Aspiration Learning in Coordination Games

Georgios C. Chasparis Jeff S. Shamma Ari Arapostathis

Abstract—We consider the problem of distributed convergence to efficient outcomes in coordination games through payoff-based learning dynamics, namely aspiration learning. The proposed learning scheme assumes that players reinforce well performed actions, by successively playing these actions, otherwise they randomize among alternative actions. Our first contribution is the characterization of the asymptotic behavior of the induced Markov chain of the iterated process by an equivalent finite-state Markov chain, which simplifies previously introduced analysis on aspiration learning. We then characterize explicitly the behavior of the proposed aspiration learning in a generalized version of so-called coordination games, an example of which is network formation games. In particular, we show that in coordination games the expected percentage of time that the efficient action profile is played can become arbitrarily large.

I. INTRODUCTION

Distributed coordination is of particular interest in many engineering systems. A few examples include distributed overlay routing [1] and topology control [2] in wireless communications. In either case, nodes need to utilize their resources efficiently so that a global objective is achieved. In these scenarios, expected utility maximization (or best reply) by each agent is not often feasible in terms of information or formulation of beliefs about other agents’ actions.

To this end, this paper is concerned about a form of distributed learning dynamics, namely aspiration learning, where agents “satisfice” rather than “optimize”. The aspiration learning scheme is based on a simple principle of “win-stay, lose-shift” [3], according to which a successful action is repeated while an unsuccessful action is dropped. The success of an action is usually determined by a simple comparison test of its prior performance with the player’s desirable return (aspiration level). The aspiration level is updated to incorporate prior experience into the agent’s success criterion.

The history of aspiration learning schemes starts with the pioneering work of [4], where satisfaction-seeking behavior used to explain social decision making. A simple aspiration learning model is presented in [3], where games of two players and two actions are considered, decisions are taken based on the “win-stay, lose-shift” rule and the success of a selected action is based on a simple comparison between the payoff received and an aspiration level.

The difficulty of the aspiration learning schemes lies in the analysis of their convergence behavior. To simplify the analysis several schemes consider an aspiration update which is slower than the action update [5]. In the case of synchronous aspiration and action update, [6] and [7] show that the Pareto-dominant state is selected with probability close to one on mutual interest games and symmetric coordination games of two players and two actions, respectively. Similar are the results in [8], [9]. However, contrary to [6] and [7], both models incorporate a small perturbation either in the aspiration update [8] or in the action update [9]. An aspiration learning scheme in games of multiple players and actions is also presented in [10], where aspiration level is updated similarly to the update rule of [3]. It is shown that a Nash equilibrium will be selected with probability arbitrarily close to one in weakly acyclic games.

In this paper, we are going to focus on games of large number of players and actions, where players apply an aspiration learning scheme similar to the one introduced in [8]. Our goal here is to characterize explicitly the asymptotic behavior of the learning process, and derive conditions under which “efficient” outcomes are selected. Our first contribution is the characterization of the asymptotic behavior of the induced Markov chain by an equivalent finite-state Markov chain, which extends prior analysis to games of multiple players and actions. We also characterize the asymptotic behavior in a class of games which is a generalized version of so-called coordination games. It will be shown that, in these games, the expected percentage of time that the learning process spends at the payoff-dominant profiles can become arbitrarily large when the experimentation probability becomes sufficiently small. We finally demonstrate the applicability of the learning scheme to a network formation game.

The remainder of the paper is organized as follows. Section II defines coordination games and presents an example in network formation. Section III presents the aspiration learning algorithm and its convergence properties in games of multiple players and actions. Section IV specializes the convergence results in coordination games. The results are illustrated through simulations in a network formation game. Finally, Section V presents concluding remarks.

Terminology: We consider the standard setup of finite strategic form games. There is a finite set of agents, \( I = \{1, 2, \ldots, n\} \), and each agent has a finite number of actions, denoted by \( A_i \). The set of action profiles is the cartesian

This work was supported by ONR project #N00014-09-1-0751 and AFOSR project #FA9550-09-1-0538.

G.C. Chasparis is a post-doctoral fellow in the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332. E-mail: chasparis@gatech.edu. URL: http://users.ece.gatech.edu/~gechas.

J.S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332. E-mail: shamma@gatech.edu. URL: http://www.prism.gatech.edu/~jshamma3.

A. Arapostathis is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712. ar@ece.utexas.edu. http://www.ece.utexas.edu/~ari. A. Arapostathis was supported in part by the Office of Naval Research through the Electric Ship Research and Development Consortium.
product $A \triangleq A_1 \times \ldots \times A_n$; $\alpha_i \in A_i$ denotes an action of agent $i$; and $\alpha = (\alpha_1, \ldots, \alpha_n) \in A$ denotes an action profile of all agents. The payoff/utility function of player $i$ is a mapping $u_i : A \to \mathbb{R}$. An action profile $\alpha^* \in A$ is a (pure) Nash equilibrium if

$$u_i(\alpha^*_i, \alpha^*_{-i}) \geq u_i(\alpha'_i, \alpha^*_{-i})$$

for all $i \in I$ and $\alpha'_i \in A_i$, where $-i$ denotes the complementary set $I \setminus \{i\}$. We will denote the set of pure Nash equilibria by $A^*$. Finally, define the set of payoff-dominant action profiles, denoted by $\overline{A}$, as follows:

$$\overline{A} \triangleq \{\alpha \in A : \alpha \in \arg \max_{\alpha \in A} u_i(\alpha), \forall i \in I\}.$$  

II. COORDINATION GAMES

A. Definition

Before defining coordination games, we first need to define the notion of better reply:

Definition 2.1 (Better reply): The better reply of agent $i \in I$ to an action profile $\alpha = (\alpha_i, \alpha_{-i}) \in A$ is a set valued map $\text{BR}_i : A \to 2^A$ such that for any $\alpha_i^* \in \text{BR}_i(\alpha)$ we have $u_i(\alpha^*_i, \alpha_{-i}) > u_i(\alpha'_i, \alpha_{-i})$.

A strict coordination game is defined as follows:

Definition 2.2 (Strict coordination game): A game of two or more agents is a strict coordination game if $\overline{A} \neq \emptyset$ and the following conditions are satisfied:

1) for any $\overline{\alpha} \in \overline{A}$ and $\alpha \notin \overline{A}$, $u_i(\overline{\alpha}) > u_i(\alpha), \forall i \in I$;
2) for any $\alpha \in A \backslash A^*$, there exist $i \in I$ such that $\text{BR}_i(\alpha) \neq \emptyset$, and action $\alpha'_i \in \text{BR}_i(\alpha)$ such that $u_j(\alpha'_i, \alpha_{-i}) > u_j(\alpha_i, \alpha_{-i})$ for all $j \neq i$;
3) for any $\alpha^* \in A^*$ and agent $j_1 \in I$, there exists a sequence of distinct agents $j_1, \ldots, j_n \in I$, and actions $\tilde{\alpha}_j \in A_{j}$, $j = j_1, \ldots, j_{n-1}$, such that

$$u_i(\tilde{\alpha}_{j_1}, \ldots, \tilde{\alpha}_{j_{n-1}}, \alpha^*_n) < u_i(\alpha^*_i),$$

for all $i \in \{j_1, j_2, \ldots, j_{n+1}\}$, $\ell = 1, 2, \ldots, n-1$.

We also define a (non-strict) coordination game as a game which satisfies all conditions of Definition 2.2 but Condition 1.

The conditions of a coordination game establish a weak form of “coincidence of interests” and define a larger class of games than the games traditionally considered as coordination games, e.g., [11], [12]. For example, Condition 2 implies that there exist an agent $i$ and a better reply for $i$ which keeps everyone satisfied. Also, Condition 3 implies that, for any Nash equilibrium profile, there exists a sequence of agents and actions that can reduce everyone’s payoff below their equilibrium payoff.

A trivial example of a strict coordination game is the Stag-Hunt Game of Table I. First, there exists a payoff-dominant profile that satisfies Condition 1 of Definition 2.2. Also, from any non-equilibrium profile, there is a better reply that improves the payoff for all agents (i.e., Condition 2 holds). Finally, for either one of the Nash equilibrium profiles, $(A, A)$ or $(B, B)$, there is a player (row or column) and an action which makes everyone’s payoff less than the equilibrium payoff (i.e., Condition 3 holds).

It is straightforward to show that:

Claim 2.1: In a coordination game, for any action profile $\alpha \notin A^*$ there exists a sequence of action profiles $\{\alpha(k)\}$, such that $\alpha(0) = \alpha$ and $\alpha_i(k) \in \text{BR}_i(\alpha(k-1))$ for some $i$, terminates at a Nash equilibrium.

Note that a direct consequence of Claim 2.1 is that coordination games are weakly acyclic games (cf., [13]). Another straightforward observation is that:

Claim 2.2: In any coordination game $\overline{A} \subseteq A^*$.

In other words, the set of payoff-dominant action profiles is a subset of the set of Nash equilibria.

B. Network formation example

To illustrate the tractability of coordination games, we consider a simple network formation game of $n$ nodes and assume that the sets of actions of each agent, $A_i$, contains all possible combinations of neighbors of $i$, denoted $\mathcal{N}_i$, with which a link can be established, i.e., $A_i = 2^{\mathcal{N}_i}$. Links are considered unidirectional, and a link established by node $i$ with node $s$, denoted $(s, i)$, starts at $s$ with the arrowhead pointing to $i$. A graph $G$ will then be defined as a collection of nodes and directed links. Define also the path from $s$ to $i$ as a sequence of nodes and directed links that starts at $s$ and ends to $i$ following the orientation of the graph, i.e.,

$$(s \rightarrow i) = \{s = s_0, (s_0, s_1), s_1, \ldots, (s_{m-1}, s_m), s_m = i\}$$

for some positive integer $m$. In a connected graph, there is a path from any node to any other node.

Let us consider the utility function $u_i : A \to \mathbb{R}, i \in I$, as:

$$u_i(\alpha) \triangleq \sum_{s \in I \setminus \{i\}} \chi_\alpha(s \rightarrow i) - \kappa_0 |\alpha_i|,$$

where $|\alpha_i|$ denotes the number of links corresponding to action $\alpha_i$ and $\kappa_0$ is a constant in $(0, 1)$. Also,

$$\chi_\alpha(s \rightarrow i) \triangleq \begin{cases} 1 & \text{if } (s \rightarrow i) \subseteq G_\alpha \\ 0 & \text{otherwise} \end{cases}$$

where $G_\alpha$ denotes the graph induced by joint action $\alpha$. The resulting Nash equilibria are usually called Nash networks [14]. As it was shown in Proposition 4.2 in [15], any Nash network $G$ is critically connected, i.e., $i$ it is connected, and ii) for any $(s, i) \in G$, $(s, i)$ is the unique path from $s$ to $i$. For example, the resulting Nash networks for $n = 3$ agents and unconstrained neighborhoods are shown in Fig. 1.

Note that payoff-dominant networks (if exist) are connected with minimum number of links. Also, not all Nash networks are payoff-dominant, e.g., the network of Fig. 1(a) payoff-dominates the network of Fig. 1(b).
The utility function (2) corresponds to the connections model of [16] and has been used to describe various economic and social contexts such as transmission of information. It has also been applied for distributed topology control in wireless networks [2].

It is straightforward to check that:

Claim 2.3: The network formation game defined by (2) is a (non-strict) coordination game if the set of payoff-dominant networks is non-empty.

The condition that payoff-dominant networks exist is not restrictive. For example, if $N_i = \mathcal{I}\setminus\{i\}$ for all $i$, then there exists a unique payoff-dominant network, namely the wheel network, cf. [15].

III. ASPIRATION LEARNING

In this section we define aspiration learning, based on [8]. For some constants $\zeta > 0$, $\epsilon > 0$, $\lambda \geq 0$, $c > 0$, $0 < h < 1$, and $\rho, \bar{\rho} \in \mathbb{R}$, such that

$$-\infty < \rho < \frac{1}{\alpha_i}, \epsilon \leq \max_{\alpha_i \in A_i} u_i(\alpha) < \bar{\rho} < \infty,$$

the aspiration learning iteration initialized at $(\alpha(0), \rho(0))$ is described by Table II. According to this algorithm, each agent $i$ keeps track of an aspiration level, $\rho_i$, which measures player $i$’s desirable return and is defined as a discounted running average of its payoffs throughout the history of play. Given the current aspiration level, $\rho_i(t)$, agent $i$ selects a new action $\alpha_i(t + 1)$. If the previous action, $\alpha_i$, provided utility higher than $\rho_i(t)$, then agent $i$ is “satisfied” and selects the same action, $\alpha_i(t + 1) = \alpha_i$. Otherwise, the new action is selected randomly over all available actions, where the probability of selecting again $\alpha_i$ depends on the level of discontent measured by the difference $u_i(\alpha) - \rho_i(t) < 0$.

The following notation is standard, e.g., [17].

- $B(\mathcal{X})$: the Borel $\sigma$-algebra on $\mathcal{X}$.
- $C(\mathcal{X})$: the Banach space of real-valued continuous functions on $\mathcal{X}$ under the sup-norm (denoted by $\|\cdot\|_\infty$) topology.
- $\mathcal{M}(\mathcal{X})$: the space of finite-signed measures on $\mathcal{X}$ under the topology of weak convergence.

At any instance $t = 0, 1, \ldots$,  
1) Agent $i$ plays $\alpha_i(t) = \alpha_i$ and measures utility $u_i(\alpha)$.  
2) Agent $i$ updates its aspiration level according to  
$$\rho_i(t + 1) = \text{sat}[\rho_i(t) + \epsilon[u_i(\alpha) - \rho_i(t)] + r_i(t)]$$

where  
$$r_i(t) \triangleq \begin{cases} 0, & \text{w.p. } 1 - \lambda; \\ \text{rand}[-\zeta, \zeta], & \text{w.p. } \lambda, \end{cases}$$

and  
$$\text{sat}[\rho] \triangleq \begin{cases} \bar{\rho}, & \rho > \bar{\rho}; \\ \rho, & \rho \in [\underline{\rho}, \bar{\rho}]; \\ \underline{\rho}, & \rho < \underline{\rho}. \end{cases}$$

3) Agent $i$ updates its action:  
$$\alpha_i(t + 1) = \begin{cases} \alpha_i \text{ rand}(A_i \setminus \alpha_i), & \text{w.p. } \phi(u_i(\alpha) - \rho_i) \\ 1 - \phi(u_i(\alpha) - \rho_i), & \text{w.p. } 1 - \phi(u_i(\alpha) - \rho_i). \end{cases}$$

4) Agent $i$ updates the time and repeats.

A. Notation and terminology

Define the state-space as $\mathcal{X} \triangleq A \times [\underline{\rho}, \bar{\rho}]^n$, i.e., pairs of i) joint actions $\alpha$ and ii) vectors of aspiration levels, $\rho_i$, $i \in \mathcal{I}$.

The following notation is standard, e.g., [17].

- $B(\mathcal{X})$: the Borel $\sigma$-algebra on $\mathcal{X}$.
- $C(\mathcal{X})$: the Banach space of real-valued continuous functions on $\mathcal{X}$ under the sup-norm (denoted by $\|\cdot\|_\infty$) topology.
- $\mathcal{M}(\mathcal{X})$: the space of finite-signed measures on $\mathcal{X}$ under the topology of weak convergence.

For $f \in C(\mathcal{X})$ and $\mu \in \mathcal{M}(\mathcal{X})$, define

$$\mu[f] \triangleq \int_{\mathcal{X}} \mu(dx)f(x).$$

Under this notation, $\mu_n$ converges weakly to $\mu_0$, denoted $\mu_n \Rightarrow \mu_0$, if for all $f \in C(\mathcal{X})$, $\mu_n[f] \rightarrow \mu_0[f]$.

$P(\mathcal{X})$: the subset of $\mathcal{M}(\mathcal{X})$ of probability measures on $\mathcal{X}$, i.e., for $\mu \in P(\mathcal{X})$, i) $\mu(\mathcal{X}) = 1$ and ii) $\mu(B) \geq 0$ for all $B \in B(\mathcal{X})$.

$T : \mathcal{X} \times B(\mathcal{X}) \rightarrow [0, 1]$ is a transition probability function if i) $T(\cdot, \cdot) \in P(\mathcal{X})$ for all $x \in \mathcal{X}$ and ii) $T(\cdot, B)$ is measurable on $\mathcal{X}$ for all $B \in B(\mathcal{X})$.

For $f \in C(\mathcal{X})$, and transition probability function, $T$,

$$Tf(x) \triangleq \int_{\mathcal{X}} T(x, dy)f(y).$$

For $\mu \in P(\mathcal{X})$ and transition probability function $T$, $\mu T \in P(\mathcal{X})$ is the probability measure defined by

$$\mu T(B) \triangleq \int_{\mathcal{X}} \mu(dx)T(x, B).$$

$\mu \in P(\mathcal{X})$ is an invariant measure for the transition probability function $T$ if $\mu = \mu T$.

For transition probability functions $T_1$ and $T_2$, the transition probability function $T_1T_2$ is defined by

$$T_1T_2(x, B) \triangleq \int_{\mathcal{X}} T_1(x, dy)T_2(y, B).$$

$\delta_x$: the Dirac measure in $\mathcal{M}(\mathcal{X})$ defined by $x \in \mathcal{X}$,
i.e., for $B \in \mathcal{B}(\mathcal{X})$,
\[
\delta_x(B) \triangleq \begin{cases} 
1, & x \in B; \\
0, & x \notin B.
\end{cases}
\]

- $1_B$: the indicator function of the set $B$,
\[
1_B(x) \triangleq \begin{cases} 
1, & x \in B; \\
0, & x \notin B.
\end{cases}
\]

- $\partial B$: the boundary of the set $B$.

B. Limiting invariant measure and ergodicity

We are interested in the asymptotic behavior of aspiration learning dynamics as the experimentation probability $\lambda$ approaches zero.

Aspiration learning defines an $\mathcal{X}$-valued Markov chain. Let $P_\lambda(x, \cdot)$ denote the corresponding transition probability function. We will refer to this process with $\lambda > 0$ as the perturbed process.

The analysis of the asymptotic behavior of aspiration learning will be related to the following class of states:

**Definition 3.1 (Pure strategy state):** A pure strategy state is a state $x = (\alpha, \rho) \in \mathcal{X}$ such that for all $i \in I, u_i(\alpha) = \rho_i$.

The set of pure strategy states will be denoted by $S$ with cardinality $|S|$. Note also that the set $S$ is isomorphic to $A$ and can be identified as such.

The main result of the paper is summarized in the following theorem:

**Theorem 3.1 (Limiting invariant measure and ergodicity):**

1) For $\lambda > 0$, $P_\lambda$ admits an invariant probability measure, $\mu_\lambda$.

2) There exists a unique vector $\pi = [\pi_s]$, with $\pi_s \geq 0$, $s \in S$, and $\sum_{s \in S} \pi_s = 1$, such that for all $x \in X$, $u_i(\pi) = \rho_i$.

3) For $\lambda > 0$ and for every $f \in L^1(\mu_\lambda)$,
\[
\lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} P^k_\lambda f(x) \to \mu_\lambda[f]
\]

for $\mu_\lambda$-a.e. $x \in \mathcal{X}$. Furthermore, for every $B \in \mathcal{B}(\mathcal{X})$ such that $\partial B \cap S = \emptyset$, we have
\[
\lim_{\lambda \to 0} \lim_{t \to \infty} \frac{1}{t} \sum_{k=0}^{t-1} P^k_\lambda(x, B) = \hat{\mu}(B).
\]

The importance of this theorem lies on the fact that it relates the asymptotic behavior of the perturbed aspiration learning with the unique invariant distribution, $\pi$, of a finite-state Markov chain. The characterization of $\pi$ will also be provided through the proof of the second claim. The third claim also provides the expected time average behavior of aspiration learning as $\lambda \to 0$ and $t \to \infty$ in relation to the limiting invariant measure $\hat{\mu}$ and therefore $\pi$.

C. Proof of Theorem 3.1

The transition probability function $P_\lambda$ satisfies the weak Feller property, i.e., $f \in C(\mathcal{X})$ implies $P_\lambda f \in C(\mathcal{X})$. Furthermore, $\mathcal{X}$ is a compact metric space. Therefore, $P_\lambda$ admits an invariant probability measure, $\mu_\lambda$, e.g., [17, Theorem 7.2.3].

Equation (4) of the third claim is a direct consequence of Birkhoff’s individual ergodic theorem, e.g., [17, Theorem 2.3.4], and the fact that $\mathcal{X}$ is the only invariant set with respect to $P_\lambda$.

Equation (5) is a direct consequence of Portmanteau’s theorem due to the weak convergence of $\mu_\lambda$ to $\hat{\mu}$ and the fact that $S$ is the support of $\hat{\mu}$ as (3) indicates.

It remains to show the second claim of Theorem 3.1. This will require a series of propositions the proofs of which have been omitted for the sake of brevity.

Let $P(x, \cdot)$ denote the transition probability function for $\lambda = 0$. We will refer to this process as the unperturbed process. Let $\Omega \triangleq \mathcal{X}^\omega$ denote the canonical path space with an element $\omega \in \Omega$ being a sequence $\{x(0), x(1), \ldots\}$.

Let $\xi : \Omega \to \mathcal{X}$ be a sequence defined by $\xi(t) \triangleq \{x(\tau) : \tau \leq t\}$.

The importance of this theorem lies on the fact that it relates the asymptotic behavior of the perturbed aspiration learning with the unique invariant distribution, $\pi$, of a finite-state Markov chain. The characterization of $\pi$ will also be provided through the proof of the second claim. The third claim also provides the expected time average behavior of aspiration learning as $\lambda \to 0$ and $t \to \infty$ in relation to the limiting invariant measure $\hat{\mu}$ and therefore $\pi$.
where \( \varphi(\lambda) \in (0, 1), \lambda > 0, \) and \( \lim_{\lambda \to 0} \varphi(\lambda) = 0, \) then
\[
\lim_{\lambda \to 0} \|R_\lambda f - \Pi f\|_\infty = 0, \text{ for all } f \in \mathcal{C}(\chi).
\]
We can express the transition probability function of the perturbed process as
\[
P_\lambda = (1 - \varphi(\lambda)) P + \varphi(\lambda) Q_\lambda,
\]
where \( 0 < \varphi(\lambda) < 1 \) satisfies \( \varphi(\lambda) \to 0 \) as \( \lambda \to 0 \) and equals the probability that at least one agent trembles.

Define the “lifted” transition probability function:
\[
P^L_\lambda = \varphi(\lambda) \sum_{t=0}^{\infty} (1 - \varphi(\lambda))^t Q_\lambda P^t = Q_\lambda R_\lambda.
\]
This transition probability function can be interpreted as an alternative representation of \( P_\lambda \). The evolution of \( P_\lambda \) looks like trembles followed by periods of no trembles. The periods of no trembles have varying length, and so the dynamics involve periods of \( Q_\lambda P \ldots P \). The definition of \( P^L_\lambda \) is a weighted sum of such strings, with the weights being the probability of the associated string.

In a similar manner to \( P_\lambda \), we can write
\[
Q_\lambda = (1 - \psi(\lambda)) Q + \psi(\lambda) Q^*,
\]
where \( Q \) is the transition probability function induced by aspiration learning where exactly one player trembles, and \( Q^* \) is the transition probability function where at least two players tremble simultaneously.

It turns out that \( P^L_\lambda \) from (7) and \( \Pi \) from Corollary 3.1 have a useful relationship for small \( \lambda \).

Proposition 3.4: The following hold,
1) For \( f \in \mathcal{C}(\chi) \), \( \lim_{\lambda \to 0} \|P^L_\lambda f - \Pi f\|_\infty = 0 \).
2) Any invariant distribution, \( \mu_s \), of \( P_\lambda \) is also an invariant distribution of \( P^L_\lambda \).
3) Let \( \mu_{\lambda_0} \) be a sequence of invariant probability measures of \( P_{\lambda_0} \) with \( \lambda_0 \to 0 \). If \( \mu_{\lambda_0} \Rightarrow \hat{\mu} \), then \( \hat{\mu} \) is an invariant probability measure of \( \Pi \).

For some \( s \in \mathcal{S} \) define the sets
\[
N_\varepsilon \triangleq [\alpha_s, \rho_s - \varepsilon, \rho_s + \varepsilon], \quad \varepsilon > 0,
\]
where \( (\alpha_s, \rho_s) \) denote the action and aspiration level of \( s \). For any two pure strategy states, \( s, s' \in \mathcal{S} \), define also
\[
\hat{P}_{ss'} = \lim_{t \to \infty} QP^t(s, N_\varepsilon(s'))
\]
for some \( \varepsilon > 0 \) sufficiently small. Given Proposition 3.2, \( \hat{P}_{ss'} \) is independent of the selection of \( \varepsilon \). It can be interpreted as the probability that under the dynamics \( QPP \ldots \) with initial condition \( s \) the process has been “captured” by \( s' \), i.e., action \( \alpha_{s'} \) is being played repeatedly after some time \( t \). Define also the matrix \( \hat{P} \triangleq [\hat{P}_{ss'}] \).

Proposition 3.5 (Invariant measure of \( \Pi \)): The unique invariant measure, \( \hat{\mu} \), of \( \Pi \) satisfies
\[
\hat{\mu}(s) = \sum_{s \in \mathcal{S}} \pi_s \delta_{s}(-),
\]
for some constants \( \pi_s \geq 0, s \in \mathcal{S} \), such that \( \pi = [\pi_s] \) define an invariant distribution of \( \hat{P} \).

IV. APPLICATION TO COORDINATION GAMES

A. Invariant measure

Let us define \( \mathcal{S} \) and \( \mathcal{S}^* \) to be the set of pure strategy states corresponding to \( \mathcal{A} \) and \( \mathcal{A}^* \), respectively. Let us assume that \( \pi \) is an invariant distribution of the stochastic matrix \( P \).

We are going to characterize the invariant distribution \( \pi \) with respect to the payoff-dominant pure strategy states and for step-size \( \epsilon \) that approaches zero. To this end, we first need to compute the transition probability under the \( QPP \ldots \) dynamics from a) a pure strategy state in \( \mathcal{S} \) to a pure strategy state in \( \mathcal{S}^* \), and b) a pure strategy state in \( \mathcal{S}^* \) to a pure strategy state in \( \mathcal{S} \). These transition probabilities are summarized in the followign two lemmas.

Lemma 4.1: In any strict coordination game and for sufficiently small \( \zeta > 0 \), the probability of a transition under the \( QPP \ldots \) dynamics from any pure strategy state in \( \mathcal{S} \) to any pure strategy state in \( \mathcal{S}^* \) goes to zero as \( \epsilon \) goes to zero.

Lemma 4.2: In any strict coordination game and for sufficiently small \( \epsilon, \zeta > 0 \), the probability of a transition under the \( QPP \ldots \) dynamics from any pure strategy state in \( \mathcal{S}^* \) to any pure strategy state in \( \mathcal{S} \) is bounded away from zero and independent of \( \epsilon \).

Using Lemmas 4.1–4.2 and Claim 2.1, it is straightforward to show the following:

Theorem 4.1: In any strict coordination game, \( \pi_{s_i} = 0 \) for all \( s_i \notin \mathcal{S} \) as \( \epsilon \to 0 \).

Theorem 4.1 combined with Theorem 3.1 provides a complete characterization of the time average asymptotic behavior of aspiration learning in strict coordination games.

B. Simulations in network formation

In this section, we demonstrate the asymptotic behavior of aspiration learning in coordination games as described by Theorems 3.1–4.1. Consider the network formation game of Section II-B which, according to Claim 2.3, is a (non-strict) coordination game. Although Theorem 4.1 was only shown for strict coordination games, our intention here is to demonstrate that it also applies to the larger class of (non-strict) coordination games.

To this end, we consider a set of six nodes deployed on the plane so that the neighbors of each node are the two immediate nodes (e.g., \( N_1 = \{2, 6\} \)). Note that in order for the average behavior to be observed \( \lambda \) and \( \epsilon \) need to be sufficiently small. We choose: \( h = 0.01, c = 0.2, \zeta = 0.01, \epsilon = \lambda = 10e - 4, \) and \( \kappa_0 = 1/8 \). In Fig. 2, we have plotted a typical response of aspiration learning for this setup, where we plotted the final graph and the aspiration level as a function of time. To illustrate better the response, define \( dist_G(j, i) \) as the distance from node \( j \) to node \( i \) as the minimum number of hops from \( j \) to \( i \). We admit the convention \( dist_G(i, i) = 0 \) and \( dist_G(j, i) = \infty \) if there is no path from \( j \) to \( i \) in \( G \). The final graph of Fig. 2 plots, for each node, the running average of the inverse of the sum of its distances from all other nodes, i.e., \( 1/\sum_{j \in \mathcal{N}} dist_G(j, i) \).
the inverse total distance converges to $1/15 = 0.067$, both of which correspond to the wheel network.

V. Conclusions

We introduced an aspiration learning algorithm and analyzed its asymptotic behavior in games of multiple players and actions. The main contribution of this analysis is the establishment of a relation between the time average behavior of the induced Markov chain with the invariant distribution of a finite-state Markov chain. We further characterized this invariant distribution in a large class of games, namely coordination games, and showed that it puts non-zero weight only on the payoff-dominant profiles. Finally, we demonstrated how aspiration learning can be implemented for distributed network formation, which is of independent interest.

References

[1] B. G. Chun, R. Fonseca, I. Stoica, and J. Kubiatowicz, “Characterizing selfishly constructed overlay routing networks,” in Proc. of IEEE INFOCOM 04, Hong-Kong, 2004.
[2] R. Komali, A. B. MacKenzie, and R. P. Gilles, “Effect of selfish node behavior on efficient topology design,” IEEE Transactions on Mobile Computing, vol. 7, no. 9, pp. 1057–1070, 2008.
[3] M. Posch, A. Pichler, and K. Sigmund, “The efficiency of adapting aspiration levels,” Biological Sciences, vol. 266, no. 1427, pp. 1427–1435, July 1999.
[4] H. A. Simon, “A behavioural model of rational choice,” Quarterly Journal of Economics, vol. 69, pp. 99–118, 1955.
[5] J. Bendor, D. Mookherjee, and D. Ray, “Reinforcement learning in repeated interaction games,” Advances in Theoretical Economics, vol. 1, no. 1, 2001.
[6] A. Pazgal, “Satisficing leads to cooperation in mutual interest games,” Int J Game Theory, vol. 26, no. 4, 1997.
[7] Y. Kim, “Satisficing and optimality in $2 \times 2$ common interest games,” Economic Theory, vol. 13, pp. 365–375, 1999.
[8] R. Karandikar, D. Mookherjee, and D. Ray, “Evolving aspirations and cooperation,” Journal of Economic Theory, vol. 80, pp. 292–331, 1998.
[9] I. K. Cho and A. Matsui, “Learning aspiration in repeated games,” Journal of Economic Theory, vol. 124, pp. 171–201, 2005.
[10] J. R. Marden, H. P. Young, G. Arslan, and J. S. Shamma, “Payoff-based dynamics for multi-player weakly acyclic games,” SIAM Journal on Control and Optimization, vol. 48, no. 1, pp. 373–396, 2009.
[11] P. Vanderschraaf, Learning and Coordination. New York, NY: Routledge, 2001.
[12] D. Lewis, Convention: A Philosophical Study. Blackwell Publishing, 2002.
[13] H. P. Young, Strategic Learning and Its Limits. New York, NY: Oxford University Press, 2004.
[14] V. Bala and S. Goyal, “A noncooperative model of network formation,” Econometrica, vol. 68, no. 5, pp. 1181–1229, 2000.
[15] G. Chasparis and J. Shamma, “Efficient network formation by distributed reinforcement,” in IEEE 47th Conference on Decision and Control, Cancun, Mexico, Dec. 2008.
[16] M. Jackson and A. Wolinsky, “A strategic model of social and economic networks,” Journal of Economic Theory, vol. 71, pp. 44–74, 1996.
[17] O. Hernandez-Lerma and J. B. Lasserre, Markov Chains and Invariant Probabilities. Birkhauser Verlag, 2003.