Open access • Journal Article • DOI:10.1109/TAC.2018.2841319

Learning Generalized Nash Equilibria in a Class of Convex Games — Source link

Tatiana Tatarenko, Maryam Kamgarpour

Institutions: ETH Zurich

Published on: 01 Apr 2019 - IEEE Transactions on Automatic Control (IEEE)

Topics: Best response, Nash equilibrium, Convex function, Stochastic game and Coordination game

Learning Generalized Nash Equilibria in a Class of Convex Games

Journal Article

Author(s):
Tatarenko, Tatiana; Kamgarpour, Maryam

Publication date:
2019

Permanent link:
https://doi.org/10.3929/ethz-b-000268429

Originally published in:
IEEE Transactions on Automatic Control, https://doi.org/10.1109/TAC.2018.2841319

Learning Generalized Nash Equilibria in a Class of Convex Games

Tatiana Tatarenko[*] Maryam Kamgarpour[*†]

October 4, 2018

Abstract

We consider multi-agent decision making where each agent optimizes its convex cost function subject to individual and coupling constraints. The constraint sets are compact convex subsets of a Euclidean space. To learn Nash equilibria, we propose a novel distributed payoff-based algorithm, where each agent uses information only about its cost value and the constraint value with its associated dual multiplier. We prove convergence of this algorithm to a Nash equilibrium, under the assumption that the game admits a strictly convex potential function. In the absence of coupling constraints, we prove convergence to Nash equilibria under significantly weaker assumptions, not requiring a potential function. Namely, strict monotonicity of the game mapping is sufficient for convergence. We also derive the convergence rate of the algorithm for strongly monotone game maps.

1 Introduction

Decision making in multi-agent systems arises in engineering applications ranging from electricity markets to telecommunication and transportation networks [2, 37, 40]. Game theory provides a powerful framework for analyzing and optimizing decisions in multi-agent systems. The notion of an equilibrium in a game characterizes stable solutions to multi-agent decision making problems. In this work, we design a distributed learning algorithm to converge to Nash equilibria for a class of non-cooperative games modeled by convex objective functions and coupling constraints.

There is a large body of work on computation of Nash equilibria. The approaches differ mainly by the particular structure of agents' cost functions as well as the information available to each agent. In a so-called *potential game*, a central optimization problem can be formulated whose minimizers coincide with a subset of the Nash equilibria of the game. One can then leverage distributed optimization algorithms to compute the minima of the potential function [21, 38], despite agents' limited information of others' cost functions or action sets. Distributed algorithms have also been designed for the class of *aggregative games* [14, 31]. In general, for implementation of the aforementioned distributed algorithms each agent needs to know the structure of its cost function or its derivative. Furthermore, agents may need to communicate with each other or with a central entity, even if their strategy spaces are decoupled.

In contrast to deterministic distributed optimization approaches, learning approaches start with the assumption that the each agent's objective function or its derivative may not be known to the agent itself nor to the other agents. They attempt to compute Nash equilibria by sampling agents' actions from a set of probability distributions. These probability distributions are updated based on the information available in the system. In particular, most of the past work has focused on algorithms that require the ability of each agent to evaluate its cost function at any feasible point, given fixed actions of all other agents, and convergence is established for the subclass of potential games [23, 33, 45, 47].

In many practical situations the agents do not know functional form of their objectives and can only access the values of their objective functions at a played action. Such situations arise, for example, in electricity markets (unknown price functions or constraints) [24, 50], network routing (unknown traffic demands/constraints) [5, 26], and sensor coverage problems (unknown density function on the mission space)

This research was gratefully funded by the European Union ERC Starting Grant CONENE.

[*] Department of Control Theory and Robotics, TU Darmstadt, Germany

[†] Automatic Control Laboratory, ETH Zürich, Switzerland

[55]. In such cases, the information structure is referred to as *payoff-based*, that is, each agent can only observe its obtained payoffs and be aware of its local actions. A payoff-based learning algorithm in potential games is proposed in [25] with the guarantee of stochastic stability of potential function minimizers. However, to implement this payoff-based algorithm agents need to have some memory. Other algorithms requiring only payoff-based information and memory are proposed in [10] and [55]. These learning procedures assume a potential game and guarantee convergence to a distribution over potential function minimizers in total variation. In [42] the idea of dynamic feedback is utilized for matrix games and an extension of fictitious play is proposed that considers empirical frequencies and their derivatives. The convergence to Nash equilibrium in this setting is established. Learning based approaches are also proposed in [34] for non-potential games, where stochastic convergence to the Nash equilibrium maximizing social welfare is guaranteed.

The above payoff-based procedures are applicable to games with finite action spaces. For games with uncountable action spaces, a payoff-based approach was developed based on extremum seeking [8]. The extremum seeking approach designs a dynamic update law for the actions based on sinusoidally perturbed measured payoffs. If the amplitude and frequency of this sinusoid are chosen properly, locally asymptotically stable equilibria of this dynamical system will correspond to Nash equilibria of the game. This approach was extended to account for stochastic noise affecting measurements of the cost functions [44]. Given strongly convex cost functions almost sure convergence to a Nash equilibrium was proven. An alternative payoff-based approach, inspired by the *logit* dynamics in finite action games [4], was proposed in [46] in a potential game setting. This approach considered sampling agents' actions from a Gaussian distribution. The result was generalized to arbitrary games (not necessarily potential) with uncoupled action sets in [48].

Despite considerable progress in learning algorithms for games, the work on payoff-based learning has not considered convex cost functions and coupling constraints on agents' actions. In several realistic scenarios in which players share resources each player's feasible strategy space depends on the other players' actions. For example, in an electricity market, there are coupling constraints due to the underlying physical electricity network. Similar constraints exist in a transportation or telecommunication network and general deregulated economy problems [41]. The Nash equilibria in a game with coupling constraints are referred to as *generalized Nash equilibria*. Ensuring uniqueness of these equilibria and computing them is a lively research topic [7].

We focus on learning equilibria in a subset of generalized Nash equilibrium problems in which the coupling constraint is shared among the agents and is jointly convex (convex in all actions). In this setting, one can formulate a variational equilibrium problem to characterize a subset of the generalized Nash equilibria, referred to as variational equilibria [7]. In addition to computational advantages, variational equilibria present few desired practical properties. For example, since in this equilibrium, the dual multiplier associated to the joint constraint is equal across all players, there is a well-defined cost associated to constraint violation. Note that variational equilibria are also a subset of the normalized equilibria, introduced in the seminal work [36], where the normalizing coefficients considered in [36] would be constant across all players. The authors in [20] further derived theoretical connections between variational equilibria and generalized Nash equilibria and showed that in the interior of the shared constraint sets, the two equilibria concepts are equivalent.

Recent research has focused on distributed algorithms for computing generalized Nash equilibria. Authors in [53] consider variational equilibria in monotone games and propose a primal-dual distributed algorithm. Similarly, [30, 9] addresses decentralized computation of variational equilibria for aggregative games. In [39] a distributed primal-dual algorithm for computing generalized Nash equilibria is proposed for a network game. In the network game setting, it is assumed that there exists a communication graph through which each player can share its strategy information with its neighbors. Hence, some coordination between agents is needed. In all the above work, each agent needs to know the functional form of their cost function or its gradient. The work in [54] suppresses this requirement and develops a primal-dual algorithm for learning generalized Nash equilibria. Nevertheless, players need to exchange information with their neighbors according to the network graph. As such, they can estimate the gradients of their cost functions online using neighborhood information.

The approach to estimate the gradient of a cost function online is well-studied in the stochastic optimization literature [28]. In the game setting however, this approach necessitates some coordination between agents. This is because for a given player to estimate the gradient of its cost function, it needs to evaluate this cost function at at least two points of its strategy space while other agents who influence the player's cost function should not change their actions (otherwise, the player cannot attribute the decrease/increase of its cost to its own actions and hence, the gradient cannot be estimated). Our goal is to develop a payoff-based algorithm that bypasses the need for coordination or information exchange during each step of the algorithm. Naturally however, similar to all past algorithms, the agents must agree to implement the algorithm.

Our contributions are as follows. First, we develop a payoff-based approach for computing Nash equilibria in a class of convex games with jointly convex coupling constraints. Second, we prove almost sure convergence of the algorithm to variational Nash equilibria, under the existence of a strictly convex potential function. Third, in the absence of coupling constraints, we prove almost sure convergence to variational Nash equilibria, relaxing the requirement of existence of a potential function. Fourth, for this latter setup, we quantify the convergence rate of the payoff-based algorithm if the game map is strongly monotone. While our setup is similar to [36, 53, 54, 30], in contrast to the above work we do not require knowledge of the cost functions, constraints or their gradients [30, 53]. Also, we require neither information exchange between players [39, 54], nor knowledge of a norm bound on the dual multipliers of the coupling constraints [54].

Our approach is detailed as follows. We extend the game to define a player corresponding to the dual multiplier of the coupling constraints, similar to [53, 54, 30, 9]. We then develop a novel sampling based approach, in which the probability distributions from which agents sample their actions are Gaussian, inspired by the literature on learning automata [51]. The mean of the distribution is updated iteratively by each agent based only on its own current payoff and local constraint set. The dual player, on the other hand, updates its action deterministically by measuring constraint violation at each time step. Notice that similar to [30, 54] the dual player is a fictitious player. It can refer to a central coordinator who measures the constraint violation at each step. Alternatively, if each agent can locally measure the constraint violation, it can update its dual variable. Furthermore, similar to primal-dual algorithms in [30, 53, 54] constraints are satisfied upon convergence of the algorithm. To prove convergence of our algorithm we leverage results on Robbins-Monro stochastic approximation [3, 29]. We quantify the convergence rate based on rate estimates in stochastic projection algorithms [43].

This paper is organized as follows. In Section 2, we set up the game under consideration. In Section 3, we propose our payoff-based approach and present its convergence result. Section 4 develops the proof of the main result using supporting theorems on stochastic random variables. In Section 5, we relax the coupling constraint and consequently, the requirements for convergence of the proposed algorithm. Furthermore, we provide a convergence rate for this latter case. A case study is provided in Section 6 based on a game arising in a classical Cournot economic model. In Section 7, we summarize the result and discuss future work.

Notations and basic definitions. The set $\{1,\dots,N\}$ is denoted by $[N]$. Boldface is used to distinguish between vectors in a multi-dimensional space and scalars. Given N vectors $\boldsymbol{x}^i \in \mathbb{R}^d$, $i \in [N]$, $[\boldsymbol{x}^i]_{i=1}^N := [{\boldsymbol{x}^1}^\top,\dots,{\boldsymbol{x}^N}^\top]^\top \in \mathbb{R}^{Nd}$; $\boldsymbol{x}^{-i} := [\boldsymbol{x}^1,\dots,\boldsymbol{x}^{i-1},\boldsymbol{x}^{i+1},\dots,\boldsymbol{x}^N] \in \mathbb{R}^{(N-1)d}$. $\mathbb{R}^d_+$ and $\mathbb{Z}_+$ denote respectively, vectors from $\mathbb{R}^d$ with non-negative coordinates and non-negative whole numbers. The standard inner product on $\mathbb{R}^d$ is denoted by $(\cdot,\cdot)$: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, with associated norm $\|\boldsymbol{x}\| := \sqrt{(\boldsymbol{x},\boldsymbol{x})}$. We let $\mathbb{R}^d_{\le K} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \le K\}$. I_d represents the d-dimensional identity matrix and $\mathbb{1}_N$ represents the N-dimensional vector of unit entries. Given some matrix $A \in \mathbb{R}^{d\times d}$, $A \succeq (\succ) 0$, if and only if $\boldsymbol{x}^\top A \boldsymbol{x} \ge (>) 0$ for all $\boldsymbol{x} \ne 0$.

Given a function $\mathbf{g}(\boldsymbol{x},\boldsymbol{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$, we define the mapping $\nabla_{\boldsymbol{x}}\mathbf{g}(\boldsymbol{x},\boldsymbol{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ component-wise as $[\nabla_{\boldsymbol{x}}\mathbf{g}(\boldsymbol{x},\boldsymbol{y})]_i := \frac{\partial \mathbf{g}(\boldsymbol{x},\boldsymbol{y})}{\partial x^i}$. We use the big-$O$ notation, that is, the function $f(x) : \mathbb{R} \to \mathbb{R}$ is $O(\mathbf{g}(x))$ as $x \to a$, $f(x) = O(g(x))$ as $x \to a$, if $\lim_{x\to a} \frac{|f(x)|}{|g(x)|} \le K$ for some positive constant K. We say that a function $f(\boldsymbol{x})$ grows not faster than a function $g(\boldsymbol{x})$ as $\boldsymbol{x} \to \infty$, if there exists a positive constant Q such that $f(\boldsymbol{x}) \le g(\boldsymbol{x})$ for any $\boldsymbol{x}$ with $\|\boldsymbol{x}\| \ge Q$.

Definition 1. *The mapping $\boldsymbol{M} : \mathbb{R}^d \to \mathbb{R}^d$ is called* pseudo-monotone *over $X \subseteq \mathbb{R}^d$, if $(\boldsymbol{M}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}) \ge 0$ implies $(\boldsymbol{M}(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{y}) \ge 0$ for every $\boldsymbol{x},\boldsymbol{y} \in X$.*

Definition 2. *The mapping $\boldsymbol{M} : \mathbb{R}^d \to \mathbb{R}^d$ is called* strongly monotone *over $X \subseteq \mathbb{R}^d$ with constant $\kappa > 0$, if $(\boldsymbol{M}(\boldsymbol{x}) - \boldsymbol{M}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}) \ge \kappa\|\boldsymbol{x}-\boldsymbol{y}\|^2$ for any $\boldsymbol{x},\boldsymbol{y} \in X$. It is* strictly monotone *if $(\boldsymbol{M}(\boldsymbol{x}) - \boldsymbol{M}(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}) > 0$ for any $\boldsymbol{x},\boldsymbol{y} \in X$.*

Definition 3. *Let $\mathcal{C}$ be a convex constraint set described by a finite set of convex inequality constraints $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^d : f^i(\boldsymbol{x}) \le 0, i \in [m]\}$. The* Slater's constraint qualification *consists in existence of a strictly feasible point $\mathbf{x}^* \in \mathcal{C}$, $f^i(\boldsymbol{x}^*) < 0$ for $i \in [m]$.*

2 Problem Formulation

2.1 Convex games with coupling constraints

We consider a game $\Gamma(N, \{A_i\}, \{J_i\}, C)$ with N players. We assume that the action of the ith player is locally constrained to $\boldsymbol{a}^i \in A_i \subset \mathbb{R}^d$ and that the vector of joint actions[1], $\boldsymbol{a} = [\boldsymbol{a}^1 \ldots, \boldsymbol{a}^N] \in \boldsymbol{A} = A_1 \times \ldots \times A_N$, has to belong to a *global coupling constraint set* C, namely

$$\boldsymbol{a} \in C = \{\boldsymbol{a} \in \boldsymbol{A} : \mathbf{g}(\boldsymbol{a}) \leq \mathbf{0}\}, \tag{1}$$

where $\mathbf{g} : \mathbb{R}^{Nd} \to \mathbb{R}^n$ with coordinates $g_i(\boldsymbol{a}), i \in [n]$. Let $\mathcal{Q} = \boldsymbol{A} \cap C$, $\mathcal{Q}^i(\boldsymbol{a}^{-i}) = \{\boldsymbol{a}^i \in A^i : \mathbf{g}(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) \leq 0\}$. The cost functions $J_i : \mathbb{R}^{Nd} \to \mathbb{R}$ indicate the cost $J_i(\boldsymbol{a})$ the agent i has to pay, given any joint action $\boldsymbol{a} \in \mathcal{Q}$. Throughout this paper we assume A_i to be compact for all $i \in [N]$.

A *generalized Nash equilibrium* (GNE) in a game Γ with coupled actions represents a joint action from which no player has any incentive to unilaterally deviate.

Definition 4. *A point $\boldsymbol{a}^* \in \mathcal{Q}$ is called a* generalized Nash equilibrium *(GNE) if for any $i \in [N]$ and $\boldsymbol{a}^i \in \mathcal{Q}^i(\boldsymbol{a}^{*-i})$*

$$J_i(\boldsymbol{a}^{*i}, \boldsymbol{a}^{*-i}) \leq J_i(\boldsymbol{a}^i, \boldsymbol{a}^{*-i}).$$

If $C = \mathbb{R}^{Nd}$ then $\mathcal{Q}^i(\boldsymbol{a}^{-i}) = \{a^i : a^i \in A_i\}$ and any $\boldsymbol{a}^*$ for which the inequality above holds is a Nash equilibrium (NE)*.*

We consider convex games as follows.

Assumption 1. *The game under consideration is* convex*. Namely, for all $i \in [N]$ the set A_i is convex and compact, the cost function $J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$ is defined on $\mathbb{R}^{Nd}$, continuously differentiable in $\boldsymbol{a}$ and convex in $\boldsymbol{a}^i$ for fixed $\boldsymbol{a}^{-i}$. The coupling constraint function $\mathbf{g} : \mathbb{R}^{Nd} \to \mathbb{R}^n$ is continuously differentiable and has convex coordinates $g_i(\boldsymbol{a}), i \in [n]$.*

Given differentiable cost functions, we define the game mapping and the extended game mapping.

Definition 5. *The mapping $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$, referred to as the* game mapping *of $\Gamma(N, \{A_i\}, \{J_i\}, C)$ is defined by*

$$\begin{aligned}
&\boldsymbol{M}(\boldsymbol{a}) = \\
&[M_{1,1}(\boldsymbol{a}), \ldots, M_{1,d}(\boldsymbol{a}), \ldots, M_{N,1}(\boldsymbol{a}), \ldots, M_{N,d}(\boldsymbol{a})]^\top, \\
&M_{i,k}(\boldsymbol{a}) = \frac{\partial J_i(\boldsymbol{a})}{\partial a^i_k}, \ \boldsymbol{a} \in \mathcal{Q} = \boldsymbol{A} \cap C, i \in [N], k \in [d].
\end{aligned} \tag{2}$$

Definition 6. *The mapping $\boldsymbol{M}^0 : \mathbb{R}^{Nd+n} \to \mathbb{R}^{Nd+n}$, referred to as the* extended game mapping *of $\Gamma(N, \{A_i\}, \{J_i\}, C)$ with coupled actions, is defined by*

$$\begin{aligned}
&\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda}) = [\boldsymbol{M}^0_1(\boldsymbol{a}, \boldsymbol{\lambda}), \ldots, \boldsymbol{M}^0_N(\boldsymbol{a}, \boldsymbol{\lambda}), -\mathbf{g}(\boldsymbol{a})]^\top, \\
&\boldsymbol{M}^0_i(\boldsymbol{a}, \boldsymbol{\lambda}) = [M^0_{i,1}(\boldsymbol{a}, \boldsymbol{\lambda}), \ldots, M^0_{i,d}(\boldsymbol{a}, \boldsymbol{\lambda})], \quad i \in [N], \\
&M^0_{i,k}(\boldsymbol{a}, \boldsymbol{\lambda}) = M_{i,k}(\boldsymbol{a}) + \frac{\partial(\boldsymbol{\lambda}, \mathbf{g}(\boldsymbol{a}))}{\partial a^i_k}, \ \boldsymbol{a} \in \mathcal{Q} = \boldsymbol{A} \cap C, \\
&\qquad\qquad\qquad\qquad\qquad i \in [N], \quad k \in [d].
\end{aligned} \tag{3}$$

To design an algorithm with bounded iterates, we need the following standard assumptions [30, 53, 54, 7].

Assumption 2. *The coordinates $\boldsymbol{M}^0_i(\boldsymbol{a}, \boldsymbol{\lambda}) : \mathbb{R}^{Nd+n} \to \mathbb{R}^d$ of extended mapping $\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda}) : \mathbb{R}^{Nd+n} \to \mathbb{R}^{Nd+n}$ of a game $\Gamma(N, \{A_i\}, \{J_i\}, C)$ with coupled actions are* Lipschitz *on $\mathbb{R}^{Nd}$ with respect to coordinates $\boldsymbol{a}$ with a linear function $L_i(\boldsymbol{\lambda})$. The function $\mathbf{g}(\boldsymbol{a})$ is Lipschitz on $\mathbb{R}^{Nd}$. Moreover, the extended game mapping $\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda})$ is* pseudo-monotone *on $\mathcal{Q} \times \mathbb{R}^n_+ = (\boldsymbol{A} \cap C) \times \mathbb{R}^n_+$.*

[1] All results below are applicable for games with different dimensions $\{d_i\}$ of the action sets $\{A_i\}$.

Assumption 3. *The sets A_i, $i \in [N]$, $\boldsymbol{A}$, and $\mathcal{Q}$ satisfy the Slater's constraint qualification (see Definition 3).*

Assumption 4. *The cost functions $J_i(\boldsymbol{a})$, $i \in [N]$, grow not faster than a quadratic function of $\boldsymbol{a}^i$ as $\|\boldsymbol{a}^i\| \to \infty$.*

Let us provide some insight on the assumptions above.

Remark 1. *If agents' cost functions are quadratic and coupling constraints are linear, the extended game mapping is affine, namely $\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda}) = M[\boldsymbol{a}, \boldsymbol{\lambda}]^\top + \boldsymbol{m}$, where $M \in \mathbb{R}^{(Nd+n)\times(Nd+n)}$ and $\boldsymbol{m} \in \mathbb{R}^{Nd+n}$. The affine mapping above is pseudo-monotone if M is positive semi-definite [11]. This is in particular fulfilled if the quadratic forms of the cost functions are positive definite or semi-definite (see [31] and [48], respectively). However, if the affine map is pseudo-monotone for every $\boldsymbol{m}$ then M is positive semi-definite and hence the map is also monotone [11]. In general, monotonicity implies pseudo-monotonicity and the former is more stringent[2].*

Remark 2. *Since the extended mapping $\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda})$ is affine in $\boldsymbol{\lambda}$, the Lipschitz condition for $\boldsymbol{M}^0(\boldsymbol{a}, \boldsymbol{\lambda})$ in Assumption 2 above holds if the coordinates $\boldsymbol{M}_i(\boldsymbol{a})$, $i \in [N]$, of game mapping $\boldsymbol{M}(\boldsymbol{a})$ and the functions $\frac{\partial g_j(\boldsymbol{a})}{\partial a^i_k}$, $j \in [n]$, $i \in [N]$, $k \in [d]$, are Lipschitz with respect to their argument $\boldsymbol{a} = (\boldsymbol{a}^1, \ldots, \boldsymbol{a}^N)$ with some constants l_i, $l^k_{j,i}$ respectively.*

Remark 3. *Given Lipschitz continuity of $\mathbf{g}$ on $\mathbb{R}^{Nd}$, the functions $g_i(\boldsymbol{a}), i \in [n]$, grow not faster than a linear function of $\boldsymbol{a}$ as $\|\boldsymbol{a}\| \to \infty$. Furthermore, since the action sets are compact, we can always approximate J_i outside A_i by a quadratic function without loss of generality.*

2.2 Generalized Nash equilibria and Variational Inequalities

Here, we prove that the set of GNE is nonempty, given fulfillment of Assumptions 1-3 for the game $\Gamma(N, \{A_i\}, \{J_i\}, C)$. This result will be obtained through connecting generalized Nash equilibria and solutions of variational inequalities. Moreover, we model an uncoupled action game associated with the game Γ and establish the relation between its Nash equilibria and the GNE of the game Γ. Existence of such an uncoupled action game will allow us to present a payoff-based approach to learning GNE in the initial game Γ.

Definition 7. *Consider a mapping $\boldsymbol{T}(\cdot)$: $\mathbb{R}^d \to \mathbb{R}^d$ and a set $Y \subseteq \mathbb{R}^d$. The* solution set *$SOL(Y, \boldsymbol{T})$ to the* variational inequality problem *$VI(Y, \boldsymbol{T})$ is a set of vectors $\mathbf{y}^* \in Y$ such that $(\boldsymbol{T}(\mathbf{y}^*), \mathbf{y} - \mathbf{y}^*) \geq 0$, for all $\mathbf{y} \in Y$.*

Given $VI(Y, \boldsymbol{T})$, suppose that the set Y is compact, convex and that the mapping $\boldsymbol{T}$ is continuous. Then, $SOL(Y, \boldsymbol{T})$ is nonempty and compact (see Corollary 2.2.5 in [32].)

For a game $\Gamma(N, \{A_i\}, \{J_i\}, C)$ with coupled actions $\mathcal{Q} = \boldsymbol{A} \cap C$, we can define $VI(\mathcal{Q}, \boldsymbol{M})$, where $\boldsymbol{M}$ is the game mapping defined in (2). Under Assumption 1, Theorem 2.1 in [6] implies that $SOL(\mathcal{Q}, \boldsymbol{M})$ is non-empty and if $\boldsymbol{a}^* \in SOL(\mathcal{Q}, \boldsymbol{M})$, then $\boldsymbol{a}^*$ is a GNE in the game Γ.

A challenge in developing a payoff-based learning based algorithm lies in the coupling constraint C. Hence, our goal is to develop a game with uncoupled actions whose equilibria can be used to find those of the original game Γ. To do so, we first define an *associated game* $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ as follows:

$$\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+) = \Gamma_a(N+1, \{J^0_i\}_{i\in[N+1]}, \{\{A_i\}_{i\in[N]}, \mathbb{R}^n_+\}), \tag{4}$$

with $N+1$ players. The first N players are called regular and the $(N+1)$th player is called dual. The action sets of the regular players coincide with the local action sets $\{A_i\}$ of the players in the initial game Γ, whereas the action set of the dual player is the set $\mathbb{R}^n_+$. The cost functions of the players in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ are defined as follows:

$$\begin{aligned} J^0_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}, \boldsymbol{\lambda}) &= J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) + (\boldsymbol{\lambda}, \mathbf{g}(\boldsymbol{a}^i, \boldsymbol{a}^{-i})), \quad i \in [N], \\ J^0_{N+1}(\boldsymbol{a}, \boldsymbol{\lambda}) &= -(\boldsymbol{\lambda}, \mathbf{g}(\boldsymbol{a})). \end{aligned} \tag{5}$$

[2] As an example, gradient of any pseudo-convex function such as $x^3 + x$ is pseudo-monotone but not necessarily monotone.

So, the cost function of each regular player $i \in [N]$ in the game $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ is composed of two terms: the original cost function from the game Γ plus an additional term that depends on the strategy $\boldsymbol{\lambda}$ of the dual player and on the influence of the current joint action in the coupling constraint. As $\boldsymbol{\lambda} \geq \mathbf{0}$, the latter can be interpreted as a term penalizing violations of the global constraint by the given joint action.

Lemma 1. *Let $\Gamma(N, \{A_i\}, \{J_i\}, C)$ be a game for which Assumptions 1-3 hold. Then,*

1) $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*] \in \boldsymbol{A} \times \mathbb{R}^n_+$ is a Nash equilibrium in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$, if and only if $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*] \in SOL(\boldsymbol{A} \times \mathbb{R}^n_+, \boldsymbol{M}^0)$,

2) if $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*]$ is a Nash equilibrium in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$, then $\boldsymbol{a}^*$ is a GNE of Γ,

3) there exists a Nash equilibrium $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*]$ in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$,

4) for any Nash equilibrium $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*]$ in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ there exists a constant $K > 0$ such that $\|\boldsymbol{\lambda}^*\| \leq K$.

Please refer to the appendix for the proof of the above lemma.

3 Payoff-Based Algorithm

Let $\mathbf{x}^i(t) = [x^i_1(t), \dots, x^i_d(t)]^\top \in \mathbb{R}^d$ denote the strategy of player i at iteration t referred to as its *state* and $\hat{J}^0_i(t) = J^0_i(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t), \boldsymbol{\lambda}(t))$ the current value of its cost at a joint state. Each regular player $i \in [N]$, "mixes" its state, namely, it chooses its state $\mathbf{x}^i(t)$ randomly according to the multidimensional normal distribution $\mathcal{N}(\boldsymbol{\mu}^i(t) = [\mu^i_1(t), \dots, \mu^i_d(t)]^\top, \sigma_i(t))$ with density:

$$
\begin{aligned}
&p_i(x^i_1, \dots, x^i_d; \boldsymbol{\mu}^i(t), \sigma_i(t)) \\
&= \frac{1}{(\sqrt{2\pi}\sigma_i(t))^d} \exp\left\{ -\sum_{k=1}^{d} \frac{(x^i_k - \mu^i_k(t))^2}{2\sigma^2_i(t)} \right\},
\end{aligned}
$$

where $i \in [N]$. Our choice of Gaussian distribution is based on the idea of Continuous Action-set Learning Automaton presented in the literature on learning automata [51].

The mean $\boldsymbol{\mu}^i(t)$ of the state's distribution is considered an action of the regular agent i and is updated as follows:

$$
\begin{aligned}
&\boldsymbol{\mu}^i(t+1) = \\
&\mathrm{Proj}_{A_i}\left[\boldsymbol{\mu}^i(t) - \gamma_i(t+1)\sigma^2_i(t+1)\hat{J}^0_i(t)\frac{\mathbf{x}^i(t) - \boldsymbol{\mu}^i(t)}{\sigma^2_i(t)} \right],
\end{aligned}
\tag{6}
$$

where $i \in [N]$, $\gamma_i(t+1)$ is a step-size parameter chosen by player i and $\mathrm{Proj}_{A_i}[\cdot]$ denotes the projection on set A_i. The initial value of $\boldsymbol{\mu}(0)$ can be set to any finite value arbitrarily.

As for the dual player $N+1$, it updates its current action $\boldsymbol{\lambda}(t)$ based only on the observation of the violation of the constraint C, namely based on the actual value $\hat{\mathbf{g}}(t)$, of the function $\mathbf{g}(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t))$ at the current joint state of the regular players as follows:

$$
\boldsymbol{\lambda}(t+1) = \mathrm{Proj}_{\mathbb{R}^n_+}[\boldsymbol{\lambda}(t) + \beta_0(t+1)\hat{\mathbf{g}}(t)], \tag{7}
$$

where $\beta_0(t+1)$ is a step-size parameter chosen by the dual player $N+1$. The initial value of $\boldsymbol{\lambda}(0)$ can be arbitrarily set to any finite value.

Note that in contrast to the approach in computing generalized Nash equilibria presented in [54], our proposed payoff-based algorithm does not rely on the specified bound K of the dual variable $\boldsymbol{\lambda}^*$ in the associated bounded game $\Gamma_{ab}(\boldsymbol{A} \times \mathbb{R}^n_{\leq K+r})$, nor does it assume a communication graph between agents to estimate local gradients.

To analyze the convergence of this algorithm, we show that this algorithm is analogous to the Robbins-Monro stochastic approximation procedure [3].

Given $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$, for any $j \in [N+1]$ define

$$\tilde{J}_j^0(\boldsymbol{\mu}^1, \dots, \boldsymbol{\mu}^N, \boldsymbol{\lambda}, \boldsymbol{\sigma}) = \int_{\mathbb{R}^{Nd}} J_j^0(\boldsymbol{x}, \boldsymbol{\lambda}) p(\boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{\sigma}) d\boldsymbol{x},$$
$$p(\boldsymbol{\mu}, \boldsymbol{x}, \boldsymbol{\sigma}) = \prod_{j=1}^{N} p_j(x_1^j, \dots, x_d^j; \boldsymbol{\mu}^j, \sigma_j).$$

For $j \in [N+1]$, $\tilde{J}_j^0$ can be interpreted as the jth player's cost function in mixed strategies of the regular players, given that the mixed strategies of these players are multivariate normal distributions $\{\mathcal{N}(\boldsymbol{\mu}^i, \sigma_i)\}_{i\in[N]}$. It follows that the second terms inside the projections in (6) and (7) are samples of the gradient of the cost function with respect to these mixed strategies. In particular, we can verify that for all $\boldsymbol{\mu} \in \boldsymbol{A}$[3]

$$\begin{aligned}
&\mathrm{E}_{\mathbf{x}(t)}\{\hat{J}_i^0(t)\frac{x_k^i(t)-\mu_k^i(t)}{\sigma_i^2(t)}\} &(8)\\
=&\mathrm{E}\{J_i^0(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t), \boldsymbol{\lambda}(t))\frac{x_k^i(t)-\mu_k^i(t)}{\sigma_i^2(t)}|\\
&\qquad\qquad x_k^i(t) \sim \mathcal{N}(\mu_k^i(t), \sigma_i(t)), i \in [N], k \in [d]\}\\
=&\frac{\partial \tilde{J}_i^0(\boldsymbol{\mu}^1(t), \dots, \boldsymbol{\mu}^N(t), \boldsymbol{\lambda}(t), \boldsymbol{\sigma}(t))}{\partial \mu_k^i}, \ \forall i \in [N], k \in [d],\\
&\mathrm{E}_{\mathbf{x}(t)}\{\hat{\mathbf{g}}(t)\} = \mathrm{E}\{\mathbf{g}(\mathbf{x}^1(t), \dots, \mathbf{x}^N(t))| &(9)\\
&\qquad x_k^i(t) \sim \mathcal{N}(\mu_k^i(t), \sigma_i(t)), i \in [N], k \in [d]\}\\
=&-\nabla_{\boldsymbol{\lambda}} \tilde{J}_{N+1}^0(\boldsymbol{\mu}^1(t), \dots, \boldsymbol{\mu}^N(t), \boldsymbol{\lambda}(t), \boldsymbol{\sigma}(t)).
\end{aligned}$$

Using the notation $\boldsymbol{\eta}(t) = [\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)]$, we can rewrite the algorithm steps (6)-(7) in the following form:

$$\begin{aligned}
&\boldsymbol{\mu}^i(t+1) = \mathrm{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}^i(t) - \gamma_i(t+1)\sigma_i^2(t+1) &(10)\\
&\times (\boldsymbol{M}_i^0(\boldsymbol{\eta}(t)) + \boldsymbol{Q}_i(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) + \boldsymbol{R}_i(\boldsymbol{\eta}(t), \mathrm{x}(t), \boldsymbol{\sigma}(t)))],\\
&\boldsymbol{\lambda}(t+1) = \mathrm{Proj}_{\mathbb{R}_+^n}[\boldsymbol{\lambda}(t) - \beta_0(t+1) \times (-\mathbf{g}(\boldsymbol{\mu}(t)) &(11)\\
&+ \boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) + \boldsymbol{R}_{N+1}(\boldsymbol{\eta}(t), \mathrm{x}(t), \boldsymbol{\sigma}(t)))],
\end{aligned}$$

where for $i \in [N]$

$$\begin{aligned}
&\boldsymbol{Q}_i(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = \tilde{\boldsymbol{M}}_i^0(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) - \boldsymbol{M}_i^0(\boldsymbol{\eta}(t)),\\
&\boldsymbol{R}_i(\mathrm{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = \boldsymbol{F}_i(\mathrm{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) - \tilde{\boldsymbol{M}}_i^0(\boldsymbol{\eta}(t), \sigma(t)),\\
&\boldsymbol{F}_i(\mathrm{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = \hat{J}_i^0(t)\frac{\mathbf{x}^i(t) - \boldsymbol{\mu}^i(t)}{\sigma_i^2(t)},
\end{aligned}$$

and $\tilde{\boldsymbol{M}}_i^0(\cdot) = [\tilde{M}_{i,1}^0(\cdot), \dots, \tilde{M}_{i,d}^0(\cdot)]^\top$ is the d-dimensional mapping with the following elements:

$$\tilde{M}_{i,k}^0(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = \frac{\partial \tilde{J}_i^0(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t))}{\partial \mu_k^i}, \text{ for } k \in [d]. \tag{12}$$

Furthermore,

$$\begin{aligned}
&\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = \tilde{\boldsymbol{M}}_{N+1}^0(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) + \mathbf{g}(\boldsymbol{\mu}(t)),\\
&\boldsymbol{R}_{N+1}(\mathrm{x}(t), \boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)) = -\hat{\mathbf{g}}(t) - \tilde{\boldsymbol{M}}_{N+1}^0(\boldsymbol{\eta}(t), \boldsymbol{\sigma}(t)),
\end{aligned}$$

[3] Assumption 4 and compact set $\boldsymbol{A}$ justify differentiation under the integral sign of $\tilde{J}_i^0(\boldsymbol{\mu}^1(t), \dots, \boldsymbol{\mu}^N(t), \boldsymbol{\lambda}(t), \boldsymbol{\sigma}(t))$.

and $\tilde{\boldsymbol{M}}^0_{N+1}(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)) = \nabla_{\boldsymbol{\lambda}}\tilde{J}^0_{N+1}(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t))$. The algorithm (10)-(11) falls under the framework of Robbins-Monro stochastic approximations procedure [3], where

$$\boldsymbol{M}^0(\boldsymbol{\eta}(t)) = [\boldsymbol{M}^0_1(\boldsymbol{\eta}(t)),\dots,\boldsymbol{M}^0_N(\boldsymbol{\eta}(t)),-\mathbf{g}(\boldsymbol{\mu}(t))],$$

corresponds to the gradient term in stochastic approximation procedures. Furthermore,

$$\begin{aligned}\boldsymbol{Q}(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)) = [\boldsymbol{Q}_1(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)),\dots,&\boldsymbol{Q}_N(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)),\\ &\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t))],\end{aligned}$$

is a disturbance of the gradient term, and

$$\begin{aligned}\boldsymbol{R}(\mathbf{x}(t),\boldsymbol{\eta}(t),&\boldsymbol{\sigma}(t)) = [\boldsymbol{R}_1(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)),\dots,\\ &\boldsymbol{R}_N(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)),\boldsymbol{R}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t))],\end{aligned}$$

is a Martingale difference. Namely, according to (8) and (9),

$$\begin{aligned}&\boldsymbol{R}_i(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)) = \boldsymbol{F}_i(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)) && (13)\\ &- \mathrm{E}_{\mathbf{x}(t)}\{\boldsymbol{F}_i(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t))\},\;\; i\in[N],\\ &\boldsymbol{R}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t),\boldsymbol{\sigma}(t)) = -\hat{\mathbf{g}}(t) + \mathrm{E}_{\mathbf{x}(t)}\{\hat{\mathbf{g}}(t)\}. && (14)\end{aligned}$$

To ensure convergence of the iterates $\boldsymbol{\eta}(t)$, the step-sizes $\beta_0(t)$, $\sigma_i(t)$, $\gamma_i(t)$, $i\in[N]$, need to satisfy certain assumptions. Let $\beta_i(t) = \gamma_i(t)\sigma_i^2(t)$, $\beta_{\min}(t) = \min_{i\in\{0,[N]\}}\beta_i(t)$, $\beta_{\max}(t) = \max_{i\in\{0,[N]\}}\beta_i(t)$, $\gamma_{\max}(t) = \max_{i\in[N]}\gamma_i(t)$.

Assumption 5. *The variance parameters $\sigma_i(t)$ and the step-size parameters $\beta_0(t)$, $\gamma_i(t)$, $i\in[N]$, are chosen such that*

1) $\sum_{t=0}^{\infty}\beta_{min}(t) = \infty$,
2) $\sum_{t=0}^{\infty}\beta_{max}(t) - \beta_{min}(t) < \infty$,
3) $\sum_{t=0}^{\infty}\gamma^2_{max}(t) < \infty$, $\sum_{t=0}^{\infty}\gamma_{max}(t)\sigma^3_{max}(t) < \infty$.

Remark 4. *Similar to [16] the agents choose algorithm parameters independently and the only coordination is in Assertion 2) above. An example of sequences $\gamma_i(t)$, $\sigma_i(t)$, $i\in[N]$, $\beta_0(t)$ is the protocol for distributed optimization schemes [16], where each regular agent picks a positive integer R_i, $i\in[N]$, the dual player picks a positive integer N_0, and $\gamma_i(t) = \frac{1}{(t+R_i)^a}$, $\sigma_i(t) = \frac{1}{(t+R_i)^b}$, $i\in[N]$, $\beta_0(t) = \frac{1}{(t+N_0)^{a+2b}}$, with $a+2b\in(0.5,1]$, $2a>1$, and $a+3b>1$.*

For our convergence results under coupling constraints, we will consider potential games as defined below.

Assumption 6. *The game $\Gamma(N,\{A_i\},\{J_i\},C)$ admits a strictly convex potential function $f:\mathbb{R}^{Nd}\to\mathbb{R}$, with $\frac{\partial f(\boldsymbol{a})}{\partial a^i_k} = M_{i,k}(\boldsymbol{a})$. This is equivalent to the Jacobian of the game mapping $\boldsymbol{M}:\mathbb{R}^{Nd}\to\mathbb{R}^{Nd}$ being symmetric (Theorem 1.3.1 in [32].)*

Theorem 1. *Let Assumptions 1-4, 6 hold in a game $\Gamma(N,\{A_i\},\{J_i\},C)$. Let the regular players choose the states $\{\mathbf{x}^i(t)\}$ at time t according to the normal distribution $\mathcal{N}(\boldsymbol{\mu}^i(t),\sigma_i(t))$, where the mean parameters are updated as in (6). Let the action $\boldsymbol{\lambda}(t)$ of the dual player is updated according to (7). Let the variance parameters and the step-size parameters satisfy Assumption 5 Then, as $t\to\infty$, the mean vector $\boldsymbol{\mu}(t)$ converges almost surely to the generalized Nash equilibrium $\boldsymbol{\mu}^* = \boldsymbol{a}^*$ of the game Γ, given any initial vector $[\boldsymbol{\mu}(0),\boldsymbol{\lambda}(0)]$, and the joint state $\mathbf{x}(t)$ converges in probability to $\boldsymbol{a}^*$.*

Remark 5. *Analogously to optimization methods based on the gradient descent iterations, condition 1) in Assumption 5, $\sum_{t=0}^{\infty}\beta_{min}(t) = \infty$, guarantees sufficient energy for the time-step parameter $\gamma_i(t)\sigma_i^2(t)$ to let the algorithm (10)-(11) get to a neighborhood of a desired stationary point for each i, whereas condition $\sum_{t=0}^{\infty}\gamma^2_{max}(t) < \infty$ ensures the algorithm converges as time goes to infinity.*

4 Analysis of the Algorithm Convergence

Our approach in proving Theorem 1 is to first prove boundedness of the iterates $\boldsymbol{\eta}(t)$. Next, we show that the limit of the iterates $\boldsymbol{\eta}(t)$ exists and satisfies the conditions for being a variational equilibrium of the game $\Gamma(N, \{A_i\}, \{J_i\}, C)$.

4.1 Boundedness of the Algorithm Iterates

First, we demonstrate that under conditions of Theorem 1 the vector $\boldsymbol{\eta}(t)$ stays almost surely bounded for any $t \in \mathbb{Z}_+$.

Lemma 2. *Let Assumptions 1-4 hold in a game* $\Gamma(N, \{A_i\}, \{J_i\}, C)$ *with coupled actions and* $\boldsymbol{\eta}(t) = [\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)]$ *be the vector updated in the run of the payoff-based algorithm* (10)-(11)*. Let the variance parameters and the step-size parameters satisfy Assumption 5. Then,* $\Pr\{\sup_{t\geq 0}\|\boldsymbol{\eta}(t)\|< \infty\} = 1$.

Proof. In the following, for simplicity in notation, we omit the argument $\boldsymbol{\sigma}(t)$ in the terms $\boldsymbol{M}^0$, $\tilde{\boldsymbol{M}}^0$, $\boldsymbol{Q}$, and $\boldsymbol{R}$. In certain derivations, for the same reason we omit the time parameter t as well. According to the vector form of the algorithm (10)-(11),

$$\begin{aligned}\boldsymbol{\mu}^i(t+1) =& \mathrm{Proj}_{A_i}[\boldsymbol{\mu}^i(t) - \beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}(t)) \\ &+ \boldsymbol{Q}_i(\boldsymbol{\eta}(t)) + \boldsymbol{R}_i(\boldsymbol{\eta}(t), \mathbf{x}(t)))],\end{aligned} \tag{15}$$

$$\begin{aligned}\boldsymbol{\lambda}(t+1) =& \mathrm{Proj}_{\mathbb{R}^n_+}[\boldsymbol{\lambda}(t) - \beta_0(t+1) \times (-\mathbf{g}(\boldsymbol{\mu}(t)) \\ &+ \boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t)) + \boldsymbol{R}_{N+1}(\boldsymbol{\eta}(t), \mathbf{x}(t)))].\end{aligned} \tag{16}$$

Let $\boldsymbol{\eta}^* = [\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*] \in \boldsymbol{A} \times \mathbb{R}^n_+$ be a Nash equilibrium of the associated game $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$. This equilibrium exists and its norm is bounded, namely $\|\boldsymbol{\eta}^*\|< \infty$, according to Lemma 1, Assertions 1) and 3).

Let us define the function $V(\boldsymbol{\eta}) = \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|^2$. We consider the generating operator of the Markov process $\boldsymbol{\eta}(t)$

$$LV(\boldsymbol{\eta}) = E[V(\boldsymbol{\eta}(t+1)) \mid \boldsymbol{\eta}(t) = \boldsymbol{\eta}] - V(t, \boldsymbol{\eta}).$$

Our goal is to show that $LV(\boldsymbol{\eta}(t))$ satisfies sufficient decay, as per results on stability of discrete-time Markov processes ([29], Theorem 2.5.2). That is,

$$LV(\boldsymbol{\eta}(t)) \leq -\alpha(t+1)\psi(\boldsymbol{\eta}(t)) + f(t)(1 + V(\boldsymbol{\eta}(t))),$$

where the functions $\psi(.) \geq 0$, $f(.) > 0$ will be identified as per requirements of ([29], Theorem 2.5.2) (repeated in Theorem 4 in the appendix for completeness).

Taking into account the iterative procedure for the update of $\boldsymbol{\eta}(t)$ above and the non-expansion property of the projection operator on a convex set, we get

$$\begin{aligned}\|\boldsymbol{\mu}_i(t+1) - \boldsymbol{\mu}_i^*\|^2 =& \|\mathrm{Proj}_{A_i}[\boldsymbol{\mu}_i(t) - \beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}(t)) \\ &+ \boldsymbol{Q}_i(\boldsymbol{\eta}(t)) + \boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)))] - \boldsymbol{\mu}_i^*\|^2 \\ \leq& \|\boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i^* - \beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}(t)) \\ &+ \boldsymbol{Q}_i(\boldsymbol{\eta}(t)) + \boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)))\|^2 \\ =& \|\boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i^*\|^2 - 2\beta_i(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i^*) \\ &- 2\beta_i(t+1)(\boldsymbol{Q}_i(\boldsymbol{\eta}(t)) + \boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)), \boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_i^*) \\ &+ \beta_i^2(t+1)\|\boldsymbol{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2,\end{aligned} \tag{17}$$

where, for ease of notation we have defined

$$\boldsymbol{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)) = \boldsymbol{M}_i^0(\boldsymbol{\eta}(t)) + \boldsymbol{Q}_i(\boldsymbol{\eta}(t)) + \boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t)). \tag{18}$$

Similarly, we can bound the dual term

$$\begin{aligned}
&\|\boldsymbol{\lambda}(t+1)-\boldsymbol{\lambda}^*\|^2\\
&=\|\mathrm{Proj}_{\mathbb{R}^n_+}[\boldsymbol{\lambda}(t)-\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}(t))+\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t))\\
&\qquad\qquad\qquad+\boldsymbol{R}_{N+1}(\boldsymbol{\eta}(t),\mathbf{x}(t)))]-\boldsymbol{\lambda}^*\|^2\\
&\le\|\boldsymbol{\lambda}(t)-\boldsymbol{\lambda}^*-\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}(t))+\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t))\\
&\qquad\qquad\qquad+\boldsymbol{R}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t)))\|^2\\
&=\|\boldsymbol{\lambda}(t)-\boldsymbol{\lambda}^*\|^2-2\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}(t)),\boldsymbol{\lambda}(t)-\boldsymbol{\lambda}^*)\\
&-2\beta_0(t+1)(\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t))+\boldsymbol{R}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t)),\boldsymbol{\lambda}(t)-\boldsymbol{\lambda}^*)\\
&\qquad\qquad+\beta_0^2(t+1)\|\boldsymbol{G}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t))\|^2,
\end{aligned}\tag{19}$$

with

$$\begin{aligned}
&\boldsymbol{G}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t))\\
&\quad=-\mathbf{g}(\boldsymbol{\mu}(t))+\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t))+\boldsymbol{R}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t)).
\end{aligned}\tag{20}$$

Thus, taking into account the Martingale properties in (13) and (14) of the terms $\boldsymbol{R}_j$, $j\in[N+1]$, we obtain

$$\begin{aligned}
LV(\boldsymbol{\eta})&=\mathrm{E}[\|\boldsymbol{\eta}(t+1)-\boldsymbol{\eta}^*\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}]-\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|^2\\
&=\sum_{i=1}^N\left(\mathrm{E}[\|\boldsymbol{\mu}_i(t+1)-\boldsymbol{\mu}_i^*\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}]-\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^*\|^2\right)\\
&\qquad+E[\|\boldsymbol{\lambda}(t+1)-\boldsymbol{\lambda}^*\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}]-\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\|\\
&\le-2\sum_{i=1}^N\beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}),\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^*)\\
&\quad-2\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)\\
&\quad-2\sum_{i=1}^N\beta_i(t+1)(\boldsymbol{Q}_i(\boldsymbol{\eta}),\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^*)\\
&\quad-2\beta_0(t+1)(\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)\\
&\quad+\sum_{i=1}^N\beta_i^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_i(\mathbf{x}(t),\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}\}\\
&\quad+\beta_0^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_{N+1}(\mathbf{x}(t),\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}\}.
\end{aligned}\tag{21}$$

Now, we bound the first two terms in the last expression above.

$$\begin{aligned}
&-2\sum_{i=1}^N\beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}),\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^*)\\
&\qquad\qquad-2\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)\\
&=-2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}),\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\
&\qquad\qquad+2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}),\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\
&\qquad\qquad-2\sum_{i=1}^N\beta_i(t+1)(\boldsymbol{M}_i^0(\boldsymbol{\eta}),\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^*)\\
&\qquad\qquad-2\beta_0(t+1)(-\mathbf{g}(\boldsymbol{\mu}),\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)\\
&\le-2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}),\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\
&\qquad+2(\beta_{\max}(t+1)-\beta_{\min}(t+1))\|\boldsymbol{M}^0(\boldsymbol{\eta})\|\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|\\
&\le-2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}),\boldsymbol{\eta}-\boldsymbol{\eta}^*)\\
&\qquad+2(\beta_{\max}(t+1)-\beta_{\min}(t+1))k_1(1+V(\boldsymbol{\eta})),
\end{aligned}\tag{22}$$

for some constant $k_1 > 0$, where the last inequality is due to the linear behavior of the mapping $\boldsymbol{M}^0(\boldsymbol{\eta})$ at infinity (see Assumption 4). Hence, (21) and (22) imply

$$\begin{aligned} LV(\boldsymbol{\eta}) \leq & -2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \\ & + 2(\beta_{\max}(t+1) - \beta_{\min}(t+1))k_1(1 + V(\boldsymbol{\eta})) \\ & + 2\sum_{i=1}^{N} \beta_i(t+1)\|\boldsymbol{Q}_i(\boldsymbol{\eta})\|\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_i^*\| \\ & + 2\beta_0(t+1)\|\boldsymbol{Q}_{N+1}(\boldsymbol{\eta})\|\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\| \\ & + \sum_{i=1}^{N} \beta_i^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\} \\ & + \beta_0^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\}. \end{aligned} \tag{23}$$

Let us analyze the terms containing $\boldsymbol{Q}_i$ for $i \in [N+1]$ in (23). First, we will show that the mapping $\tilde{\boldsymbol{M}}_i^0(\boldsymbol{\eta}(t))$ (see (12)) evaluated at $\boldsymbol{\eta}(t)$ is equivalent to the extended game mapping (see Definition 6) in mixed strategies, that is, for $i \in [N+1]$

$$\tilde{\boldsymbol{M}}_i^0(\boldsymbol{\eta}(t)) = \int_{\mathbb{R}^{Nd}} \boldsymbol{M}_i^0(\boldsymbol{x}, \boldsymbol{\lambda}) p(\boldsymbol{\mu}(t), \boldsymbol{x}) d\boldsymbol{x}. \tag{24}$$

Indeed, using the notations

$$\begin{aligned} &\mu_{-k}^i = (\mu_1^i, \dots, \mu_{k-1}^i, \mu_{k-1}^i, \dots \mu_d^i) \in \mathbb{R}^{d-1}, \\ &x_{-k}^i = (x_1^i, \dots, x_{k-1}^i, x_{k-1}^i, \dots x_d^i) \in \mathbb{R}^{d-1}, \\ &p(\mu_{-k}^i, x_{-k}^i) = \frac{1}{(\sqrt{2\pi}\sigma_i)^{d-1}} \exp\left\{ -\sum_{j \neq k} \frac{(x_j^i - \mu_j^i)^2}{2\sigma_i^2} \right\} \\ &p(\boldsymbol{\mu}^{-i}, \boldsymbol{x}^{-i}) = \prod_{j \neq i, j=1}^{N} \frac{1}{(\sqrt{2\pi}\sigma_j)^d} \exp\left\{ -\sum_{k=1}^{d} \frac{(x_k^j - \mu_k^j)^2}{2\sigma_j^2} \right\}, \end{aligned}$$

we can show that for any $i \in [N]$, $k \in [d]$, $\tilde{M^0}_{i,k}(\boldsymbol{\eta})$

$$\begin{aligned} &\tilde{M^0}_{i,k}(\boldsymbol{\eta}) = \tilde{M^0}_{i,k}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \\ &= \frac{1}{\sigma_i^2} \int_{\mathbb{R}^{Nd}} J_i^0(\boldsymbol{x}, \boldsymbol{\lambda})(x_k^i - \mu_k^i) p(\boldsymbol{\mu}, \boldsymbol{x}) d\boldsymbol{x} \\ &= -\int_{\mathbb{R}^{Nd}} J_i^0(\boldsymbol{x}, \boldsymbol{\lambda}) p(\mu_{-k}^i, x_{-k}^i) p(\boldsymbol{\mu}^{-i}, \boldsymbol{x}^{-i}) \frac{1}{\sqrt{2\pi}\sigma_i} \\ &\qquad \times d\left(e^{-\frac{(x_k^i - \mu_k^i)^2}{2\sigma_i^2}} \right) d\boldsymbol{x}^{-i} \\ &= -\int_{\mathbb{R}^{Nd-1}} \left(J_i^0(\boldsymbol{x}, \boldsymbol{\lambda}) e^{-\frac{(x_k^i - \mu_k^i)^2}{2\sigma_i^2}} \right) \Bigg|_{-\infty(x_k^i)}^{\infty(x_k^i)} \\ &\qquad \times p(\mu_{-k}^i, x_{-k}^i) p(\boldsymbol{\mu}^{-i}, \boldsymbol{x}^{-i}) \frac{1}{\sqrt{2\pi}\sigma_i} d\boldsymbol{x}^{-i} \\ &\qquad + \int_{\mathbb{R}^{Nd}} \frac{\partial J_i^0(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial x_k^i} p(\boldsymbol{\mu}, \boldsymbol{x}) d\boldsymbol{x} \\ &= \int_{\mathbb{R}^{Nd}} \frac{\partial J_i^0(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial x_k^i} p(\boldsymbol{\mu}, \boldsymbol{x}) d\boldsymbol{x}. \end{aligned} \tag{25}$$

The above holds since according to Assumption 4,

$$\lim_{x_k^i\to\infty(-\infty)} J_i^0(\boldsymbol{x},\boldsymbol{\lambda})e^{-\frac{(x_k^i-\mu_k^i)^2}{2\sigma_i^2}} = 0,$$

for any fixed μ_k^i, $\boldsymbol{x}^{-i}$, and $\boldsymbol{\lambda}$. Thus, (24) holds for each regular player $i \in [N]$. Moreover, for the dual player

$$\begin{aligned}\tilde{\boldsymbol{M}}_{N+1}^0(\boldsymbol{\eta}) &= \int_{\mathbb{R}^{Nd}} \nabla_{\boldsymbol{\lambda}} J_{N+1}^0(\boldsymbol{x},\boldsymbol{\lambda})p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x} \\ &= -\int_{\mathbb{R}^{Nd}} \mathbf{g}(\boldsymbol{x})p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x}. \end{aligned} \tag{26}$$

Since $\boldsymbol{Q}_i(\boldsymbol{\eta}(t)) = \tilde{\boldsymbol{M}}_i^0(\boldsymbol{\eta}(t)) - \boldsymbol{M}_i^0(\boldsymbol{\eta}(t))$ and due to Assumption 2 and equation (24), we obtain the following:

$$\begin{aligned}\|\boldsymbol{Q}_i(\boldsymbol{\eta})\| &= \|\int_{\mathbb{R}^{Nd}}[\boldsymbol{M}_i^0(\boldsymbol{x},\boldsymbol{\lambda}) - \boldsymbol{M}_i^0(\boldsymbol{\mu},\boldsymbol{\lambda})]p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x}\| \\ &\le \int_{\mathbb{R}^{Nd}} \|\boldsymbol{M}_i^0(\boldsymbol{x},\boldsymbol{\lambda}) - \boldsymbol{M}_i^0(\boldsymbol{\mu},\boldsymbol{\lambda})\|p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x} \\ &\le \int_{\mathbb{R}^{Nd}} L_i(\boldsymbol{\lambda})\|\boldsymbol{x}-\boldsymbol{\mu}\|p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x} \\ &\le \int_{\mathbb{R}^{Nd}} L_i(\boldsymbol{\lambda})\left(\sum_{i=1}^N\sum_{k=1}^d |x_k^i-\mu_k^i|\right)p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x} \\ &= O(\sum_{i=1}^N \sigma_i)(1+\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|), \end{aligned} \tag{27}$$

where the last equality is due to the fact that the first central absolute moment of a random variable with a normal distribution $\mathcal{N}(\mu,\sigma)$ is $O(\sigma)$ and $L_i(\boldsymbol{\lambda})$ is a linear function of $\boldsymbol{\lambda}$ (see Assumption 2) and, hence, $L_i(\boldsymbol{\lambda}) \le k(1+\|\boldsymbol{\eta}-\boldsymbol{\eta}^*\|)$ for some constant k. Thus,

$$\|\boldsymbol{Q}_i(\boldsymbol{\eta})\|\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_i^*\| \le O(\sum_{i=1}^N \sigma_i)(1+V(\boldsymbol{\eta})). \tag{28}$$

Similarly, using Assumption 2 and equality (26), we have

$$\begin{aligned}\|\boldsymbol{Q}_{N+1}(\boldsymbol{\eta})\| &= \|\tilde{\boldsymbol{M}}_{N+1}^0(\boldsymbol{\eta}) + \mathbf{g}(\boldsymbol{\mu})\| \le O(\sum_{i=1}^N \sigma_i), \\ \|\boldsymbol{Q}_{N+1}(\boldsymbol{\eta})\|\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^*\| &\le O(\sum_{i=1}^N \sigma_i)(1+V(\boldsymbol{\eta})). \end{aligned} \tag{29}$$

Finally, we bound the last two terms in (23). Since $\mathrm{E}(\xi - \mathrm{E}\xi)^2 \le \mathrm{E}\xi^2$ and taking into account (13), we have

$$\begin{aligned}&\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t),\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}\} \\ &\le \sum_{k=1}^d \int_{\mathbb{R}^{Nd}} J_i^{0^2}(\boldsymbol{x},\boldsymbol{\lambda})\frac{(x_k^i-\mu_k^i)^2}{\sigma_i^4(t)}p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x}. \end{aligned} \tag{30}$$

Thus, we can use Assumption 4, Remark 3, and the fact that $J_i^0(\boldsymbol{x},\boldsymbol{\lambda})$ is affine in $\boldsymbol{\lambda}$ to get the next inequality:

$$\begin{aligned}&\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t),\boldsymbol{\eta}(t))\|^2|\boldsymbol{\eta}(t)=\boldsymbol{\eta}\} \\ &\qquad\le f(\boldsymbol{\mu},\boldsymbol{\sigma}(t))\left(\frac{1}{\sigma_i^4(t)} + k_2V(\boldsymbol{\eta})\right), \end{aligned} \tag{31}$$

where $f(\boldsymbol{\mu}, \boldsymbol{\sigma}(t))$ is a polynomial of μ_i and $\sigma_i(t)$, $i \in [N]$, and k_2 is some positive constant.

Furthermore, taking into account boundedness of $\boldsymbol{\mu}(t)$ and affinity of the mapping $\boldsymbol{M}^0(\boldsymbol{\eta})$ with respect to $\boldsymbol{\lambda}$, we obtain the following bound for any $i \in [N]$:

$$\begin{aligned}
&\beta_i^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\} \\
&\leq \beta_i^2(t+1)(\|\boldsymbol{M}_i^0(\boldsymbol{\eta})\|^2 + \|\boldsymbol{Q}_i(\boldsymbol{\eta}(t))\|^2) \\
&\qquad + 2\beta_i^2(t+1)\|\boldsymbol{M}_i^0(\boldsymbol{\eta})\|\|\boldsymbol{Q}_i(\boldsymbol{\eta})\| \\
&\qquad + \beta_i^2(t+1)(\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\}) \\
&\leq \beta_i^2(t+1)(k_3 + k_4\|\boldsymbol{\lambda}\|^2 + O(\sigma_i^2(t))(1 + \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|)^2) \\
&\qquad + 2\beta_i^2(t+1)(k_5 + k_6\|\boldsymbol{\lambda}\|)O(\sigma_i(t)(1 + \|\boldsymbol{\eta} - \boldsymbol{\eta}^*\|)) \\
&\qquad\qquad + O(\gamma_i^2(t+1)) + O(\beta_i^2(t+1))V(\boldsymbol{\eta}) \\
&\leq O(\beta_i^2(t+1)(1 + \sigma_i^2(t) + \sigma_i(t)) + \gamma_i^2(t+1)) \\
&\qquad\qquad \times (1 + V(\boldsymbol{\eta})),
\end{aligned} \tag{32}$$

where k_j, $j = 3, \dots, 6$, are some positive constants.

For the term $\mathrm{E}\{\|\boldsymbol{G}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\}$ we can first derive the following bound:

$$\begin{aligned}
&\mathrm{E}\{\|\boldsymbol{R}_{N+1}(\mathbf{x}(t), \boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\} \\
&\qquad \leq \int_{\mathbb{R}^{Nd}} \|\mathbf{g}(\boldsymbol{x})\|^2 p(\boldsymbol{\mu}, \boldsymbol{x}) d\boldsymbol{x} = \phi(\boldsymbol{\mu}, \boldsymbol{\sigma}(t)),
\end{aligned}$$

where $\phi(\boldsymbol{\mu}, \boldsymbol{\sigma}(t))$ is a polynomial of μ_i and $\sigma_i(t)$, $i \in [N]$ (see Remark 3). Hence, according to boundedness of $\boldsymbol{\mu}(t)$ and the fact that $\sigma(t)$ goes to zero, we obtain

$$\begin{aligned}
&\beta_0^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_{N+1}(\boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\} \\
&\leq \beta_0^2(t+1)(\|\mathbf{g}(\boldsymbol{\mu})\|^2 + \|\boldsymbol{Q}_{N+1}(\boldsymbol{\eta}(t))\|^2) \\
&\qquad + 2\beta_0^2(t+1)\|\mathbf{g}(\boldsymbol{\mu})\|\|\boldsymbol{Q}_{N+1}(\boldsymbol{\eta})\| \\
&\qquad + \beta_0^2(t+1)(\mathrm{E}\{\|\boldsymbol{R}_{N+1}(\boldsymbol{\eta}(t))\|^2 | \boldsymbol{\eta}(t) = \boldsymbol{\eta}\}) \\
&\leq \beta_0^2(t+1)(k_7 + O(\sigma_{\max}^2(t)) + O(\sigma_{\max}(t))),
\end{aligned} \tag{33}$$

where k_7 is some positive constant.

Since $\boldsymbol{\eta}^*$ is a NE in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}_+^n)$, Assertion 1) in Theorem 1 implies that

$$(\boldsymbol{M}^0(\boldsymbol{\eta}^*), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \geq 0,$$

for any $\boldsymbol{\eta} \in \boldsymbol{A} \times \mathbb{R}_+^n$. According to pseudo-monotonicity of $\boldsymbol{M}^0$ in Assumption 2, the inequality above implies

$$(\boldsymbol{M}^0(\boldsymbol{\eta}), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \geq 0 \text{ for any } \boldsymbol{\eta} \in \boldsymbol{A} \times \mathbb{R}_+^n. \tag{34}$$

Thus, bringing (23), (28), (29), (32), and (33) together, we get

$$\begin{aligned}
LV(\boldsymbol{\eta}) \leq &- 2\beta_{\min}(t+1)(\boldsymbol{M}^0(\boldsymbol{\eta}), \boldsymbol{\eta} - \boldsymbol{\eta}^*) \\
&+ p(t)(1 + V(\boldsymbol{\eta})), \\
p(t) = &O(\beta_{\max}(t+1) - \beta_{\min}(t+1)) \\
&\quad + O(\beta_{\max}(t+1)\sigma_{\max}(t) + \gamma_{\max}^2(t+1)).
\end{aligned} \tag{35}$$

Hence, using conditions on parameters in Assumption 5, we get $\sum_{t=0}^{\infty} p(t) < \infty$. Finally, taking into account (34), (35), $\sum_{t=0}^{\infty} \beta_{\min}(t) = \infty$, and Theorem 4, we conclude that $\boldsymbol{\eta}(t)$ is finite almost surely for any $t \in \mathbb{Z}_+$ during the run of the algorithm irrespective of $\boldsymbol{\eta}(0)$. □

4.2 Convergence of the Algorithm Iterates

Having established boundedness of the iterates in the algorithm (6)-(7), in this subsection we prove Theorem 1.

Proof of Theorem 1. Under Assumption 6, the game can be reformulated as a constrained optimization problem. Indeed, if the game Γ is potential with a strictly convex potential function $f(\boldsymbol{a})$, then $\boldsymbol{M}(\boldsymbol{a}) = \nabla f(\boldsymbol{a})$ and the problem of finding a variational equilibrium is equivalent to solving

$$\begin{aligned} &\text{minimize } f(\boldsymbol{a}) \\ &\text{subject to } g_j(\boldsymbol{a}) \leq 0, \quad j = 1, \dots, m \\ &\qquad \boldsymbol{a} \in \boldsymbol{A} = A_1 \times \ldots \times A_N \subseteq \mathbb{R}^{Nn}. \end{aligned} \tag{36}$$

This equivalence follows from the fact that, due to the definition of potential games, the unique minimum solving the problem above is a variational Nash equilibrium in the game, whereas strict monotonicity of the game mapping $\boldsymbol{M}$ implies uniqueness of the solution of $VI(\mathcal{Q}, \boldsymbol{M})$ (see, for example, [32]) and, thus, uniqueness of the variational generalized Nash equilibrium. We call the above the primal problem.

Consider the Lagrangian function for the primal problem

$$\mathcal{L}(\boldsymbol{a}, \boldsymbol{\lambda}) = f(\boldsymbol{a}) + (\boldsymbol{\lambda}, \mathbf{g}(\boldsymbol{a})). \tag{37}$$

Then the dual function is defined as $\sup_{\boldsymbol{\lambda}\in\mathbb{R}^n_+} \inf_{\boldsymbol{a}\in\boldsymbol{A}} \mathcal{L}(\boldsymbol{a}, \boldsymbol{\lambda})$. Given Assumption 6, we conclude that for any primal-dual optimal pair $(\boldsymbol{a}^*, \boldsymbol{\lambda}^*)$ the vector $\boldsymbol{a}^*$ is the unique (due to strict monotonicity) variational Nash equilibrium in the game Γ. Under Assumption 3 the Karush-Kuhn-Tucker (KKT) [18] conditions hold and consequently, $(\boldsymbol{a}^*, \boldsymbol{\lambda}^*)$ is a primal-dual optimal pair if and only if:

(1) $\boldsymbol{a}^*$ is primal feasible and $\boldsymbol{\lambda}^* \in \mathbb{R}^n_+$ ($\boldsymbol{\lambda}^*$ is dual feasible);

(2) $\boldsymbol{a}^*$ attains the minimum in $\inf_{\boldsymbol{a}\in\boldsymbol{A}} \mathcal{L}(\boldsymbol{a}, \boldsymbol{\lambda}^*)$;

(3) $\boldsymbol{\lambda}^*$ attains the maximum in $\sup_{\boldsymbol{\lambda}\in\mathbb{R}^n_+} \mathcal{L}(\boldsymbol{a}^*, \boldsymbol{\lambda})$. Using the above characterization with the bounds derived in the previous section, we can establish convergence of the algorithm. First, from Inequality (35) we obtain that for any Nash equilibrium $\boldsymbol{\eta}^*$

$$\begin{aligned} &\mathrm{E}\{\|\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}^*\|^2 | \mathcal{F}_t\} \leq \|\boldsymbol{\eta}(t) - \boldsymbol{\eta}^*\|^2 \\ &-2\beta_{\min}(t)(\boldsymbol{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}^*) + h(t), \end{aligned} \tag{38}$$

where $\mathcal{F}_t$ is the σ-algebra generated by the random variables $\{\boldsymbol{\eta}(k), k \leq t\}$ and $h(t) = p(t)(1 + V(\boldsymbol{\eta}(t)))$. Due to Lemma 2, $\boldsymbol{\eta}(t)$ is bounded almost surely and, thus, according to the estimations for $p(t)$ in the proof of Lemma 2, $\sum_{t=0}^{\infty} h(t) < \infty$. Second, we will bound the term $(\boldsymbol{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}^*)$ using the Lagrangian of the game $\mathcal{L}(\boldsymbol{a}, \boldsymbol{\lambda})$.

Due to definition of the mapping $\boldsymbol{M}^0$ and Assumption 6,

$$\begin{aligned} (\boldsymbol{M}^0(\boldsymbol{\eta}(t)), \boldsymbol{\eta}(t) - \boldsymbol{\eta}^*) =& (\nabla_{\boldsymbol{\mu}}\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \\ &- (\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*), \end{aligned}$$

where $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is the Lagrangian function defined in (37). Due to convexity of f and $g_1, \dots, g_m$, according to KKT optimal conditions above, we get for any $\boldsymbol{\mu} \in \boldsymbol{A}$ and $\boldsymbol{\lambda} \in \mathbb{R}^n_+$

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \geq 0, \tag{39}$$

$$\mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}) \geq 0. \tag{40}$$

Due to convexity of $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ in $\boldsymbol{\mu}$ for any $\boldsymbol{\lambda} \in \mathbb{R}^n_+$

$$(\nabla_{\boldsymbol{\mu}}\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \geq \mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t)) \tag{41}$$

and due to linearity of $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ in $\boldsymbol{\lambda}$, we obtain

$$(\nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)), \boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*) = \mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}(t)) - \mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}^*). \tag{42}$$

Bringing (41) and (42) together and taking into account (39), (40), adding and subtracting $\mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ we conclude that

$$\begin{aligned}\mathrm{E}\{\|\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}^*\|^2|\mathcal{F}_t\} \leq \|\boldsymbol{\eta}(t) - \boldsymbol{\eta}^*\|^2 - &2\beta_{\min}(t) \\ \times [\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) + \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) - &\mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t))] \\ &+ h(t).\end{aligned} \tag{43}$$

Thus, we can apply the result of Robbins and Siegmund (Lemma 5) to the inequality (43) to conclude that almost surely

1) $\|\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}^*\|$ converges,

2) $\sum_{t=0}^{\infty} \beta_{\min}(t)(\mathcal{L}(\boldsymbol{\mu}(t), \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) + \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t))) < \infty$.

As $\sum_{t=0}^{\infty} \beta_{\min}(t) = \infty$, 2) implies that there exists a subsequence $\boldsymbol{\eta}(t_l) = (\boldsymbol{\mu}(t_l), \boldsymbol{\lambda}(t_l))$ such that almost surely

$$\begin{aligned}\lim_{l\to\infty} \{[\mathcal{L}(\boldsymbol{\mu}(t_l), \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)] \\ +[\mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t_l))]\} = 0.\end{aligned}$$

Since $\mathcal{L}(\boldsymbol{\mu}(t_l), \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \geq 0$ and $\mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) - \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t_l)) \geq 0$ for any l, the above holds if and only if

$$\begin{aligned}\lim_{l\to\infty} \mathcal{L}(\boldsymbol{\mu}(t_l), \boldsymbol{\lambda}^*) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \\ \lim_{l\to\infty} \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t_l)) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*).\end{aligned}$$

As a subsequence of $\{\boldsymbol{\eta}(t)\}$, the sequence $\boldsymbol{\eta}(t_l) = (\boldsymbol{\mu}(t_l), \boldsymbol{\lambda}(t_l))$ is bounded almost surely. Hence, we can choose an almost surely convergent subsequence $\boldsymbol{\eta}(t_{l_s}) = (\boldsymbol{\mu}(t_{l_s}), \boldsymbol{\lambda}(t_{l_s}))$ such that $\lim_{s\to\infty} \boldsymbol{\eta}(t_{l_s}) = \hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\lambda}})$ with probability 1, where $\hat{\boldsymbol{\mu}} \in \boldsymbol{A}$ and $\hat{\boldsymbol{\lambda}} \in \mathbb{R}^n_+$ (since $\boldsymbol{A}$ and $\{\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \mathbb{R}^n_+\}$ are closed). Due to the last two equalities above, almost surely

$$\begin{aligned}\lim_{s\to\infty} \mathcal{L}(\boldsymbol{\mu}(t_{l_s}), \boldsymbol{\lambda}^*) = \mathcal{L}(\hat{\boldsymbol{\mu}}, \boldsymbol{\lambda}^*) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) \\ \lim_{s\to\infty} \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}(t_{l_s})) = \mathcal{L}(\boldsymbol{\mu}^*, \hat{\boldsymbol{\lambda}}) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*).\end{aligned}$$

Since f is strictly convex and g_i, $i = 1, \ldots, m$ are convex over $\boldsymbol{A}$, the equality $\mathcal{L}(\hat{\boldsymbol{\mu}}, \boldsymbol{\lambda}^*) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \min_{\boldsymbol{\mu}\in\boldsymbol{A}} \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\lambda}^*)$ implies that $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}^*$. Moreover, due to dual feasibility of $\hat{\boldsymbol{\lambda}}$, the equality $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}^*$ together with the equality $\mathcal{L}(\boldsymbol{\mu}^*, \hat{\boldsymbol{\lambda}}) = \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}\in\mathbb{R}^n_+} \mathcal{L}(\boldsymbol{\mu}^*, \boldsymbol{\lambda})$ imply that $\hat{\boldsymbol{\eta}} = (\boldsymbol{\mu}^*, \hat{\boldsymbol{\lambda}})$ is an optimal primal dual pair and hence, a Nash equilibrium of the game. Since assertion 1) above holds for any Nash equilibrium $\boldsymbol{\eta}^*$, replacing $\boldsymbol{\eta}^*$ by $\hat{\boldsymbol{\eta}}$ in (43), we conclude that $\|\boldsymbol{\eta}(t+1) - \hat{\boldsymbol{\eta}}\|$ converges. Since there exists a subsequence $\boldsymbol{\eta}(t_{l_s})$ which converges to $\hat{\boldsymbol{\eta}}$, we get that almost surely $\lim_{t\to\infty} \boldsymbol{\eta}(t) = \hat{\boldsymbol{\eta}}$. Hence, $\Pr\{\lim_{t\to\infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*\} = 1$.

Finally, Assumption 5 implies that $\lim_{t\to\infty} \sigma_i(t) = 0$ for all $i \in [N]$. Taking into account that $\mathbf{x}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t))$, we conclude that $\mathbf{x}(t)$ converges weakly to a Nash equilibrium $\boldsymbol{a}^* = \boldsymbol{\mu}^*$ as time runs. Moreover, according to Portmanteau Lemma [17], this convergence is also in probability. □

5 Convergence with only local constraints

5.1 Convergence to Nash equilibria under relaxed conditions

We consider the game without coupling constraints, namely $C = \mathbb{R}^{Nd}$ (or equivalently $\mathbf{g} \equiv 0$), and relax the assumption existence of a potential function in the game Γ. In particular, Assumptions 1 and 6 are replaced by the following one.

Assumption 7. *For all $i \in [N]$ the set A_i is convex and compact, the cost function $J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})$ is defined on $\mathbb{R}^{Nd}$, continuously differentiable in $\boldsymbol{a}$ and the game mapping $\boldsymbol{M}$ is strictly monotone.*

In this case, as $C = \mathbb{R}^{Nd}$, the payoff-based procedure (6)-(7) is modified as follows:

$$\boldsymbol{\mu}^i(t+1) = \tag{44}$$
$$\mathrm{Proj}_{A_i}\left[\boldsymbol{\mu}^i(t) - \gamma_i(t+1)\sigma_i^2(t+1)\hat{J}_i(t)\frac{\mathbf{x}^i(t) - \boldsymbol{\mu}^i(t)}{\sigma_i^2(t)}\right],$$

where $i \in [N]$, $\gamma_i(t+1)$ is a step-size parameter chosen by player i and $\hat{J}_i(t) = J_i(\mathbf{x}(t))$ is the current observation of the ith player's cost function in the game Γ.

We make the following assumption regarding parameters $\gamma_i(t)$, $\sigma_i(t)$, and $\beta_i(t) = \gamma_i(t)\sigma_i^2(t)$ in the procedure (44).

Assumption 8. *The variance parameters $\sigma_i(t)$ and the step-size parameters $\gamma_i(t)$, $i \in [N]$, are chosen such that*
1) $\sum_{t=0}^{\infty} \beta_{min}(t) = \infty$,
2) $\sum_{t=0}^{\infty} \beta_{max}(t) - \beta_{min}(t) < \infty$,
3) $\sum_{t=0}^{\infty} \gamma_{max}^2(t)\sigma_{max}^2(t) < \infty$, $\sum_{t=0}^{\infty} \gamma_{max}(t)\sigma_{max}^3(t) < \infty$.

Theorem 2. *Let Assumptions 2-4, 7, and 8 hold in a game $\Gamma(N, \{A_i\}, \{J_i\}, \mathbb{R}^{nN})$. Let the players choose the states $\{\mathbf{x}^i(t)\}$ at time t according to the normal distribution $\mathcal{N}(\boldsymbol{\mu}^i(t), \sigma_i(t))$, where the mean parameters are updated as in (44). Then, as $t \to \infty$, the mean vector $\boldsymbol{\mu}(t)$ converges almost surely to the Nash equilibrium $\boldsymbol{\mu}^* = \boldsymbol{a}^*$ of the game Γ, given any initial vector $\boldsymbol{\mu}(0)$, and the joint state $\mathbf{x}(t)$ converges in probability to $\boldsymbol{a}^*$.*

Proof. As before, let $V(\boldsymbol{\mu}) = \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|^2$, where $\boldsymbol{\mu}^*$ is the unique Nash equilibrium in the game Γ (existence and uniqueness of $\boldsymbol{\mu}^*$ follows from Proposition 2.3.3 in [32]). Following the discussion in the proof of Lemma 2, we can rewrite the inequality (23) as follows

$$\begin{aligned} LV(\boldsymbol{\mu}) \le& -2\beta_{\min}(t+1)(\boldsymbol{M}(\boldsymbol{\mu}), \boldsymbol{\mu} - \boldsymbol{\mu}^*) \\ &+ 2(\beta_{\max}(t+1) - \beta_{\min}(t+1))k_1(1 + V(\boldsymbol{\mu})) \\ &+ 2\sum_{i=1}^{N}\beta_i(t+1)\|\boldsymbol{Q}_i^0(\boldsymbol{\mu})\|\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_i^*\| \\ &+ \sum_{i=1}^{N}\beta_i^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_i^0(\mathbf{x}(t), \boldsymbol{\mu}(t))\|^2 | \boldsymbol{\mu}(t) = \boldsymbol{\mu}\}, \text{where} \end{aligned} \tag{45}$$

$$\boldsymbol{G}_i^0(\mathbf{x}(t), \boldsymbol{\mu}(t)) = \boldsymbol{M}_i(\boldsymbol{\mu}(t)) + \boldsymbol{Q}_i^0(\boldsymbol{\mu}(t)) + \boldsymbol{R}_i^0(\mathbf{x}(t), \boldsymbol{\mu}(t)), \tag{46}$$

and $\boldsymbol{Q}_i^0(\boldsymbol{\mu}(t))$, $\boldsymbol{R}_i^0(\mathbf{x}(t), \boldsymbol{\mu}(t))$ are obtain from $\boldsymbol{Q}_i(\boldsymbol{\eta}(t))$, $\boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\eta}(t))$ by letting $\mathbf{g} \equiv 0$. Similarly to (27) and (28),

$$\|\boldsymbol{Q}_i^0(\boldsymbol{\mu})\| = O(\sum_{i=1}^{N}\sigma_i) \tag{47}$$

$$\|\boldsymbol{Q}_i^0(\boldsymbol{\mu})\|\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_i^*\| \le O(\sum_{i=1}^{N}\sigma_i)(1 + V(\boldsymbol{\mu})). \tag{48}$$

Moreover, similarly to (30), we conclude that

$$\begin{aligned} &\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\mu}(t))\|^2 | \boldsymbol{\mu}(t) = \boldsymbol{\mu}\} \\ &\le \sum_{k=1}^{d}\int_{\mathbb{R}^{Nd}} {J_i}^2(\boldsymbol{x})\frac{(x_k^i - \mu_k^i)^2}{\sigma_i^4(t)}p(\boldsymbol{\mu}, \boldsymbol{x})d\boldsymbol{x}. \end{aligned} \tag{49}$$

Thus, we can use Assumption 4 to get

$$\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t),\boldsymbol{\mu}(t))\|^2 | \boldsymbol{\mu}(t) = \boldsymbol{\mu}\} \le \frac{f(\boldsymbol{\mu}, \boldsymbol{\sigma}(t))}{\sigma_i^2(t)}, \tag{50}$$

where $f(\boldsymbol{\mu}, \boldsymbol{\sigma}(t))$ is a quadratic function of μ_i and polynomial in $\sigma_i(t)$, $i \in [N]$. Taking into account (46)-(50) and Assumption 4, we conclude that for some positive constants k_2, k_3

$$\begin{aligned}
&\beta_i^2(t+1)\mathrm{E}\{\|\boldsymbol{G}_i(\mathbf{x}(t), \boldsymbol{\mu}(t))\|^2 | \boldsymbol{\mu}(t) = \boldsymbol{\mu}\} \\
&\leq \beta_i^2(t+1)(\|\boldsymbol{M}_i(\boldsymbol{\mu})\|^2 + \|\boldsymbol{Q}_i^0(\boldsymbol{\mu}(t))\|^2) \\
&\qquad\qquad + 2\beta_i^2(t+1)\|\boldsymbol{M}_i(\boldsymbol{\mu})\|\|\boldsymbol{Q}_i^0(\boldsymbol{\mu})\| \\
&\qquad\qquad + \beta_i^2(t+1)(\mathrm{E}\{\|\boldsymbol{R}_i(\mathbf{x}(t), \boldsymbol{\mu}(t))\|^2 | \boldsymbol{\mu}(t) = \boldsymbol{\mu}\}) \\
&\leq \beta_i^2(t+1)(k_2 + O(\sigma_i^2(t))(1 + \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|)^2) \\
&\qquad\quad + 2\beta_i^2(t+1)k_3(1 + \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|)O(\sigma_i(t)) \\
&\qquad\qquad\qquad + O(\gamma_i^2(t+1)\sigma_i^2(t+1))(1 + V(\boldsymbol{\mu})) \\
&\leq O(\beta_i^2(t+1) + \gamma_i^2(t+1)\sigma_i^2(t+1))(1 + V(\boldsymbol{\mu})).
\end{aligned} \tag{51}$$

Next, bringing estimations (48) and (51) into (45), we obtain

$$\begin{aligned}
LV(\boldsymbol{\mu}) \leq &-2\beta_{\min}(t+1)(\boldsymbol{M}(\boldsymbol{\mu}), \boldsymbol{\mu} - \boldsymbol{\mu}^*) \\
&+ O(s(t))(1 + V(\boldsymbol{\mu})),
\end{aligned} \tag{52}$$

where $s(t) = \beta_{\max}(t) - \beta_{\min}(t) + \gamma_{\max}(t)\sigma^3_{\max}(t) + \gamma^2_{\max}(t)\sigma^2_{\max}(t)$. Taking into account that $\boldsymbol{A}$ is compact, we conclude that

$$\begin{aligned}
\mathrm{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}^*\|^2 | \mathcal{F}_t\} \leq \|&\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2 \\
&- 2\beta_{\min}(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \\
&+ O(s(t)).
\end{aligned} \tag{53}$$

Moreover, for any t we have $(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) > (\boldsymbol{M}(\boldsymbol{\mu}^*)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \geq 0$, due to pseudo-monotonicity of $\boldsymbol{M}$ and the fact that $\boldsymbol{\mu}^*$ is the Nash equilibrium. Thus, using Theorem 5 we conclude that almost surely

1) $\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}^*\|$ converges,

2) $\sum_{t=0}^{\infty} \beta_{\min}(t)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) < \infty$.

As $\sum_{t=0}^{\infty} \beta_{\min}(t) = \infty$, 2) implies that there exists a subsequence $\boldsymbol{\mu}(t_l)$ such that almost surely

$$\lim_{l \to \infty} (\boldsymbol{M}(\boldsymbol{\mu}(t_l)), \boldsymbol{\mu}(t_l) - \boldsymbol{\mu}^*) = 0. \tag{54}$$

Since $\boldsymbol{\mu}(t_l)$ is bounded almost surely for any l, we can choose a convergent subsequence $\boldsymbol{\mu}(t_{l_s})$ such that $\lim_{s\to\infty} \boldsymbol{\mu}(t_{l_s}) = \boldsymbol{\mu}'$ for some $\boldsymbol{\mu}'$. Hence, due to (54),

$$(\boldsymbol{M}(\boldsymbol{\mu}'), \boldsymbol{\mu}' - \boldsymbol{\mu}^*) = 0, \tag{55}$$

which together with strict monotonicity of $\boldsymbol{M}$ implies $\boldsymbol{\mu}' = \boldsymbol{\mu}^*$. Thus, as $\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}^*\|$ converges almost surely and there exists a subsequence $\boldsymbol{\mu}(t_{l_s})$ which converges to $\boldsymbol{\mu}^*$ almost surely, we get that $\Pr\{\lim_{t\to\infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*\} = 1$. Finally, Assumption 5 implies that $\lim_{t\to\infty} \sigma_i(t) = 0$ for all $i \in [N]$. Taking into account that $\mathbf{x}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t))$, we conclude that $\mathbf{x}(t)$ converges weakly to a Nash equilibrium $\boldsymbol{a}^* = \boldsymbol{\mu}^*$ as time runs. Moreover, according to Portmanteau Lemma [17], this convergence is also in probability. □

Remark 6. *In a prior work [48], we showed convergence to Nash equilibria under pseudo-monotonicity of the game mapping, leveraging the proof technique in [54]. Recently, in [12] it was shown with a counterexample that pseudo-monotonicity is not sufficient for the convergence results in [54]. This implies that our payoff-based algorithm also would not converge to a Nash equilibrium under merely the pseudo-monotonicity assumption. However, a closer look at the required conditions following Equation* (55) *reveals that the proof remains valid if instead of strictly monotone, the game mapping is a) pseudo-monotone and additionally satisfies b) $\forall \boldsymbol{\mu} \in \boldsymbol{A}$ and $\boldsymbol{\mu}^*$ Nash equilibrium, $(\boldsymbol{M}(\boldsymbol{\mu}'), \boldsymbol{\mu}' - \boldsymbol{\mu}^*) = 0 \Rightarrow \boldsymbol{\mu}' = \boldsymbol{\mu}^*$. This latter condition holds for example, when $\boldsymbol{M}$ is pseudo-monotone and co-coercive. Hence, our new proof method of Theorem 2 corrects our mistake in [48].*

The challenge in generalizing the proof of Theorem 1 to non-potential games lies in the fact that the extended game mapping $\boldsymbol{M}^0$ cannot be strictly monotone, nor can it be co-coercive. This implies that convergence of a subsequence of $\{\boldsymbol{\eta}(t)\}$ will not suffice to establish convergence of the sequence to a Nash equilibrium. Nevertheless, given a strictly convex potential function in the game we could use equivalence of the variational Nash equilibrium to an optimal primal dual pair for the Lagrangian, and establish convergence of the sequence of iterates to the variational Nash equilibrium.

5.2 Convergence rate of the algorithm

Below, we show that if the strict monotonicity condition for the game mapping in Assumption 7 is strengthened to strong monotonicity, we obtain a convergence rate for the procedure (44) as a function of the stepsize and variance parameters.

Theorem 3. *Let Assumptions 2-4, and 7 hold and $\boldsymbol{M}$ be strongly monotone with strong monotonicity constant $\kappa > 1$. Furthermore, assume the time step and variance parameters are chosen as $\gamma_i(t) = \frac{1}{(t+R_i)^a}$, $\sigma_i(t) = \frac{1}{(t+R_i)^b}$, $i \in [N]$, where $a + 2b \in (0.5, 1]$, $2a > 1$, and $a + 3b > 1$. Then in the long run of the algorithm* (10)-(11)

$$\mathrm{E}\{\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2\} \leq \frac{C}{t^{2(a+b)-1}} = O(1/t^{2(a+b)-1}), \tag{56}$$

where C is some positive constant and $\boldsymbol{\mu}^*$ is the unique equilibrium of the game Γ to which the vector $\boldsymbol{\mu}(t)$ converges almost surely.

Proof. We we will use the following lemma.

Lemma 3. *Let the sequence $\{a_t\}$, $a_t \geq 0$ $t \in \mathbb{Z}_+$, satisfy the following iteration:*

$$a_{t+1} \leq (1 - \kappa/t)a_t + \psi/t^c,$$

for some constants $1 < c \leq 2$, $\kappa > 1$, $\psi > 0$. Then, $a_t \leq \frac{C}{t^{c-1}}$, where $C = \max\{a_0, \frac{\psi}{\kappa-1}\}$.

Please see the appendix for the proof.

Using estimation (53) and the fact that $(\boldsymbol{M}(\boldsymbol{\mu}^*), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \geq 0$ for any t, we obtain

$$\begin{aligned}\mathrm{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}^*\|^2|\mathcal{F}_t\} &\leq \|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2 \\ &\quad - 2\beta_{\min}(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)) - \boldsymbol{M}(\boldsymbol{\mu}^*), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \\ &\qquad + O(s(t)) \\ &\leq \left(1 - \frac{\kappa}{t}\right)\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2 + \frac{\psi}{t^{2(a+b)}},\end{aligned} \tag{57}$$

where in the last inequality we used the strong monotonicity of the map $\boldsymbol{M}$ over $\boldsymbol{A}$, conditions for the parameters $\gamma_i(t)$, $\sigma_i(t)$, and definition of β_i, which implies that there exists t_0 such that $\beta_{\min}(t) \geq \frac{1}{2t}$ for any $t \geq t_0$. Finally, taking into account that $1 < 2(a+b) \leq 2$ and using Lemma 3 for $c = 2(a+b)$, we conclude that $\mathrm{E}\{\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2\} \leq C/t^{2(a+b)-1}$, where $C = \max\{\mathrm{E}\{\|\boldsymbol{\mu}(0) - \boldsymbol{\mu}^*\|^2\}, \frac{\psi}{\kappa-1}\}$. □

The result above demonstrates that the convergence rate of the proposed payoff-based learning procedure is sublinear. This is consistent with the results on related optimization algorithms based on the stochastic approximation techniques [15, 49]. Note also that Theorem 3 presents the asymptotic estimation of the convergence rate, whereas its tightness and more details on characterization of the constant C need to be analyzed separately and are subject of our future work.

6 Numerical Case Study

We illustrate the proposed payoff-based learning approach through a game arising in a classical Cournot economic model. There are N firms, each producing a good and each needs to determine its production amount. Each firm (referred to also as a player or an agent) has an individual production cost $Q^i(\boldsymbol{a}^i)$ and receives a payment $p(\boldsymbol{a})\boldsymbol{a}^i$ for the quantity produced $\boldsymbol{a}^i$. The price $p(\boldsymbol{a})$ depends on the total production of all firms. The production of the firms is coupled by the fact that there is a network capacity constraint [1]. Such constraints may arise from the amount that can be delivered through a link to the consumers (consider for example an electricity network with line limits). In contrast to past approaches on computing Nash equilibria, we consider a scenario in which the form of the price function p is unknown to agents and there is no communication graph between the agents.

Let $\boldsymbol{a}^i = [a^i_1, \dots, a^i_d]^\top \in \mathbb{R}^d$ denote the decision variable of firm i (also referred to as player or agent), $i \in [N]$, which is its production level over a horizon of d steps[4]. Each player has a limit on maximum production at each step

$$0 \leq a^i_k \leq \bar{a}^i \quad \text{for } k = 1, \dots, d. \tag{58}$$

The convex and compact set defined by the constraints above is considered the action set A_i for player i. The coupling constraints arising from a network capacity limit is

$$\sum_{i=1}^{N} a^i_k \leq \bar{a}^k \quad \text{for } k = 1, \dots, d. \tag{59}$$

For the simulations, we consider a linear price function and quadratic production cost functions, which are standard assumptions in Cournot models [19, 13, 22, 30]. The function to be minimized by each agent can then be compactly written as

$$\begin{aligned} J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) &= Q^i(\boldsymbol{a}^i) - p(\boldsymbol{a})\boldsymbol{a}^i \\ &= \boldsymbol{a}^{i\top} Q^i \boldsymbol{a}^i + 2(C\frac{1}{N}\sum_{j=1}^{N} \boldsymbol{a}^j + \mathbf{c})^\top \boldsymbol{a}^i, \end{aligned} \tag{60}$$

with $Q^i, C \in \mathbb{R}^{d\times d}$, $\mathbf{c} \in \mathbb{R}^d$ for all $i \in [N]$. The production cost is assumed convex and hence $Q^i \in \mathbb{R}^n_+$, whereas $C \in \mathbb{R}^n_+$ follows from the fact that the price is a decreasing function of total production [13, 52, 1, 19, 27, 32]. It is readily verified that the resulting game mapping (see Definition (2)) is affine and, hence, Lipschitz on $\mathbb{R}^{Nd}$. Moreover, the game mapping is symmetric positive definite and hence, the game admits a strongly convex potential function. Consequently, Assumptions 1-6 hold.

We let the agents follow the payoff-based algorithm described by (6)-(7) to find their Nash equilibrium strategies. Each player submits its proposed production profile over time horizon of d units, $\mathbf{x}^i(t) = [x^i_1(t), \dots, x^i_d(t)]^\top$ at iteration t. It then observes J^0_i consisting of the cost functions corresponding to prices of the good, the violation of the coupling constraint, as well as its individual production cost.

For the simulation, we let $d = 4$, the matrices Q^i, $i \in [N]$ and C, in (60) are the identity matrices in $\mathbb{R}^{d\times d}$, and the vector $\mathbf{c} \in \mathbb{R}^d$ is chosen randomly according to a normal distribution. The action set A_i for each player $i \in [N]$ is defined by (58)-(59), where $\bar{a}^i = 9$ and $\bar{a}^k$ is a random variable taking values in the interval $(3N, 3N + 100)$. The initial vector $(\boldsymbol{\mu}(0), \boldsymbol{\lambda}(0))$ is chosen from a uniform distribution on $\boldsymbol{A} \times [0, 5]$.

Figure 1 presents the relative error $\frac{\|\boldsymbol{\mu}(t)-\boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$ during the algorithm's run for $N = 3, 10, 30$, where $\boldsymbol{a}^*$ is the unique generalized Nash equilibrium of the game. We see that after the first iteration the iterates quickly approach the Nash equilibrium. However, convergence of the error to zero is slow. The slow decrease of the relative error after the first iteration can be explained by the choice of the rapidly decreasing parameter $\sigma(t)$ and $\gamma(t)$. The convergence is also slower for increasing number of players in the game. It is interesting to derive explicit dependence of the convergence rate derived based on the number of players.

7 Conclusion

We proposed a novel payoff-based learning approach for convergence to variational Nash equilibria in convex games with jointly convex coupling constraints. In this approach, each agent determined its next state by

[4] The formulation here also can be interpreted as production levels of each firm at d different locations [1].

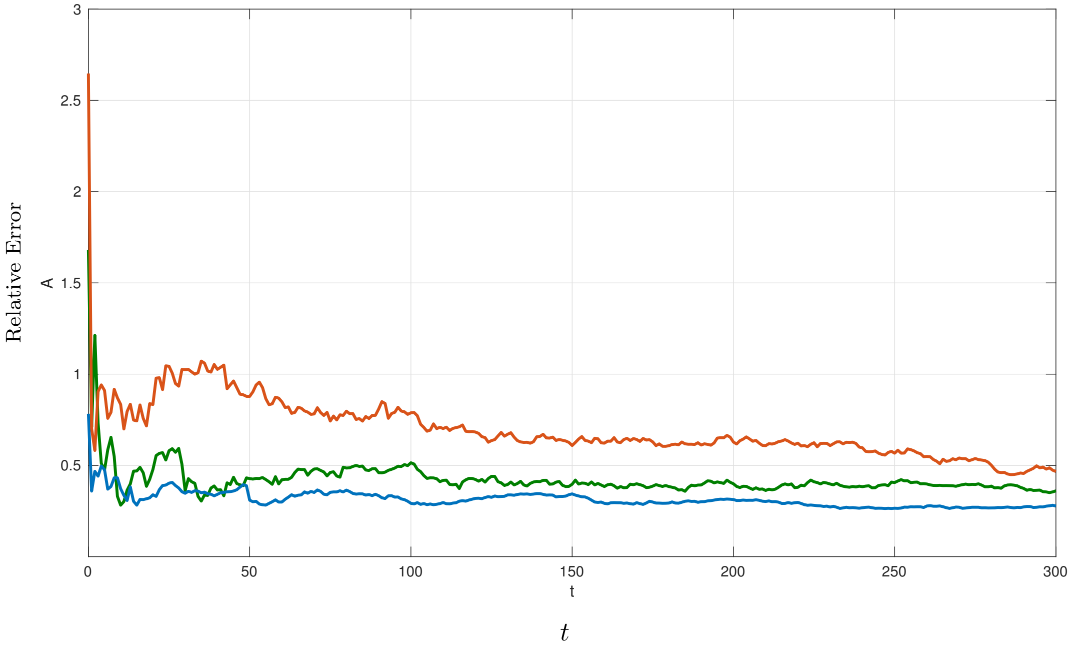

Figure 1: Relative error $\frac{\|\boldsymbol{\mu}(t)-\boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$ during the payoff-based algorithm, $N = 3$ (blue line), $N = 10$ (green line), $N = 30$ (red line).

sampling from a Gaussian distribution, whose mean was updated using the payoff information. We proved almost sure convergence of the means of the distributions to a variational Nash equilibrium, given appropriate choice of algorithms' step-sizes and variances of the distributions. The convergence result relied on existence of a strictly convex potential function. In the absence of coupling constraint, convergence to a Nash equilibrium was established based on strict monotonicity of the game mapping. Furthermore, in this case, under strong monotonicity of the game mapping, the convergence rate of the algorithm was derived. Further relaxing conditions for convergence of the payoff-based algorithm with and without coupling constraint is subject of our current work. We are also developing algorithms that ensure constraint satisfaction during the algorithm iterates.

8 Acknowledgement

We thank Sergio Grammatico for informing us of the recent publication [12], in which it was shown that monotonicity is not a sufficient condition for convergence of a class of forward-backward algorithms. This implied a bug in our convergence proof for the payoff-based algorithm proposed in [48]. In the current paper, we corrected this mistake by deriving stronger convergence conditions in Theorem 2.

References

[1] M. Abolhassani, M. H. Bateni, M..T. Hajiaghayi, H. Mahini, and A. Sawant. Network Cournot competition. In *International Conference on Web and Internet Economics*, pages 15–29. Springer, 2014.

[2] G. Arslan, J. R. Marden, and J. S. Shamma. Autonomous vehicle-target assignment: a game theoretical formulation. *ASME Journal of Dynamic Systems, Measurement and Control*, 129:584–596, September 2007.

[3] B. Bharath and V. S. Borkar. Stochastic approximation algorithms: Overview and recent trends. *Sadhana*, 24(4):425–452, 1999.

[4] L. E. Blume. The statistical mechanics of strategic interaction. *Games and economic behavior*, 5(3):387–424, 1993.

[5] P. S. Dutta, N. R. Jennings, and L. Moreau. Cooperative information sharing to improve distributed learning in multi-agent systems. *Journal of Artificial Intelligence Research*, 24:407–463, 2005.

[6] F. Facchinei, A. Fischer, and V. Piccialli. On generalized Nash games and variational inequalities. *Operations Research Letters*, 35(2):159 – 164, 2007.

[7] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *4OR*, 5(3):173–210, 2007.

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

[9] B. Gentile, F. Parise, D. Paccagnan, M. Kamgarpour, and J. Lygeros. Nash and wardrop equilibria in aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, 2018. to appear.

[10] T. Goto, T. Hatanaka, and M. Fujita. Payoff-based inhomogeneous partially irrational play for potential game theoretic cooperative control: Convergence analysis. In *American Control Conference (ACC), 2012*, pages 2380–2387, June 2012.

[11] M. S. Gowda. Affine pseudomonotone mappings and the linear complementarity problem. *SIAM Journal on Matrix Analysis and Applications*, 11(3):373–380, July 1990.

[12] S. Grammatico. Comments on distributed robust adaptive equilibrium computation for generalized convex games (automatica 63(2016) 82-91). *Automatica*, 97:186 – 188, 2018.

[13] B. F. Hobbs. LCP models of Nash-Cournot competition in bilateral and POOLCO-based power markets. In *Power Engineering Society 1999 Winter Meeting, IEEE*, volume 1, pages 303–308. IEEE, 1998.

[14] M. K. Jensen. Aggregative games and best-reply potentials. *Economic Theory*, 43(1):45–66, 2010.

[15] A. Juditsky, A. Nemirovski, and C. Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stoch. Syst.*, 1(1):17–58, 2011.

[16] A. Kannan and U. V. Shanbhag. Distributed Computation of Equilibria in Monotone Nash Games via Iterative Regularization Techniques. *SIAM Journal on Optimization*, 22(4):1177–1205, 2012.

[17] A. Klenke. *Probability theory: a comprehensive course.* Springer, London, 2008.

[18] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492. University of California Press, 1951.

[19] N. Kukushkin. *Cournot Oligopoly with "almost" Identical Convex Costs.* Instituto Valenciano de Investigaciones Económicas, 1993.

[20] A. A. Kulkarni and U. V. Shanbhag. New insights on generalized Nash games with shared constraints: Constrained and variational equilibria. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 151–156. IEEE, 2009.

[21] N. Li and J. R. Marden. Designing games for distributed optimization. *IEEE Journal of Selected Topics in Signal Processing*, 7(2):230–242, 2013. Special issue on adaptation and learning over complex networks.

[22] Z. Ma, D. Callaway, and I. Hiskens. Decentralized charging control for large populations of plug-in electric vehicles. In *49th IEEE conference on decision and control*, pages 206–212. IEEE, 2010.

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

[24] J. R. Marden, S. D. Ruben, and L. Y. Pao. A model-free approach to wind farm control using game theoretic methods. *IEEE Trans. Contr. Sys. Techn.*, 21(4):1207–1214, 2013.

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

[26] J. R. Marden, H. P. Young, Gürdal Arslan, and J. S. Shamma. Payoff-based dynamics for multiplayer weakly acyclic games. *SIAM J. Control and Optimization*, 48(1):373–396, 2009.

[27] C. Metzler, B. F. Hobbs, and J.-S. Pang. Nash-Cournot equilibria in power markets on a linearized DC network with arbitrage: Formulations and properties. *Networks and Spatial Economics*, 3(2):123–150, 2003.

[28] Y. Nesterov. Random gradient-free minimization of convex functions. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2011. No. 2011001.

[29] M. B. Nevelson and R. Z. Khasminskii. *Stochastic approximation and recursive estimation [translated from the Russian by Israel Program for Scientific Translations ; translation edited by B. Silver]*. American Mathematical Society, 1973.

[30] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros. Distributed computation of generalized nash equilibria in quadratic aggregative games with affine coupling constraints. In *IEEE Conference on Decision and Control*, pages 6123–6128, Dec 2016.

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

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

[33] S. Perkins, P. Mertikopoulos, and D. S. Leslie. Mixed-strategy learning with continuous action sets. *IEEE Transactions on Automatic Control*, (open access), 2015.

[34] B. Pradelski and H. P. Young. Learning efficient Nash equilibria in distributed systems. *Games and Economic behavior*, 75(2):882–897, 2012.

[35] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Herbert Robbins Selected Papers*, pages 111–135. Springer, 1985.

[36] J. B. Rosen. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, 33(3):520–534, 1965.

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

[38] F. Salehisadaghiani and L. Pavel. Nash equilibrium seeking by a gossip-based algorithm. In *53rd IEEE Conference on Decision and Control*, pages 1155–1160, Dec 2014.

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

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

[41] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang. Monotone games for cognitive radio systems. In *Distributed Decision Making and Control*, pages 83–112. Springer, 2012.

[42] J. S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3):312–327, March 2005.

[43] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2014.

[44] M. S Stankovic, K. H. Johansson, and D. M. Stipanovic. Distributed seeking of Nash equilibria with applications to mobile sensor networks. *IEEE Transactions on Automatic Control*, 57(4):904–919, 2012.

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

[46] T. Tatarenko. Stochastic payoff-based learning in multi-agent systems modeled by means of potential games. In *IEEE 55th Conference on Decision and Control*, pages 5298–5303, Dec 2016.

[47] T. Tatarenko. Stochastic stability of potential function maximizers in continuous version of independent log-linear learning. In *European Control Conference*, pages 210–215, June 2016.

[48] T. Tatarenko and M. Kamgarpour. Payoff-Based Approach to Learning Nash Equilibria in Convex Games. In *The 20th World Congress of the International Federation of Automatic Control, IFAC*, 2017. submitted.

[49] T. Tatarenko and B. Touri. Non-convex distributed optimization. *IEEE Transactions on Automatic Control*, PP(99):1–1, 2017.

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

[51] A. L. Thathachar and P. S. Sastry. *Networks of Learning Automata: Techniques for Online Stochastic Optimization*. Springer US, 2003.

[52] J. W. Weibull. Price competition and convex costs. Technical report, SSE/EFI Working Paper Series in Economics and Finance, 2006. No. 622.

[53] H. Yin, U. V. Shanbhag, and P. G. Mehta. Nash equilibrium problems with scaled congestion costs and shared constraints. *IEEE Transactions on Automatic Control*, 56(7):1702–1708, July 2011.

[54] M. Zhu and E. Frazzoli. Distributed robust adaptive equilibrium computation for generalized convex games. *Automatica*, 63:82 – 91, 2016.

[55] M. Zhu and S. Martínez. Distributed coverage games for energy-aware mobile sensor networks. *SIAM J. Control and Optimization*, 51(1):1–27, 2013.

.1 Proofs

Proof of Lemma 1. Assume the set Y is compact and is expressed by the following inequality constraints: $Y = \{\mathbf{y} \in \mathbb{R}^d : \boldsymbol{h}(\mathbf{y}) = [h_1(\mathbf{y}), \dots, h_m(\mathbf{y})]^T \leq \mathbf{0}\}$, where $h_i : \mathbb{R}^d \to \mathbb{R}$, $i \in [m]$, are convex functions defined on R^d. Then, using Slater's condition for the set Y and continuity of $\boldsymbol{T}$, conditions analogous to the Karush-Kuhn-Tucker ones can be formulated for $SOL(Y, \boldsymbol{T})$ (see Proposition 1.3.4 in [32]). Namely, $\mathbf{y}^* \in SOL(Y, \boldsymbol{T})$ if and only if there exists $\boldsymbol{\nu} \in \mathbb{R}^m$ such that

$$\begin{aligned} \mathbf{0} &= \boldsymbol{T}(\boldsymbol{y}^*) + \sum_{i=1}^{m} (\nu_i, \nabla h_i(\boldsymbol{y}^*)), \\ 0 &= (\boldsymbol{\nu}, \boldsymbol{h}(\mathbf{y}^*)), \quad \boldsymbol{\nu} \geq \mathbf{0}, \quad \boldsymbol{h}(\mathbf{y}^*) \leq \mathbf{0}. \end{aligned} \tag{61}$$

Thus, we can associate a multiplier $\boldsymbol{\nu}$ with any solution $\mathbf{y}^*$ of $VI(Y, \boldsymbol{T})$. We further call the vector $(\mathbf{y}^*, \boldsymbol{\nu})$ a *KKT tuple*.

Now consider the associated game $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ defined in (4). Note that the variational equilibria in game Γ are characterized as $SOL(\mathcal{Q}, \boldsymbol{M})$. Hence, $\boldsymbol{a}^*$ is a variational equilibrium in Γ if and only if $(\boldsymbol{a}^*, \boldsymbol{\nu})$ is a KKT tuple for $VI(\mathcal{Q}, \boldsymbol{M})$. And from Lemma 3 in [30], under Assumptions 1 and 3, $(\boldsymbol{a}^*, \boldsymbol{\nu})$ is a KKT tuple for the $VI(\mathcal{Q}, \boldsymbol{M})$ if and only if the pair $[\boldsymbol{a}^*, \boldsymbol{\lambda}^*]$ is a Nash equilibrium of the game $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$, where $\boldsymbol{\lambda}^*$ denotes the coordinate of the multiplier $\boldsymbol{\nu}$ corresponding to the constraint $\mathbf{g}(\boldsymbol{a}) \leq \mathbf{0}$. Thus, Assertion 1) and 2) are proven. Furthermore, taking into account Propositions 1.3.4 and 1.4.2 in [32] such a KKT tuple exists since $VI(\mathcal{Q}, \boldsymbol{M})$ has a solution. It follows that there exists a Nash equilibrium in $\Gamma_a(\boldsymbol{A} \times \mathbb{R}^n_+)$ and Assertion 3) is proven. Finally, Assertion 4) holds since under Assumptions 1 and 3, Lemma 5.1 in [54] shows that $\|\boldsymbol{\lambda}^*\|$ is bounded.

□

Proof of Lemma 3. The proof is based on a standard rate estimate, analogous to the result in (5.292) in [43]. We prove the claim by induction. Let us assume that $a_0 \geq \frac{\psi}{\kappa - 1}$. Then, according to the induction step,

$$a_{t+1} \leq \left(1 - \frac{\kappa}{t}\right) \frac{a_0}{t^{c-1}} + \frac{\psi}{t^c}.$$

Thus, it suffices to show that the right-hand-side of the inequality above is not more than $\frac{a_0}{(t+1)^{c-1}}$. As $\frac{\psi}{a_0} - \kappa \leq -1$,

$$\begin{aligned}\left(1 - \frac{\kappa}{t}\right) \frac{a_0}{t^{c-1}} + \frac{\psi}{t^c} = \frac{a_0}{t^c} \left(t - \kappa + \frac{\psi}{a_0}\right) &\leq \frac{a_0}{t^c}(t-1) \\ &\leq \frac{a_0}{(t+1)^{c-1}},\end{aligned}$$

since $a_0 > 0$ and $\left(1 + \frac{1}{t}\right)^{c-1} \leq \left(1 + \frac{1}{t-1}\right)$. The case $a_0 \leq \frac{\psi}{\kappa - 1}$ can be considered analogously.

□

.2 Supporting Theorems

Let $\{\mathbf{X}(t)\}_t$, $t \in \mathbb{Z}_+$, be a discrete-time Markov process on some state space $E \subseteq \mathbb{R}^d$, namely $\mathbf{X}(t) = \mathbf{X}(t, \omega) : \mathbb{Z}_+ \times \Omega \to E$, where Ω is the sample space of the probability space on which the process $\mathbf{X}(t)$ is defined. The transition function of this chain, namely $\Pr\{\mathbf{X}(t+1) \in \Gamma | \mathbf{X}(t) = \mathbf{X}\}$, is denoted by $P(t, \mathbf{X}, t+1, \Gamma)$, $\Gamma \subseteq E$.

Definition 8. *The operator L defined on the set of measurable functions $V : \mathbb{Z}_+ \times E \to \mathbb{R}$, $\mathbf{X} \in E$, by*

$$\begin{aligned}LV(t, \mathbf{X}) &= \int P(t, \mathbf{X}, t+1, dy)[V(t+1, y) - V(t, \mathbf{X})] \\ &= E[V(t+1, \mathbf{X}(t+1)) \mid \mathbf{X}(t) = \mathbf{X}] - V(t, \mathbf{X}),\end{aligned}$$

is called a generating operator *of a Markov process $\{\mathbf{X}(t)\}_t$.*

Next, we formulate the following theorem for discrete-time Markov processes, which is proven in [29], Theorem 2.5.2.

Theorem 4. *Consider a Markov process $\{\mathbf{X}(t)\}_t$ and suppose that there exists a function $V(t, \mathbf{X}) \geq 0$ such that $\inf_{t \geq 0} V(t, \mathbf{X}) \to \infty$ as $\|\mathbf{X}\| \to \infty$ and*

$$LV(t, \mathbf{X}) \leq -\alpha(t+1)\psi(t, \mathbf{X}) + f(t)(1 + V(t, \mathbf{X})),$$

where $\psi \geq 0$ on $\mathbb{R} \times \mathbb{R}^d$, $f(t) > 0$, $\sum_{t=0}^{\infty} f(t) < \infty$. Let $\alpha(t)$ be such that $\alpha(t) > 0$, $\sum_{t=0}^{\infty} \alpha(t) = \infty$. Then, almost surely $\sup_{t \geq 0} \|\mathbf{X}(t, \omega)\| = R(\omega) < \infty$.

The following is a well-known result of Robbins and Siegmund on non-negative random variables [35].

Theorem 5. *Let (Ω, F, P) be a probability space and $F_1 \subset F_2 \subset \ldots$ a sequence of sub-σ-algebras of F. Let z_t, b_t, ξ_t, and ζ_t be non-negative F_t-measurable random variables satisfying*

$$\mathrm{E}(z_{t+1} | F_t) \leq z_t(1 + b_t) + \xi_t - \zeta_t.$$

Then, almost surely $\lim_{t \to \infty} z_t$ exists and is finite for the case in which $\{\sum_{t=1}^{\infty} b_t < \infty, \sum_{t=1}^{\infty} \xi_t < \infty\}$. Moreover, in this case, $\sum_{t=1}^{\infty} \zeta_t < \infty$ almost surely.