, T_2\}$$

    $\boldsymbol{u}_{i,k+1} = \Pi_{\mathcal{X}_i}\|\boldsymbol{u}_{i,k} - h_k g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k})\|$ with g_i^0 specified in (14); {Primal update}
  end for
  $\boldsymbol{\lambda}_{k+1} = \Pi_{\mathbb{R}^m_{\ge 0}}\|\boldsymbol{\lambda}_k + h_k \cdot (A\boldsymbol{u}_k - b - r_k\boldsymbol{\lambda}_k)\|$; {Dual update}
  $k \leftarrow k + 1$
end while
return $\boldsymbol{u}_{i,k}, \forall i \in \mathcal{N}$ and $\boldsymbol{\lambda}_k$

Theorem III.1. *Assume that Assumptions 1-4 hold. If set the step size $h_k = \frac{1}{k^a}$, Gaussian smoothing parameter $\sigma_k = \frac{1}{k^b}$ and relugarized term $r_k = \frac{1}{k^c}$ with $0 < a, b, c < 1$, such that*

$$\sum_{k=0}^{\infty} h_k r_k = \infty, \sum_{k=0}^{\infty} h_k^2 < \infty, \sum_{k=0}^{\infty} h_k \sigma_k < \infty, \sum_{k=0}^{\infty} \frac{h_k^2}{\sigma_k^2} < \infty.$$

Then, the decision variable $\boldsymbol{\eta}_k$ converges to a Nash Equilibrium $\boldsymbol{\eta}^* = [\boldsymbol{u}^*; \boldsymbol{\lambda}^*]$ of the extended game $\mathcal{G}^0$ and $\boldsymbol{u}^*$ is the unique VSGNE of the original game $\mathcal{G}$.

Proof. The proof is given in Appendix B. □

For example, the values of a, b, c could be set as $a = 0.85, b = 0.3, c = 0.15$ respectively, which satisfy the conditions in Theorem III.1. We also adopt these values in the simulation.

Corollary III.1. *The convergence rate of Algorithm 1 is related to the magnitude of stochastic effects D_1, the bound D_2 of $\|A(\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k})\|$, the number of noisy observations T_1 and T_2, the Lipschitz parameters L_1, the strongly monotone parameter τ and the dimension of the decision variable d.*

Proof. The proof is given in Appendix C. □

Compared with using a single noisy observation for estimation in [19], [24], two-point estimation, even though the cost values at two points are affected by different stochastic effects, would achieve a better performance, i.e. reducing the variance of the algorithm. This observation can be explained from the point of view of variance reduction in the stochastic gradient descent (SGD). Equation (14) can be rewritten as follows,

$$g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k})$$

$$= \frac{\tilde{f}_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) - f_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)}{\sigma_k} \cdot \boldsymbol{v}_{i,k} \tag{22}$$

$$+ \frac{f_{i,k}^0(\boldsymbol{x}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) - f_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)}{\sigma_k} \cdot \boldsymbol{v}_{i,k} \tag{23}$$

$$+ \frac{f_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k) - \hat{f}_{i,k}^0(\boldsymbol{u}_{i,k}, \boldsymbol{u}_{-i,k}, \boldsymbol{\lambda}_k)}{\sigma_k} \cdot \boldsymbol{v}_{i,k}. \tag{24}$$

Here, (22), (24) are related to the stochastic effects and disappear under the expectation $\mathbb{E}_{\xi_i}[g_i^0(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k})]$. Since (23) reduces to 0 when the algorithm converges, the variance of the SGD interpretation (18) relies only on the second moment of stochastic effects, which can be regulated by the averaging processes of (15) and (16).

IV. Application Scenarios & Simulation Results

To demonstrate the efficacy of the proposed approach, we consider a resource allocation problem where a set of factories produce a bundle of goods and compare the simulation results of the proposed RDZO algorithm with the algorithm developed in [24], which we refer to as PBZO. Moreover, we consider a case that the distribution of the stochastic effects is known, under which the SNE problem turns into a deterministic one by taking the expectation of player's cost and the gradient information could be obtained. The NE of such a game can be implemented by applying a gradient-based algorithm. We also compare the value of this case with above two algorithms and refer to it as "Expectation".

Consider N factories that manufacture d different products and let $\boldsymbol{u}_i \in \mathbb{R}^d$ denote the production of the i-th factory. Each factory has a maximum production capacity $C_i \in \mathbb{R}_+^d$, leading to local constraints of the form $\boldsymbol{u}_i \in [0, C_i]$. Moreover, limited demand induces coupling constraints among the productions of the form $\mathcal{C} = \{\boldsymbol{u} | A\boldsymbol{u} \le \boldsymbol{b}, A \in \mathbb{R}^{d \times Nd}, \boldsymbol{b} \in \mathbb{R}^d\}$, where $A = [\mathbf{I}_d| \cdots |\mathbf{I}_d]$ is the concatenation of N identity matrices and, in the notation of earlier sections, $m = d$. For the cost of each factory, we use a quadratic production cost and linear price function with respect to the production, which are standard assumptions in Cournot game [24]. Therefore, the cost function for each player i is expressed by the difference between the production cost and the price times production

$$f_i(\boldsymbol{u}_i, \boldsymbol{u}_{-i}) = \boldsymbol{u}_i^\top \boldsymbol{u}_i - 2\underbrace{\Big(\boldsymbol{p}_0 - \frac{1}{N}\sum_{j \in \mathcal{N}} \boldsymbol{u}_j\Big)^\top}_{p(\boldsymbol{u})} \boldsymbol{u}_i, \tag{25}$$

where $p(\boldsymbol{u})$ is the linear price function related to the production of all the factories and $\boldsymbol{p}_0 \in \mathbb{R}_{>0}^d$ is the upper bound price.

Regarding the stochastic effects, the cost of each factory is affected by some unexpected disturbances, like bad weather or delivery losses. For the simplicity of simulation, all stochastic effects are represented by an additive Gaussian noise to the cost function f_i with the distribution $\mathcal{N} \sim (\xi_u, \xi_s)$.

A. Numerical results

For the initialization of our simulation, we set $N = 8$, $d = 4$, $C_i = \mathbf{10}$. $\boldsymbol{b}$ is chosen as $[50 \quad 95 \quad 67 \quad 110]$ and each entry of $\boldsymbol{p}_0$ is chosen as 12. The number of iterations is set to $K = 50000$. In the following, we consider 4 cases for simulation.

Case 1: The distribution of stochastic effects is known.

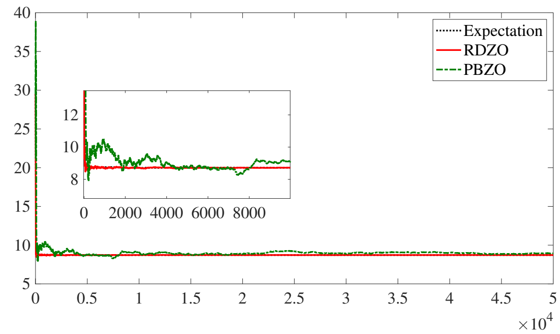
In this case, assume that the distribution of the stochastic effects is known and set $\xi_u = 8, \xi_s = 4$. Taking the expectation of the cost function turns the stochastic optimization into a deterministic one. Then the approximation g_i^0 is obtained by using the true expected cost instead of noisy observations in (15) and (16). At each iteration, as specified in Algorithm 1, it generates an $x_{i,k}$ randomly for computing g_i^0. If we implement the proposed algorithm for multiple times, the iteration process may be different even though all the settings are the same. Here we give two implementations in Fig. 1 to show the evolution of the RDZO and PBZO respectively.

From the Fig. 1(a), both RDZO and PBZO converge to the neighborhood of the SGNE, while PBZO has drifting issues that oscillates near the equilibrium. Moreover, comparing Fig. 1(a) and Fig. 1(b), we find the practical performance of PBZO varies a lot even under the same game setting, while RDZO does not have such an issue. In order to explore this phenomenon, we repeat RDZO and PBZO for 100 times under the same setting and record the average value at each iteration, as shown in Fig. 2. For purpose of simplification, the number of iteration is set to be 5000 in these repetitions.

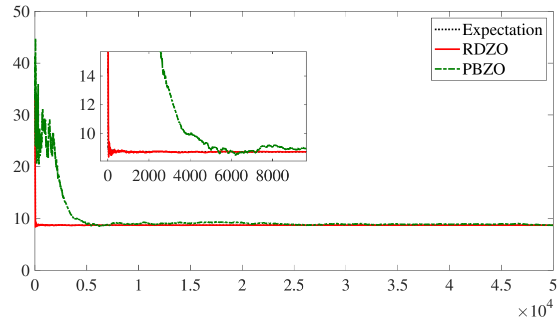
Case 2: Unknown stochastic effects without averaging process.

We generate the unknown stochastic effects by adding some random noise to the expected cost. In this case, we only observe the noisy cost once, i.e. $T_1 = T_2 = 1$. If the added random noise level is relatively small, RDZO still works better than PBZO, as shown in Fig. 3. When the noise level is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3151225, IEEE Transactions on Automatic Control



(a) An implementation of RDZO and PBZO



(b) Another implementation of RDZO and PBZO under the same setting

Fig. 1. Known stochastic effect.

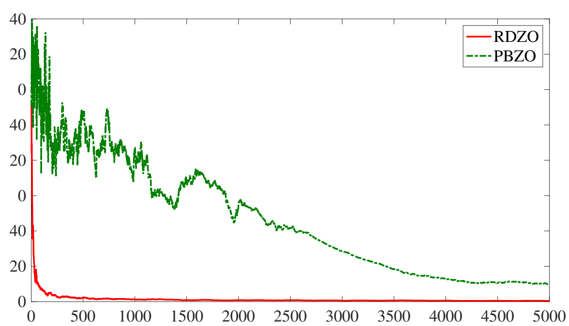


Fig. 2. Average performance under the same game setting.

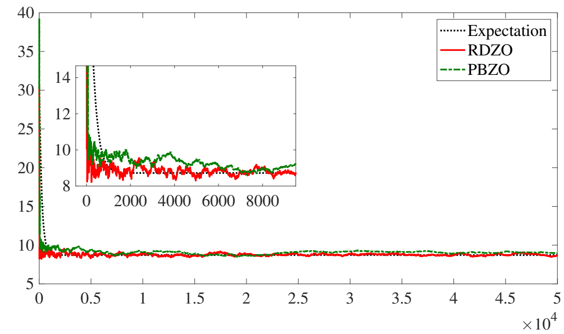


Fig. 3. Low noise level, no averaging.

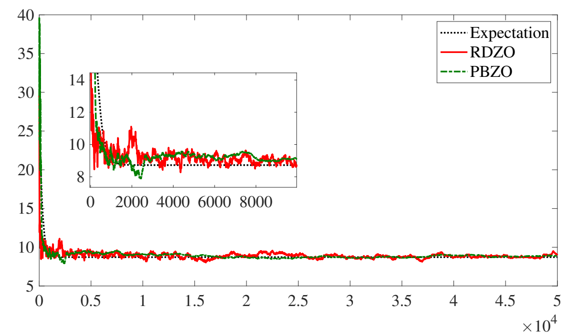


Fig. 4. High noise level, no averaging.

increased, RDZO also suffers from the drifting issue and may be worse than PBZO as shown in Fig. 4 that gives the results when the noise value is comparable with the value of the optimal cost.

Case 3: Unknown stochastic effects with averaging process (20 times).

If we add averaging process to the RDZO algorithm, e.g. $T_1 = T_2 = 20$, the drifting issue is alleviated dramatically, by comparing the red curves in Fig. 4 and Fig. 5.

Figure 5, on the other hand, shows that the drifting issue of PBZO is not alleviated even when the averaging process is applied. This is consistent to the theoretical analysis that the variance of the gradient estimation in PBZO depends also on the magnitude of player's cost.

Case 4: Scalability.

We increase the number of the players to $N = 30$ and keep the other game settings the same. From Fig. 6, RDZO converges to SGNE after thousands of iterations while PBZO seems not to converge. In fact, PBZO would also converge but it needs millions of iterations.

V. Conclusions and Future work

We studied the Nash equilibrium seeking for stochastic generalized games wherein the objective functions of players are affected by some unknown uncertainties. This results in unknown form and gradients of the objectives, which makes the existed gradient-based method inapplicable. Hence, a semi-decentralized zeroth-order algorithm is presented with the idea of estimating the gradient by observing noisy cost values. Different from the methods in the literature, the proposed method applies a two-point estimation and reduces the stochasticity by introducing the averaging process. Moreover, we add a regularized term to the dual update to ensure the convergence to the unique VSGNE of the strongly monotone game. Future work will focus on finding a better way to regularize stochastic effects, potentially testing some common methods for SGD, like SGD with momentum.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 62003037, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement OCAL, No. 787845, and the Beijing Institute of Technology Research Fund Program for Young Scholars. The authors would like to express their appreciation to Mr. Yuwen Chen, for helpful discussions on this work.

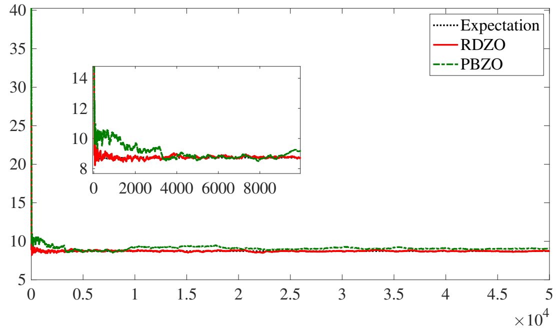


Fig. 5. High noise level, averaging 20 times.

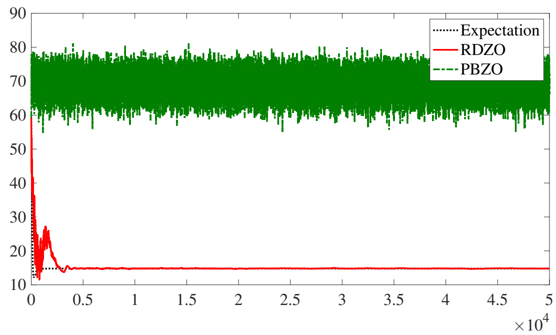


Fig. 6. High noise level, averaging 20 times, N = 30, d = 4.

REFERENCES

[1] G. Belgioioso and S. Grammatico, "Semi-decentralized nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions," *IEEE control systems letters*, vol. 1, no. 2, pp. 400–405, 2017.

[2] O. Stein and N. Sudermann-Merx, "The noncooperative transportation problem and linear generalized Nash games," *European Journal of Operational Research*, vol. 266, no. 2, pp. 543–553, 2018.

[3] D. Ardagna, M. Ciavotta, and M. Passacantando, "Generalized nash equilibria for the service provisioning problem in multi-cloud systems," *IEEE Transactions on Services Computing*, vol. 10, no. 3, pp. 381–395, 2017.

[4] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *4OR*, vol. 5, no. 3, pp. 173–210, 2007.

[5] ——, "Penalty methods for the solution of generalized Nash equilibrium problems," *SIAM Journal on Optimization*, vol. 20, no. 5, pp. 2228–2253, 2010.

[6] C. Kanzow and D. Steck, "Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems," *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2034–2058, 2016.

[7] K. Zhang, W. Shi, H. Zhu, and T. Basar, "Distributed equilibrium-learning for power network voltage control with a locally connected communication network," in *American Control Conference*, Wisconsin Center, Milwaukee, USA, 27-29 June 2018, pp. 3092–3097.

[8] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized nash equilibria computation," *Automatica*, vol. 102, no. 4, pp. 111–121, 2019.

[9] D. Gadjov and L. Pavel, "Single-timescale distributed gne seeking for aggregative games over networks via forward-backward operator splitting," *arXiv e-prints*, p. arXiv: 1908.00107, 2019.

[10] J. Lei and U. Shanbhag, "Linearly convergent variable sample-size schemes for stochastic nash games: Best-response schemes and distributed gradient-response schemes," in *IEEE Conference on Decision and Control (CDC)*, 2018, pp. 3547–3552.

[11] C. Yu, M. van der Schaar, and A. Sayed, "Distributed learning for stochastic generalized nash equilibrium problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3893–3908, 2017.

[12] R. Cominetti, E. Melo, and S. Sorin, "A payoff-based learning procedure and its application to traffic games," *Games and Economic Behavior*, vol. 70, no. 1, pp. 71–83, 2010.

[13] R. S. Sutton, *Reinforcement Learning: An Introduction, 2nd edition*, ser. Adaptive computation and machine learning. The MIT Press, Cambridge, 2018.

[14] X. Wang and T. Sandholm, "Reinforcement learning to play an optimal nash equilibrium in team markov games," in *Proceedings of the 15th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, 2002, pp. 1603–1610.

[15] M. Thathachar and P. Sastry, *Networks of learning automata: techniques for online stochastic optimization*. Kluwer Academic, Boston, 2004.

[16] S. Perkins, P. Mertikopoulos, and D. Leslie, "Mixed-strategy learning with continuous action sets," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 379–384, 2017.

[17] H. J. Kushner and G. G. Yin, *Stochastic approximation and recursive algorithms and applications*, ser. Applications of mathematics. Springer, New York, 2003, vol. 35, Ed. 2.

[18] A. Nagurney and D. Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*, ser. International Series in Operations Research and Management Science. Springer US, Boston, MA, 1996, vol. 2.

[19] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Foundations of Computational Mathematics*, vol. 17, no. 2, pp. 527–566, 2017.

[20] J. Duchi, M. Jordan, M. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2788–2806, 2015.

[21] S. Liu, J. Chen, P.-Y. Chen, and A. O. Hero, "Zeroth-order online alternating direction method of multipliers: Convergence analysis and applications," *arXiv e-prints*, p. arXiv:1710.07804v2, 2018.

[22] A. K. Sahu, D. Jakovetic, D. Bajovic, and S. Kar, "Zeroth-order stochastic variance reduction for nonconvex optimization," *arXiv e-prints*, p. arXiv:1803.07844, 2018.

[23] Z. Yu, D. W. Ho, and D. Yuan, "Distributed randomized gradient-free mirror descent algorithm for constrained optimization," *arXiv e-prints*, p. arXiv:1903.04157, 2019.

[24] T. Tatarenko and M. Kamgarpour, "Learning generalized nash equilibria in a class of convex games," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1426–1439, 2019.

[25] ——, "Learning nash equilibria in monotone games," *arXiv e-prints*, p. arXiv:1904.01882, 2019.

[26] F. Facchinei and J. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, ser. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2003.

[27] J. Koshal, A. Nedic, and U. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 594–609, 2013.

[28] ——, "Single timescale regularized stochastic approximation schemes for monotone nash games under uncertainty," in *IEEE 49th Annual Conference on Decision and Control*, 15-17 December 2010, pp. 231–236.

APPENDIX

A. Proof of Lemma III.1

For each $i \in \mathcal{N}, j \in \{1, \cdots, d\}$, we have

$$\begin{aligned}\tilde{M}^0_{i,j}(\boldsymbol{\eta}) = &-\int_{\mathbb{R}^{d-1}} p(x_{i,-j}|\boldsymbol{u}_{i,-j}, \sigma^2\mathbf{I}_{d-1})\frac{1}{\sqrt{2\pi}\sigma}d(x_{i,-j})\\ &\times\left(f^0_i(\boldsymbol{x}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda})e^{-\frac{(x_{i,j}-\boldsymbol{u}_{i,j})^2}{2\sigma^2}}\right)\Big|^{+\infty(x_{i,j})}_{-\infty(x_{i,j})}\\ &+\int_{\mathbb{R}^d}\frac{\partial f^0_i(\boldsymbol{x}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda})}{\partial x_{i,j}}p(\boldsymbol{x}_i|\boldsymbol{u}_i, \sigma^2\mathbf{I}_d)d\boldsymbol{x}_i\\ = &\int_{\mathbb{R}^d} M^0_{i,j}(\boldsymbol{x}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda})p(\boldsymbol{x}_i|\boldsymbol{u}_i, \sigma^2\mathbf{I}_d)d\boldsymbol{x}_i \qquad (26)\end{aligned}$$

The second last equality holds under Assumption 3,

$$\lim_{x_{i,j}\to\infty(-\infty)} f^0_i(\boldsymbol{x}_i, \boldsymbol{u}_{-i}, \boldsymbol{\lambda})e^{-\frac{(x_{i,j}-\boldsymbol{u}_{i,j})^2}{2\sigma^2}} = 0 \qquad (27)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3151225, IEEE Transactions on Automatic Control

Also, the same result holds for $\tilde{M}_{i,j}(\boldsymbol{u})$:

$$\tilde{M}_{i,j}(\boldsymbol{u}) = \int_{\mathbb{R}^d} M_{i,j}(\boldsymbol{x}_i, \boldsymbol{u}_{-i}) p(\boldsymbol{x}_i|\boldsymbol{u}_i, \sigma^2 \mathbf{I}_d) d\boldsymbol{x}_i \tag{28}$$

B. Proof of Theorem III.1

Under Assumption 2, $\mathcal{X}^0 \triangleq \mathcal{X} \times \mathbb{R}^m_{\geq 0}$ is closed and convex. The idea is to prove the expectation of $\|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|$ converges almost surely to 0 and subsequently obtain the expectation of $\|\boldsymbol{\eta}_{k+1} - \boldsymbol{\eta}^*\|$ converges to 0 given $\lim_{k\to\infty} \boldsymbol{y}_k = \boldsymbol{\eta}^*$.

By $\boldsymbol{y}_k \in \text{SOL}(\mathcal{Q}, M^0_{reg,k})$, we have

$$\boldsymbol{y}_{i,k} = \Pi_{Q_i} \|\boldsymbol{y}_{i,k} - h_k M^0_{reg,k,i}(\boldsymbol{y}_{i,k})\|, \quad \forall i \in [N+1]. \tag{29}$$

For any $i \in [N]$, the following is derived based on the contractive property of the projection and $M^0_{reg,k,i} = M^0_i$:

$$\begin{aligned} &\|\boldsymbol{\eta}_{i,k+1} - \boldsymbol{y}_{i,k}\|^2 \\ \leq\ &\|\boldsymbol{\eta}_{i,k} - \boldsymbol{y}_{i,k}\|^2 + h_k^2 \left\|g^0_i(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k}) - M^0_i(\boldsymbol{y}_k)\right\|^2 \\ &- 2h_k \left\langle \boldsymbol{\eta}_{i,k} - \boldsymbol{y}_{i,k}, g^0_i(\boldsymbol{\eta}_k, \boldsymbol{x}_{i,k}) - M^0_i(\boldsymbol{\eta}_k)\right\rangle \\ &- 2h_k \left\langle \boldsymbol{\eta}_{i,k} - \boldsymbol{y}_{i,k}, M^0_i(\boldsymbol{\eta}_k) - M^0_i(\boldsymbol{y}_k)\right\rangle \end{aligned} \tag{30}$$

Taking the telescopic sum of (30) over N players and then the expectation over $\Xi_k = \{\boldsymbol{v}_{i,k}, \xi_{i,k,1}, \cdots, \xi_{i,k,T_1+T_2}\}$, the following is derived

$$\begin{aligned} &\mathbb{E}_{\Xi_k} \sum_{i=1}^N \|\boldsymbol{\eta}_{i,k+1} - \boldsymbol{y}_{i,k}\|^2 \\ \leq &\sum_{i=1}^N \|\boldsymbol{\eta}_{i,k} - \boldsymbol{y}_{i,k}\|^2 + 2h_k \langle \boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k}, -A(\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k})\rangle \\ &- 2h_k\tau\|\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k}\|^2 + 2h_k\sigma_k L_1 N^{\frac{1}{2}} d^{\frac{1}{2}} \|\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k}\| \\ &+ 5\frac{h_k^2}{\sigma_k^2} \sum_{i=1}^N \left(\frac{D_1}{T_1} + \frac{D_1}{T_2}\right) d + 10h_k^2 L_1^2 \|\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k}\|^2 \\ &+ 5h_k^2 \sum_{i=1}^N \Bigg[2L_0^2(d+4)^2 + \frac{\sigma_k^2}{4} L_1^2 (d+3)^3 + \|\boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k}\|^2 \\ &\times 2\|A_i\|^2 + 4\|A_i\|^2 (d+4)^2 \left(\|\boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k}\|^2 + \|\boldsymbol{y}_{\boldsymbol{\lambda},k}\|^2\right)\Bigg], \end{aligned}$$

where $\|A(\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k})\| \leq D_2$ since both $\boldsymbol{u}_k$ and $\boldsymbol{y}_{-\boldsymbol{\lambda},k}$ are bounded. Discussing the update for the dual player gives,

$$\begin{aligned} &\|\boldsymbol{\lambda}_{k+1} - \boldsymbol{y}_{\boldsymbol{\lambda},k}\|^2 \\ \leq &- 2h_k \langle \boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k}, -A(\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k})\rangle + h_k^2 r_k \|\boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k}\|^2 \\ &+ \|(1 - h_k r_k)(\boldsymbol{\lambda}_k - \boldsymbol{y}_{\boldsymbol{\lambda},k})\|^2 + (h_k^2 r_k + h_k^2) D_2^2 \end{aligned}$$

Hence, we can derive that

$$\begin{aligned} &\mathbb{E}_{\Xi_k} \|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|^2 \\ \leq &(1 - 2h_k t_k + h_k^2 r_k^2 + h_k^2 r_k + c_1 h_k^2) \|\boldsymbol{\eta}_k - \boldsymbol{y}_k\|^2 \\ &+ (h_k^2 r_k + h_k^2) D_2^2 + 2h_k\sigma_k L_1 N^{\frac{1}{2}} d^{\frac{1}{2}} \|\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k}\| \\ &+ 5\frac{h_k^2}{\sigma_k^2} \sum_{i=1}^N \left(\frac{D_1}{T_1} + \frac{D_1}{T_2}\right) d + c_2 h_k^2 \end{aligned} \tag{31}$$

where $t_k = \min\{\tau, r_k\}, \forall k \geq 0$ and c_1, c_2 are the coefficients of $h_k^2 \|\boldsymbol{\eta}_k - \boldsymbol{y}_k\|^2$ and h_k^2 respectively.

By discussing the term $\|\boldsymbol{\eta}_k - \boldsymbol{y}_k\|^2$, it gives

$$\begin{aligned} &\mathbb{E}_{\Xi_k} \|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|^2 \\ \leq &\big[1 - h_k t_k - h_k^2 r_k^2 + h_k^3 r_k^3 + h_k^2 (r_k^2 - t_k^2 + r_k + c_1) \\ &(1 + h_k r_k)\big] \|\boldsymbol{\eta}_k - \boldsymbol{y}_{k-1}\|^2 \\ &+ \left((1 - h_k t_k)^2 + 4h_k^2 r_k + c_1 h_k^2\right) \left(1 + \frac{1}{h_k r_k}\right) \\ &\left(\frac{|r_{k-1} - r_k|}{\kappa r_k} M_{\boldsymbol{y},\boldsymbol{\lambda}}\right)^2 \\ &+ (h_k^2 r_k + h_k^2) D_2^2 + 2h_k\sigma_k L_1 N^{\frac{1}{2}} d^{\frac{1}{2}} \|\boldsymbol{u}_k - \boldsymbol{y}_{-\boldsymbol{\lambda},k}\| \\ &+ 5\frac{h_k^2}{\sigma_k^2} \sum_{i=1}^N \left(\frac{D_1}{T_1} + \frac{D_1}{T_2}\right) d + c_2 h_k^2 \end{aligned} \tag{32}$$

Suppose $h_k = \frac{1}{k^a}, \sigma_k = \frac{1}{k^b}, r_k = \frac{1}{k^c}$ where $0 < a, b, c < 1$, we have the following inequalities:

$$\frac{|r_{k-1} - r_k|}{r_k} = (1 + \frac{1}{k-1})^c - 1 \leq \frac{c}{k-1} \tag{33}$$

$$h_k t_k = h_k \min\{\tau, r_k\} = h_k r_k \min\{\frac{\tau}{r_k}, 1\} \geq \kappa h_k r_k \tag{34}$$

If $\sum_{k=0}^\infty h_k r_k = \infty$, $\sum_{k=0}^\infty h_k^2 < \infty$, $\sum_{k=0}^\infty h_k \sigma_k < \infty$, $\sum_{k=0}^\infty \frac{h_k^2}{\sigma_k^2} < \infty$, it is easy to verify that $\|\boldsymbol{\eta}_k - \boldsymbol{y}_{k-1}\|^2$ converges almost surely to 0, by [28][Lemma 1]. Also, by the uniqueness of the solution, we know $\lim_{k\to\infty} \boldsymbol{y}_k = \boldsymbol{\eta}^* \in \text{SOL}(\mathcal{X} \times \mathbb{R}^m_{\geq 0}, M^0)$, which is the solution for the extended game. Therefore, we obtain that $\boldsymbol{\eta}_k$ converges to $\boldsymbol{\eta}^*$ almost surely, which finishes the proof. A valid selection of parameters are $h_k = \frac{1}{k^{0.85}}, \sigma_k = \frac{1}{k^{0.3}}, r_k = \frac{0.4}{k^{0.15}}$.

C. Proof of Corollary III.1

We could analyze the convergence rate of Algorithm 1 by finding out those parameters relating to the convergence of $\|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|$. Rewrite (32),

$$\mathbb{E}_{\Xi_k} \|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|^2 \leq (1 - \alpha_k) \|\boldsymbol{\eta}_k - \boldsymbol{y}_{k-1}\|^2 + \beta_k \tag{35}$$

Hence, the convergence rate of the expectation of $\|\boldsymbol{\eta}_{k+1} - \boldsymbol{y}_k\|$ is affected by the values of α_k and β_k, which is Corollary III.1.