Payoff-Based Approach to Learning Nash Equilibria in Convex Games [*]

T. Tatarenko [*] **M. Kamgarpour** [**]

[*] *Department of Control Theory and Robotics, TU Darmstadt, Germany (e-mail: tatarenk@rmr.tu-darmstadt.de)*
[**] *Automatic Control Laboratory, Swiss Federal Institute of Technology, Zurich, Switzerland (e-mail: mkamgar@control.ee.ethz.ch)*

Abstract: We consider multi-agent decision making, where each agent optimizes its cost function subject to constraints. Agents' actions belong to a compact convex Euclidean space and the agents' cost functions are coupled. We propose a distributed payoff-based algorithm to learn Nash equilibria in the game between agents. Each agent uses only information about its current cost value to compute its next action. We prove convergence of the proposed algorithm to a Nash equilibrium in the game leveraging established results on stochastic processes. The performance of the algorithm is analyzed with a numerical case study.

Keywords: Multi-agent decision making, game theory, payoff-based algorithm

1. INTRODUCTION

Decision-making in multi-agent systems arises in applications ranging from electricity market to communication and transportation networks (Arslan et al., 2007; Saad et al., 2012; Scutari et al., 2006). Game theory provides a powerful framework for formulating optimisation problems corresponding to competing or collaborative multi-agent systems. The various notions of equilibria in games characterise desirable and stable solutions to multi-agent optimisation problems. The focus of our paper is on distributed computation of Nash equilibria for a class of multi-agent decision making modeled by non-cooperative games.

There is a large body of work on computation of Nash equilibria in multi-agent games. The approaches differ by the particular structure of agents' cost functions as well as information available to each agent. In a *potential game*, a central optimization problem can be formulated whose minimizers correspond to Nash equilibria of the game. One can then use distributed algorithms for computing the minima of the potential function (Li and Marden, 2013; Salehisadaghiani and Pavel, 2014) to converge to Nash equilibria. Distributed algorithms are also proposed for *aggregative games* (Jensen, 2010; Paccagnan et al., 2016). In general, for implementation of these distributed algorithms communication is needed between individual agents or between each agent and a central coordinator.

Alternative to distributed optimization approaches, learning approaches to computing Nash equilibria proceed by sampling agents' actions from a set of probability distributions. These probability distributions are updated based on the information available in the system. Most of the past work has focused on algorithms that require knowledge of the structure of the cost functions. For example, (Perkins et al., 2015; Marden et al., 2009; Tatarenko, 2014,

2016b) have dealt with learning procedures requiring the so-called oracle information, where each agent can calculate its current cost given any action from its action set.

There are many practical situations in which agents do not know functional form of the objectives. Rather, each agent can only observe their obtained payoffs and be aware of their local actions. In this case, the information structure is referred to as *payoff-based*. A payoff-based learning in potential games is proposed in (Marden and Shamma, 2012) with the guarantee of stochastic stability of potential function minimizers, which coincide with Nash equilibria in potential games. 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 (Goto et al., 2012) and (Zhu and Martínez, 2013). These learning procedures also guarantee stochastic stability of potential function minimizers. Moreover, by tuning a time-dependent parameter the learning procedures converge to a distribution over potential function minimizers in total variation.

All aforementioned payoff-based procedures are applicable to games with finite action space. In our recent work, we proposed a payoff-based approach to learn local optima of potential functions in potential continuous action games without memory (Tatarenko, 2016a). In that work, we addressed the case in which agents' actions live in $\mathbb{R}$.

Our contributions in this paper are as follows. We develop a payoff-based approach for computing Nash equilibria in a class of games with pseudo-monotone maps. In contrast to past work, we consider action spaces being compact subsets of a multidimensional Euclidean space. Given the constraints on action sets in a non-potential game setting, the previously proposed learning methods are no longer applicable. Thus, we develop a sampling based approach, in which the probability distributions from which agents sample their actions are Gaussian. Motivated by projec-

tion based algorithms, the mean of the distribution is updated iteratively by each agent based only on its current payoff and projected on the constraint set. This has an interpretation of a stochastic projected gradient algorithm. Using established results on sequence of random variables, and utilizing monotonicity of the game map, we prove that through appropriate choices of step size, the actions converge in probability to Nash equilibria.

This paper is organized as follows. In Section 2 we formulate the game under consideration. In Section 3 we present our payoff-based approach and prove its convergence to a Nash equilibrium. In Section 4 we present a numerical case study, motivated by games arising in electricity markets. In Section 5 we summarize the results.

Notations and basic definitions: The set $\{1, \ldots, N\}$ is denoted by $[N]$. Boldface is used to distinguish between the vectors in a multi-dimensional space and scalars. Given N vectors $\boldsymbol{x}^i \in \mathbb{R}^d$, $i \in [N]$, $[\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N] := [\boldsymbol{x}^i]_{i=1}^N := [\boldsymbol{x}^{1\top}, \ldots, \boldsymbol{x}^{N\top}]^\top$; $\boldsymbol{x}^{-i} := [\boldsymbol{x}^1, \ldots, \boldsymbol{x}^{i-1}, \boldsymbol{x}^{i+1}, \ldots, \boldsymbol{x}^N]$. 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 $\|\boldsymbol{x}\| := \sqrt{(\boldsymbol{x}, \boldsymbol{x})}$. 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 $\boldsymbol{x}^\top A \boldsymbol{x} \geq (>)0$ for all $\boldsymbol{x} \neq 0$. $A \otimes B$ denotes the Kronecker product. Given a function $g(\boldsymbol{x}, \boldsymbol{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$, we define the mapping $\nabla_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{y}) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ component wise as $[\nabla_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{y})]_i := \frac{\partial g(\boldsymbol{x}, \boldsymbol{y})}{\partial x^i}$. We will use the big-$O$ notation. Namely, the function $f(x) : \mathbb{R} \to \mathbb{R}$ is $O(g(x))$ as $x \to a$, $f(x) = O(g(x))$ as $x \to a$, if $\lim_{x \to a} \frac{|f(x)|}{|g(x)|} \leq K$ for some positive K.

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

Definition 2. The mapping $\boldsymbol{M} : \mathbb{R}^d \to \mathbb{R}^d$ is called *Lipschitz* with a constant L on $X \subseteq \mathbb{R}^d$, if $\|\boldsymbol{M}(\boldsymbol{x}) - \boldsymbol{M}(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|$ for any $\boldsymbol{x}, \boldsymbol{y} \in X$.

2. PROBLEM FORMULATION

2.1 Learning Nash Equilibria in Games

We are focused here on a game $\Gamma(N, \{A_i\}, \{J_i\})$ with N players, the sets of players' actions $A_i \subset \mathbb{R}^d$, $i \in [N]$, and the cost functions $J_i : \boldsymbol{A} \to \mathbb{R}$, where $\boldsymbol{A} = A_1 \times \ldots \times A_N$ is the set of joint actions [1]. We make the following assumptions regarding the game Γ.

Assumption 1. The game under consideration is *convex*. Namely, for all $i \in [N]$ the set 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 $\boldsymbol{a}^{-i}$.

Assumption 2. The mapping $\boldsymbol{M} : \mathbb{R}^{Nd} \to \mathbb{R}^{Nd}$, referred to as the game mapping, defined by

$$\boldsymbol{M}(\boldsymbol{a}) = [\nabla_{\boldsymbol{a}^i} J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i})]_{i=1}^N$$
$$= [M_{1,1}, \ldots, M_{1,d}, \ldots, M_{N,1}, \ldots, M_{N,d}]^\top,$$
$$M_{i,k}(\boldsymbol{a}) = \frac{\partial J_i(\boldsymbol{a})}{\partial a_k^i} \qquad \boldsymbol{a} \in \boldsymbol{A}, \quad i \in [N], \quad k \in [d] \quad (1)$$

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

is *Lipschitz on* $\mathbb{R}^{Nd}$ with a constant L and *pseudo-monotone on* $\boldsymbol{A}$ (see Definition 1, Definition 2).

Assumption 3. The cost functions $J_i(\boldsymbol{a})$, $i \in [N]$, grow not faster than a polynomial function as $\|\boldsymbol{a}\| \to \infty$.

A Nash equilibrium in a game $\Gamma(N, \{A_i\}, \{J_i\})$ represents a joint action from which no player has any incentive to unilaterally deviate.

Definition 3. A point $\boldsymbol{a}^* \in \boldsymbol{A}$ is called a *Nash equilibrium* if for any $i \in [N]$ and $\boldsymbol{a}^i \in A_i$
$$J_i(\boldsymbol{a}^{i*}, \boldsymbol{a}^{-i*}) \leq J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i*}).$$

In this paper, we focus on learning such a stable state in a game. We are interested in designing a *payoff-based algorithm*, which converges to a Nash equilibrium in any game for which Assumptions 1-3 hold.

2.2 Convex Games and Variational Inequalities

In this subsection, we prove existence of Nash equilibria for $\Gamma(N, \{A_i\}, \{J_i\})$, through connecting Nash equilibria and solutions of Variational Inequalities.

Definition 4. Consider a mapping $\boldsymbol{T}(\cdot)$: $\mathbb{R}^d \to \mathbb{R}^d$ and a set $Y \subseteq \mathbb{R}^d$. A *solution $SOL(Y, \boldsymbol{T})$ to the variational inequality problem $VI(Y, \boldsymbol{T})$* is a set of vectors $\boldsymbol{y}^* \in Y$ such that $(\boldsymbol{T}(\boldsymbol{y}^*), \boldsymbol{y} - \boldsymbol{y}^*) \geq 0$, for any $\boldsymbol{y} \in Y$.

The following theorem is the well-known result on the existence of $SOL(Y, \boldsymbol{T})$, see Corollary 2.2.5 in (Pang and Facchinei, 2003).

Theorem 1. 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.

Next, we formulate the result that establishes the connection between Nash equilibria in a game and solution vectors of a certain Variational Inequality, Proposition 1.4.2 in (Pang and Facchinei, 2003).

Theorem 2. Given a game $\Gamma(N, \{A_i\}, \{J_i\})$, suppose that the action sets $\{A_i\}$ are closed and convex, the cost functions $\{J_i\}$ are continuously differentiable in $\boldsymbol{a}$ and convex in $\boldsymbol{a}^i$ for every fixed $\boldsymbol{a}^{-i}$ on the interior of $\boldsymbol{A}$. Then, some vector $\boldsymbol{a}^* \in \boldsymbol{A}$ is a Nash equilibrium in Γ, if and only if $\boldsymbol{a}^* \in SOL(\boldsymbol{A}, \boldsymbol{M})$, where $\boldsymbol{M}$ is the game mapping in (1).

Note that games for which Assumptions 1 and 2 hold, satisfy all conditions in Theorem 2. Thus, any solution of $VI(\boldsymbol{A}, \boldsymbol{M})$ is also a Nash equilibrium in such games and vice versa. Moreover, according to Theorem 1, Assumption 1 guarantees non-emptiness of $SOL(\boldsymbol{A}, \boldsymbol{M})$. Formally, we get the following result.

Corollary 3. Let $\Gamma(N, \{A_i\}, \{J_i\})$ be a game for which Assumptions 1 and 2 hold. Then, there exists at least one Nash equilibrium in Γ. Moreover, any Nash equilibrium in Γ belongs to the set $SOL(\boldsymbol{A}, \boldsymbol{M})$, where $\boldsymbol{M}$ is the game mapping (see (1)).

Note that Assumptions 1 and 2 do not imply uniqueness of the Nash equilibrium in $\Gamma(N, \{A_i\}, \{J_i\})$. To guarantee uniqueness, one needs to consider a more restrictive assumption, for example, strong monotonicity of the game mapping (Pang and Facchinei, 2003). In our paper we do not restrict our attention to such case, but deal with a broader class of games admitting multiple Nash equilibria.

3. SOLUTION APPROACH

3.1 Payoff-Based Algorithm

In this subsection we formulate the payoff-based approach for the distributed learning of a Nash equilibrium $\boldsymbol{a}^*$ in a game $\Gamma(N, \{A_i\}, \{J_i\})$ satisfying Assumptions 1-2.

Having access to information about the current state $\mathbf{x}^i(t) = [x_1^i, \ldots, x_d^i]^\top \in \mathbb{R}^d$ at iteration t and the current cost value $\hat{J}_i(t)$ at the joint state $\mathbf{x}(t)$, $\hat{J}_i(t) = J_i(\mathbf{x}^1(t), \ldots, \mathbf{x}^N(t))$, each agent "mixes" its next state $\mathbf{x}^i(t+1)$, namely it chooses its next state $\mathbf{x}^i(t+1)$ randomly according to the multidimensional normal distribution $\mathcal{N}(\boldsymbol{\mu}^i(t) = [\mu_1^i(t), \ldots, \mu_d^i(t)]^\top, \sigma(t))$ with the density:

$$p_i(x_1^i, \ldots, x_d^i; \boldsymbol{\mu}^i(t+1), \sigma(t+1))$$
$$= \frac{1}{(\sqrt{2\pi}\sigma(t+1))^d} \exp\left\{ -\sum_{k=1}^d \frac{(x_k^i - \mu_k^i(t+1))^2}{2\sigma^2(t+1)} \right\}.$$

Our choice of Gaussian distribution is based on the idea of CALA (continuous action-set learning automaton), presented in the literature on learning automata (Thathachar and Sastry, 2003). The mean parameter $\boldsymbol{\mu}^i(t)$ for the state's distribution is updated as follows:

$$\boldsymbol{\mu}^i(t+1) =$$
$$\text{Proj}_{A_i}\left[\boldsymbol{\mu}^i(t) - \gamma(t+1)\sigma^2(t+1)\hat{J}_i(t)\frac{\mathbf{x}^i(t) - \boldsymbol{\mu}^i(t)}{\sigma^2(t)} \right].$$

In the above, $\text{Proj}_C[\cdot]$ denotes the projection operator on set C. The initial finite value of $\boldsymbol{\mu}(0)$ can be defined arbitrarily. We emphasize the difference between states and actions. In particular, states are intermediary values $\mathbf{x}(t) = [\mathbf{x}^1(t), \ldots, \mathbf{x}^N(t)]$ updated during the payoff-based algorithm under consideration. They need not belong to the set of joint actions $\boldsymbol{A}$. We will show that upon convergence of the algorithm, the states will also belong to the joint action set.

Our goal is now to analyze convergence of the proposed algorithm. First, we show that this algorithm is analogous to the Robbins-Monro stochastic approximation procedure (Bharath and Borkar, 1999). Next, we show the convergence of the random vector $\boldsymbol{\mu}(t) = [\boldsymbol{\mu}^1(t), \ldots, \boldsymbol{\mu}^N(t)]$ using existing results on stochastic processes and by properly choosing $\{\sigma(t), \gamma(t)\}_{t=0}^\infty$.

It is straightforward to show that under Assumption 1

$$E_{\mathbf{x}(t)}\{\hat{J}_i(t)\frac{x_k^i(t) - \mu_k^i(t)}{\sigma^2(t)}\}$$
$$= E\{\hat{J}_i(t)\frac{x_k^i(t) - \mu_k^i(t)}{\sigma^2(t)}|x_k^i(t) \sim \mathcal{N}(\mu_k^i(t), \sigma(t))\}$$
$$= \frac{\partial \tilde{J}_i(\boldsymbol{\mu}^1(t), \ldots, \boldsymbol{\mu}^N(t), \sigma(t))}{\partial \mu_k^i} \quad (2)$$

for any $i \in [N]$, $k \in [d]$, where

$$\tilde{J}_i(\boldsymbol{\mu}^1, \ldots, \boldsymbol{\mu}^N, \sigma) = \int_{\mathbb{R}^{Nd}} J_i(\boldsymbol{x})p(\boldsymbol{\mu}, \boldsymbol{x})d\boldsymbol{x},$$
$$p(\boldsymbol{\mu}, \boldsymbol{x}) = \prod_{j=1}^N p_j(x_1^j, \ldots, x_d^j; \boldsymbol{\mu}^j, \sigma).$$

Note that $\tilde{J}_i$ can be interpreted as the ith player's cost function in mixed strategies, given that the mixed strategies are multivariate normal distributions $\{\mathcal{N}(\boldsymbol{\mu}^i, \sigma)\}_i$.

We can rewrite the algorithm in the following vector form:
$$\boldsymbol{\mu}(t+1) = \text{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}(t) - \gamma(t+1)\sigma^2(t+1)$$
$$\times (\boldsymbol{M}(\boldsymbol{\mu}(t)) + \boldsymbol{Q}(\boldsymbol{\mu}(t), \sigma(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t), \mathbf{x}(t), \sigma(t)))], (3)$$
where
$$\boldsymbol{Q}(\boldsymbol{\mu}(t), \sigma(t)) = \tilde{\boldsymbol{M}}(\boldsymbol{\mu}(t)) - \boldsymbol{M}(\boldsymbol{\mu}(t)),$$
$$\boldsymbol{R}(\mathbf{x}(t), \boldsymbol{\mu}(t), \sigma(t)) = \boldsymbol{F}(\mathbf{x}(t), \boldsymbol{\mu}(t), \sigma(t)) - \tilde{\boldsymbol{M}}(\boldsymbol{\mu}(t)),$$
$$\boldsymbol{F}(\mathbf{x}(t), \boldsymbol{\mu}(t), \sigma(t))$$
$$= [\hat{J}_1(t)\frac{\mathbf{x}^1(t) - \boldsymbol{\mu}^1(t)}{\sigma^2(t)}, \ldots, \hat{J}_N(t)\frac{\mathbf{x}^N(t) - \boldsymbol{\mu}^N(t)}{\sigma^2(t)}],$$

and
$$\tilde{\boldsymbol{M}}(\cdot) = [\tilde{M}_{1,1}(\cdot), \ldots, \tilde{M}_{1,d}(\cdot), \ldots, \tilde{M}_{N,1}(\cdot), \ldots, \tilde{M}_{N,d}(\cdot)]^\top$$
is the Nd-dimensional vector, where $i \in [N]$, $k \in [d]$, and
$$\tilde{M}_{i,k}(\boldsymbol{\mu}(t)) = \frac{\partial \tilde{J}_i(\boldsymbol{\mu}^1(t), \ldots, \boldsymbol{\mu}^N(t), \sigma(t))}{\partial \mu_k^i}.$$

The algorithm above falls under the framework of Robbins-Monro stochastic approximations procedures (Bharath and Borkar, 1999). Indeed, the vector $\boldsymbol{M}(\boldsymbol{\mu}(t))$ corresponds to the gradient term in stochastic approximations procedures, $\boldsymbol{Q}(\boldsymbol{\mu}(t), \sigma(t))$ is a disturbance of the gradient term, whereas $\{\boldsymbol{R}(\mathbf{x}(t), \boldsymbol{\mu}(t), \sigma(t))\}_t$, according to (2), is a martingale difference. To prove convergence of the algorithm, we will use the following theorem that is a well-known result of Robbins and Siegmund on non-negative random variables, see, for example, Lemma 10 in (Poljak, 1987).

Theorem 4. 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, β_t, ξ_t, and ζ_t be non-negative F_t-measurable random variables such that
$$E(z_{t+1}|F_t) \leq z_t(1 + \beta_t) + \xi_t - \zeta_t.$$
Then almost surely $\lim_{t\to\infty} z_t$ exists and is finite. Moreover, $\sum_{t=1}^\infty \zeta_t < \infty$ almost surely on $\{\sum_{t=1}^\infty \beta_t < \infty, \sum_{t=1}^\infty \xi_t < \infty\}$.

Now, we are ready to formulate the main result of this section.

Theorem 5. Let players in a game $\Gamma(N, \{A_i\}, \{J_i\})$ update their states $\{\mathbf{x}^i(t)\}$ at time t according to the normal distribution $\mathcal{N}(\boldsymbol{\mu}^i(t), \sigma(t))$, where the mean parameters are updated as in (3). Let Assumptions 1-3 hold and the variance parameter $\sigma(t)$ and the step-size parameter $\gamma(t)$ be chosen such that $\sum_{t=0}^\infty \gamma(t)\sigma^2(t) = \infty$, $\sum_{t=0}^\infty \gamma(t)\sigma^3(t) < \infty$, and $\sum_{t=0}^\infty \gamma^2(t) < \infty$. Then, as $t \to \infty$, the mean vector $\boldsymbol{\mu}(t)$ converges almost surely to a Nash equilibrium $\boldsymbol{\mu}^* \in \boldsymbol{A}$ of the game Γ, given any initial mean vector $\boldsymbol{\mu}(0)$, and the joint state $\mathbf{x}(t)$ converges in probability to $\boldsymbol{a}^* = \boldsymbol{\mu}^*$.

The theorem above claims almost sure convergence of the sequence of the mean vectors $\{\boldsymbol{\mu}(t)\}$ and weak convergence of the sequence of the agents' states $\{\mathbf{x}(t)\}$ to a Nash equilibrium in the game under consideration.

Remark 6. Recall that we distinguish between the states $\{\mathbf{x}^i\}_{i\in[N]}$ and actions $\{\boldsymbol{a}^i\}_{i\in[N]}$ of players in games. During the run of the algorithm, players choose their states $\{\mathbf{x}^i\}_{i\in[N]}$ according to the normal distributions

$\{\mathcal{N}(\boldsymbol{\mu}^i,\sigma)\}_{i\in[N]}$ and have access to the current value $\{\hat{J}_i(t)\}_{i\in[N]}$ of their cost functions, given the actual joint state: $\hat{J}_i(t) = J_i(\mathbf{x}^1(t),\ldots,\mathbf{x}^N(t))$. However, feasibility of the mean vectors $\{\boldsymbol{\mu}^i(t)\}_{i\in[N]}$ in the proposed procedure justifies the following choice for the actions: $\boldsymbol{a}^i(t) = \boldsymbol{\mu}^i(t)$ for all $i \in [N]$. Thus, under such setting and according to Theorem 5, the players' joint action $\boldsymbol{a}(t) = [\boldsymbol{a}^1(t),\ldots,\boldsymbol{a}^N(t)]$ in long run of the payoff-based algorithm converges to a Nash equilibrium almost surely.

These convergences take place under an appropriate choice of the parameters $\gamma(t)$ and $\sigma(t)$. Note that, analogously to optimization methods based on the gradient descent iterations, the condition $\sum_{t=0}^\infty \gamma(t)\sigma^2(t) = \infty$ guarantees sufficient energy for the time-step parameter $\gamma(t)\sigma^2(t)$ to let the algorithm (3) get to a neighborhood of a desired stationary point, whereas the condition $\sum_{t=0}^\infty \gamma^2(t) < \infty$ does not allow the iteration under the projection operator to be unbounded as time goes to infinity.

3.2 Proof of Main Result (Theorem 5)

Our approach is as follows. Firstly, we estimate the distance between the mean vector $\boldsymbol{\mu}(t+1)$ in the run of the algorithm and some other $\boldsymbol{\mu} \in \boldsymbol{A}$ by this distance on the previous step, namely by $\|\boldsymbol{\mu}(t)-\boldsymbol{\mu}\|$. After that, we analyse each term in this estimation to demonstrate applicability of Theorem 4 to the sequence $\{\boldsymbol{\mu}(t)\}_t$. Finally, we use the properties of Nash equilibria in games satisfying Assumptions 1-3 (see Corollary 3) to demonstrate that the almost sure limit of the sequence $\{\boldsymbol{\mu}(t)\}_t$ is a Nash equilibrium in the game under consideration.

Let $\beta(t) = \gamma(t)\sigma^2(t)$. Then[2]
$$\boldsymbol{\mu}(t+1) = \text{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}(t) - \beta(t+1) \\ \times (\boldsymbol{M}(\boldsymbol{\mu}(t)) + \boldsymbol{Q}(\boldsymbol{\mu}(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t)))]. \quad (4)$$

Let $\boldsymbol{\mu} \in \boldsymbol{A}$ be any point from the joint action set of the game Γ. Then, taking into account the iterative procedure for the update of $\boldsymbol{\mu}(t)$ above and the non-expansion property of the projection operator on a convex set, we get

$$\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}\|^2 \\ = \|\text{Proj}_{\boldsymbol{A}}[\boldsymbol{\mu}(t) - \beta(t+1) \\ \qquad \times (\boldsymbol{M}(\boldsymbol{\mu}(t)) + \boldsymbol{Q}(\boldsymbol{\mu}(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t)))] - \boldsymbol{\mu}\|^2 \\ \leq \|\boldsymbol{\mu}(t) - \beta(t+1) \\ \qquad \times (\boldsymbol{M}(\boldsymbol{\mu}(t)) + \boldsymbol{Q}(\boldsymbol{\mu}(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t))) - \boldsymbol{\mu}\|^2 \\ = \|\boldsymbol{\mu}(t) - \boldsymbol{\mu}\|^2 - 2\beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) \\ \qquad - 2\beta(t+1)(\boldsymbol{Q}(\boldsymbol{\mu}(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) \\ \qquad + \beta^2(t+1)\|g(\boldsymbol{\mu}(t))\|^2, \quad (5)$$

where $g(\boldsymbol{\mu}(t)) = \boldsymbol{M}(\boldsymbol{\mu}(t)) + \boldsymbol{Q}(\boldsymbol{\mu}(t)) + \boldsymbol{R}(\boldsymbol{\mu}(t))$.

Let $\mathcal{F}_T$ be the σ-algebra generated by the random variables $\{\boldsymbol{\mu}(k), k \leq T\}$. By taking the conditional expectation with respect to $\mathcal{F}_T$ of the both sides in the inequality above, we obtain that for any $T > 0$, almost surely

---
[2] We omit further the argument $\sigma(t)$ in terms $\boldsymbol{Q}$ and $\boldsymbol{R}$ for the sake of notation simplicity.

$$2\sum_{t=0}^T \beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) \\ \leq \|\boldsymbol{\mu}(0) - \boldsymbol{\mu}\|^2 - \text{E}\{\|\boldsymbol{\mu}(T+1) - \boldsymbol{\mu}\|^2|\mathcal{F}_T\} \\ + 2\sum_{t=0}^T \beta(t+1)\|\boldsymbol{Q}(\boldsymbol{\mu}(t))\|\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}\| \\ + \sum_{t=0}^T \beta^2(t+1)\text{E}_{\mathbf{x}(t)}\|g(\boldsymbol{\mu}(t))\|^2. \quad (6)$$

In inequality (6) we used the property of the conditional expectation, namely $\text{E}\{\boldsymbol{\mu}(t_1)|\mathcal{F}_{t_2}\} = \boldsymbol{\mu}(t_1)$ almost surely for any $t_1 \leq t_2$, as well as the fact that $\text{E}\{\boldsymbol{R}(\boldsymbol{\mu}(t))|\mathcal{F}_T\} = 0$ for all $t \leq T$, which is implied by (2).

According to Assumption 3, we can show that
$$\tilde{M}_{i,k}(\boldsymbol{\mu}) = \frac{1}{\sigma^2}\int_{\mathbb{R}^{Nd}} J_i(\boldsymbol{x})(x_k^i - \mu_k^i)p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x} \\ = \int_{\mathbb{R}^{Nd}} \frac{\partial J_i(\boldsymbol{x})}{\partial x_k^i}p(\boldsymbol{\mu},\boldsymbol{x})d\boldsymbol{x}. \quad (7)$$

Thus,
$$\tilde{\boldsymbol{M}}(\boldsymbol{\mu}(t)) = \int_{\mathbb{R}^{Nd}} M(\boldsymbol{x})p(\boldsymbol{\mu}(t),\boldsymbol{x})d\boldsymbol{x}. \quad (8)$$

Since $\boldsymbol{Q}(\boldsymbol{\mu}(t)) = \tilde{\boldsymbol{M}}(\boldsymbol{\mu}(t)) - \boldsymbol{M}(\boldsymbol{\mu}(t))$ and due to Assumption 2 and equation (8), we can write the following:

$$\|\boldsymbol{Q}(\boldsymbol{\mu}(t))\| \leq \int_{\mathbb{R}^{Nd}} \|\boldsymbol{M}(\boldsymbol{\mu}(t)) - \boldsymbol{M}(\boldsymbol{x})\|p(\boldsymbol{\mu}(t),\boldsymbol{x})d\boldsymbol{x} \\ \leq \int_{\mathbb{R}^{Nd}} L\|\boldsymbol{\mu}(t) - \boldsymbol{x}\|p(\boldsymbol{\mu}(t),\boldsymbol{x})d\boldsymbol{x} \\ \leq \int_{\mathbb{R}^{Nd}} L\left(\sum_{i=1}^N \sum_{k=1}^d |\mu_k^i(t) - x_k^i|\right) p(\boldsymbol{\mu}(t),\boldsymbol{x})d\boldsymbol{x} \\ = O(\sigma(t)), \quad (9)$$

where L is the Lipschitz constant defined in Assumption 2. The last equality in (9) is due to the fact that the first central absolute moment of a random variable with a normal distribution $\mathcal{N}(\mu,\sigma)$ is $O(\sigma)$.

Obviously, $\|\boldsymbol{\mu}(t)-\boldsymbol{\mu}\|$ is bounded for any t, since $\boldsymbol{\mu}(t) \in \boldsymbol{A}$ for any t and $\boldsymbol{\mu} \in \boldsymbol{A}$. Now we proceed with estimating the terms $\text{E}_{\mathbf{x}(t)}\|g(\boldsymbol{\mu}(t))\|^2$ in (6):

$$\text{E}_{\mathbf{x}(t)}\|g(\boldsymbol{\mu}(t))\|^2 \leq \|\boldsymbol{M}(\boldsymbol{\mu}(t))\|^2 + \|\boldsymbol{Q}(\boldsymbol{\mu}(t))\|^2 \\ + 2\|\boldsymbol{M}(\boldsymbol{\mu}(t))\|\|\boldsymbol{Q}(\boldsymbol{\mu}(t))\| \\ + \text{E}_{\mathbf{x}(t)}\|\boldsymbol{R}(\boldsymbol{\mu}(t))\|^2. \quad (10)$$

Note that
$$\text{E}_{\mathbf{x}(t)}\|\boldsymbol{R}(\boldsymbol{\mu}(t))\|^2 \\ \leq \sum_{i=1}^N \sum_{k=1}^d \int_{\mathbb{R}^{Nd}} J_i^2(\boldsymbol{x})\frac{(x_k^i - \mu_k^i(t))^2}{\sigma^4(t)}p(\boldsymbol{\mu}(t),\boldsymbol{x})d\boldsymbol{x}.$$

Thus, we can use Assumption 3 to conclude that
$$\text{E}_{\mathbf{x}(t)}\|\boldsymbol{R}(\boldsymbol{\mu}(t))\|^2 \leq \frac{1}{\sigma^4(t)}f(\boldsymbol{\mu}(t),\sigma(t)),$$

where $f(\boldsymbol{\mu}(t),\sigma(t))$ is a polynomial of $\boldsymbol{\mu}(t)$ and $\sigma(t)$. Hence, taking into account boundedness of $\boldsymbol{\mu}(t)$ for all t, we conclude that
$$\beta^2(t+1)\text{E}_{\mathbf{x}(t)}\|\boldsymbol{R}(\boldsymbol{\mu}(t))\|^2 \leq k_1\gamma^2(t) \quad (11)$$

for some constant k_1. Moreover, according to boundedness of $\boldsymbol{\mu}(t)$ for all t, we can conclude that the first term on the right hand side of (10) is bounded.

Bringing (9) - (11) together and taking into account conditions on the parameters $\gamma(t)$, $\sigma(t)$, we conclude that the right hand side of inequality (6) stays finite almost surely, if $T \to \infty$ and, thus, almost surely

$$\sum_{t=0}^{\infty} \beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) < \infty. \qquad (12)$$

Next, we demonstrate that almost surely

$$\varliminf_{t \to \infty} (\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) \leq 0. \qquad (13)$$

Indeed, let us assume that, on the contrary, there exists such $\epsilon > 0$ and $t_0 > 0$ that almost surely

$$(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) \geq \epsilon$$

for any $t \geq t_0$. In this case, taking into account that $\sum_{t=0}^{\infty} \beta(t+1) = \infty$, we obtain

$$\sum_{t=0}^{\infty} \beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu})$$
$$\geq \sum_{t=0}^{t_0} \beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}) + \epsilon \sum_{t=t_0}^{\infty} \beta(t+1) = \infty$$

almost surely, which contradicts (12). Thus, (13) holds. Since $\boldsymbol{\mu}(t)$ is bounded for any t, there exists such a limit point $\boldsymbol{\mu}^* \in \boldsymbol{A}$ that $\varlimsup_{t \to \infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*$ and, according to (13),

$$(\boldsymbol{M}(\boldsymbol{\mu}^*), \boldsymbol{\mu} - \boldsymbol{\mu}^*) \geq 0. \qquad (14)$$

Since we did not specify the choice of $\boldsymbol{\mu} \in \boldsymbol{A}$, the inequality above holds for any $\boldsymbol{\mu} \in \boldsymbol{A}$. Thus, according to Corollary 3, $\boldsymbol{\mu}^*$ is a Nash equilibrium in the game Γ.

Next, we notice that, if $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ in (5), this inequality (5) together with (9) - (11) imply that

$$\mathrm{E}\{\|\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}^*\|^2 | \mathcal{F}_t\} \leq \|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\|^2$$
$$- 2\beta(t+1)(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*)$$
$$+ h(t), \qquad (15)$$

where $h(t) = k_2 \beta(t+1)\sigma(t) + k_3 \beta^2(t+1) + k_4 \beta^2(t+1)\sigma(t) + (\beta(t+1)\sigma(t))^2 + k_1 \gamma^2(t)$. According to the properties of $\sigma(t)$ and $\gamma(t)$,

$$\sum_{t=0}^{\infty} h(t) < \infty.$$

Moreover, since M is pseudo-monotone, (14) implies $(\boldsymbol{M}(\boldsymbol{\mu}(t)), \boldsymbol{\mu}(t) - \boldsymbol{\mu}^*) \geq 0$ for any t. Thus, we can apply Theorem 4 to conclude that

$$\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}^*\| \text{ converges almost surely as } t \to \infty.$$

Since $\varlimsup_{t \to \infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^*$ almost surely,

$$\lim_{t \to \infty} \boldsymbol{\mu}(t) = \boldsymbol{\mu}^* \text{ almost surely.}$$

Since $\sum_{t=0}^{\infty} \gamma(t)\sigma^2(t) = \infty$ and $\sum_{t=0}^{\infty} \gamma(t)\sigma^3(t) < \infty$, $\lim_{t \to \infty} \sigma(t) = 0$. Taking into account that $\mathbf{x}(t) \sim \mathcal{N}(\boldsymbol{\mu}(t), \sigma(t))$, we conclude that $\mathbf{x}(t)$ converges weakly to a Nash equilibrium $\boldsymbol{a}^* = \boldsymbol{\mu}^* \in \boldsymbol{A}^*$ as time runs. Moreover, according to Portmanteau Lemma (Klenke, 2008), this convergence is also in probability.

4. NUMERICAL CASE STUDY

We illustrate the proposed approach of payoff-based learning to a game arising from electricity market. The agents optimize their power consumption in response to a price signal. The price signal in turn depends on the total consumption of the agents. In contrast to most past work, we consider the case in which the form of the price function is unknown to agents and the agents do not share information via communication with each other. They can only observe their cost function for every strategy they play. A similar information structure was considered in (Gan et al., 2013). However, the problem was decentralized computation of the optimizers of the centralized social welfare objective, rather than a game theoretic formulation.

4.1 Problem setup

The problem setup is motivated by the game theoretic formulation of plug-in-electric vehicle (PEV) charging considered in several previous work including (Ma et al., 2010; Grammatico* et al., To Appear Nov. 2016; Couillet et al., 2012). The aforementioned problem can be abstracted as a game as follows. There are N market participants (users), also referred to as players or agents. Let $\boldsymbol{a}^i = [a_1^i, \ldots, a_d^i]^\top \in \mathbb{R}^d$ denote the decision variable of the player i, $i \in [N]$, that is the vector corresponding to its consumption profile over d periods. The constraints for each player i are

$$0 \leq a_k^i \leq \bar{a} \quad \text{for } k = 1, \ldots, d,$$
$$\sum_{k=1}^{d} a_k^i = \bar{a}^i. \qquad (16)$$

These constraints indicate that for each player the electricity consumption at each time instance is limited and the total electricity consumption over the considered period of time needs to match a desired amount. The convex and compact set defined by the constraints in (16) is considered the action set A_i for the corresponding player i.

The cost function is the price paid for electricity consumption by each agent (Paccagnan et al., 2016)

$$J_i(\boldsymbol{a}^i, \boldsymbol{a}^{-i}) = \boldsymbol{a}^{i\top} Q^i \boldsymbol{a}^i + 2 \left(C^i \frac{1}{N} \sum_{j=1}^{N} \boldsymbol{a}^j + \mathbf{c}^i \right)^\top \boldsymbol{a}^i \quad (17)$$

with $Q^i, C^i \in \mathbb{R}^{d \times d}$, $\mathbf{c}^i \in \mathbb{R}^d$ for all $i \in [N]$. In the above, the first term presents each agent's private value, while the second term corresponds to price of electricity and its functional form may not be known to the agents.

Consider the following setup. At iteration t, each player submits its proposed consumption profile over time horizon of d units, $\mathbf{x}^i(t) = [x_1^i(t), \ldots, x_d^i(t)]^\top$. How should players update their profiles, using only values of the function $J_i(\mathbf{x})$, in order to make the sequence of the joint profiles convergent to a Nash equilibrium? Note that Q^i can in general be known by individual agents, while the second term (17) is assumed unknown. Furthermore, Q^i can also equal zero.

For the above game, the game mapping $\boldsymbol{M}(\boldsymbol{a})$ (see (1)) is:

$$\boldsymbol{M}(\boldsymbol{a}) = \hat{M}\boldsymbol{a} + \boldsymbol{m},$$

$$\hat{M} := \mathrm{diag}\left(Q^i + Q^{i\top} + \frac{C^i}{N}\right) + \begin{bmatrix} C^1 \\ C^2 \\ \vdots \\ C^N \end{bmatrix} (\mathbb{1}_N^\top \otimes I_d)$$

$$\boldsymbol{m} := [\mathbf{c}^1,\ \mathbf{c}^2,\ \ldots,\ \mathbf{c}^N].$$

We assume the matrices Q^i and C^i to be such that $Q^i + \frac{C^i}{N} \succeq 0$ on $\mathbb{R}^{Nd}$ and $\hat{M} \succeq 0$ on $\boldsymbol{A}$. These conditions imply the fulfillment of Assumption 1. Obviously, the mapping $\hat{M}\boldsymbol{a} + \boldsymbol{m}$ is affine on $\mathbb{R}^{Nd}$ and, hence, Lipschitz on $\mathbb{R}^{Nd}$. Moreover, the positive semidefinite matrix $\hat{M}$ on $\boldsymbol{A}$ implies that $\hat{M}\boldsymbol{a} + \boldsymbol{m}$ is pseudo-monotone on $\mathbb{R}^{Nd}$ (Gowda, 1990). Thus, under such setting for the matrices Q^i and C^i, $i \in [N]$, Assumptions 1-3 hold in the game under consideration.

4.2 Simulations

Let us assume that there are N users, whose strategies are their consumption profiles for $d = 4$ periods, the matrices Q^i and C^i, $i \in [N]$, in their cost functions (17) are the identity matrices of the size 4×4, and the vector $\mathbf{c}^i$, $i \in [N]$, is a 4-dimensional vector, whose coordinates are some random variables taking values in the interval $(0,5)$. We assume that the action set A_i for each user $i \in [N]$ is defined by (16), where $\bar{a} = 6$ and $\bar{a}_i$ is a random variable taking values in the interval $(0.5, 10)$. The initial mean vector $\boldsymbol{\mu}(0)$ is a random vector with the uniform distribution on $\boldsymbol{A} = A_1 \times \ldots \times A_N$. Let the agents follow the payoff-based algorithm described by (3).

Figures 1-2 present behavior of the relative error $\frac{\|\boldsymbol{\mu}(t) - \boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$ during the algorithm's run with different initial values for $\boldsymbol{\mu}(0)$, $\gamma(t) = \frac{1}{t^{0.51}}$, $\sigma(t) = \frac{0.1}{t^{0.2}}$, $N = 10$ and $N = 100$ respectively, where $\boldsymbol{a}^*$ is the unique Nash equilibrium of the corresponding game. The uniqueness of the Nash equilibrium is due to the fact that the game mapping is strongly monotone in this example (Pang and Facchinei, 2003). We can see that already after the first iteration the algorithm gives an approximation for the Nash equilibrium in the game, irrespectively to the initial vector $\boldsymbol{\mu}(0)$. In the case of $N = 10$ the relative error is $27\% - 31\%$ (see the blue lines on Fig. 1, Fig. 2) after the first iteration. In the case of $N = 100$, this error drops, however, from 98% and 88% only to 62% and 51% respectively (see the red lines on Fig. 1, Fig. 2). At the next iterations the value of the vector $\boldsymbol{\mu}(t)$ approaches $\boldsymbol{a}^*$ slowly. After 10 iterations the relative error decreases to 26% and 28%, given $N = 10$ (see the blue lines on Fig. 1, Fig. 2), and to 56% and 45%, given $N = 100$ (see the red lines on Fig. 1, Fig. 2).

From these figures, we observe that convergence of the error to zero is slow. The blue line on Figure 2 demonstrates that the mean vector $\boldsymbol{\mu}(t)$ is within 85% (the relative error is 15%) of the Nash equilibrium in the game with 10 players after 100 iterations, whereas this vector is within 59% (see the red line on Fig. 2) in the case $N = 100$ after 100 iterations. 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 $\gamma(t)$ as well as by the projection step. These factor prevent a significant change of the projected mean vectors and of the states' values chosen according to the normal distribution with the variance $\sigma(t)$. Hence, the relation between the settings for the algorithm's parameter, number of agents, and the convergence rate of the procedure should be analyzed in the future.

5. CONCLUSION

This paper presented a new payoff-based algorithm for learning Nash equilibria in games with pseudo-monotone maps. To investigate the convergence properties of the proposed procedure we used the theory of discrete-time stochastic processes. We proved that in the run of the algorithm the joint actions in the game under consideration converge weakly and in probability to a Nash equilibrium. This payoff-based approach is demonstrated to be applicable to games between users at electrical markets, where the functional form of the electricity price may be unknown to users. Our current and future work focuses on estimation of convergence rate of the algorithm and improvement of convergence rate by adjustment of the algorithm's parameters.
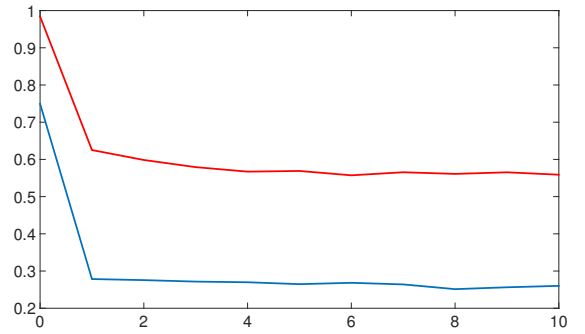

Fig. 1. Relative error $\frac{\|\boldsymbol{\mu}(t) - \boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$ during the payoff-based algorithm, $N = 10$ (blue line), $N = 100$ (red line), $\gamma(t) = \frac{1}{t^{0.51}}$, $\sigma(t) = \frac{0.1}{t^{0.2}}$.
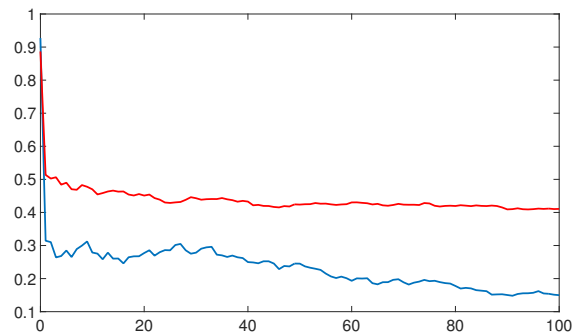

Fig. 2. Relative error $\frac{\|\boldsymbol{\mu}(t) - \boldsymbol{a}^*\|}{\|\boldsymbol{a}^*\|}$ during the payoff-based algorithm, $N = 10$ (blue line), $N = 100$ (red line), $\gamma(t) = \frac{1}{t^{0.51}}$, $\sigma(t) = \frac{0.1}{t^{0.2}}$.

REFERENCES

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

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

Couillet, R., Perlaza, S.M., Tembine, H., and Debbah, M. (2012). A mean field game analysis of electric vehicles in the smart grid. In *Computer Communications Workshops (INFOCOM WKSHPS)*, 79–84. IEEE.

Gan, L., Topcu, U., and Low, S.H. (2013). Optimal decentralized protocol for electric vehicle charging. *IEEE Transactions on Power Systems*, 28(2), 940–951.

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

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

Grammatico*, S., Parise*, F., Colombino*, M., and Lygeros, J. (To Appear Nov. 2016). Decentralized convergence to Nash equilibria in constrained deterministic mean field control. accepted to *IEEE Transactions on Automatic Control*. URL http://arxiv.org/pdf/1410.4421.pdf.

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

Klenke, A. (2008). *Probability theory: a comprehensive course*. Springer, London.

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

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

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

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

Paccagnan, D., Kamgarpour, M., and Lygeros, J. (2016). On Aggregative and Mean Field Games with Applications to Electricity Markets. In *European Control Conference*.

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

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

Poljak, B.T. (1987). *Introduction to optimization*. Optimization Software.

Saad, W., Zhu, H., Poor, H.V., and Basar, T. (2012). 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.

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

Scutari, G., Barbarossa, S., and Palomar, D.P. (2006). 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, 241–244.

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

Tatarenko, T. (2016a). Stochastic payoff-based learning in multi-agent systems modeled by means of potential games. In *55th IEEE Conference on Decision and Control*. accepted.

Tatarenko, T. (2016b). Stochastic stability of potential function maximizers in continuous version of independent log-linear learning. In *European Control Conference (ECC), 2016*.

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

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