TRACKING CORRELATED EQUILIBRIA IN CLUSTERED MULTI-AGENT NETWORKS VIA ADAPTIVE FILTERING ALGORITHMS

Omid Namvar Gharehshiran and Vikram Krishnamurthy

Department of Electrical and Computer Engineering
University of British Columbia
2332 Main Mall, Vancouver, BC V6T 1Z4

ABSTRACT

We present a decentralized adaptive filtering algorithm in a clustered network of agents. Agents receive payoffs partly due to performing localized tasks in clusters and partly due to strategic interaction with agents outside clusters. Each agent is only aware of the actions of others within its cluster and is oblivious to the actions, or even existence, of agents outside the cluster. We show that the global behavior of the network converges to the set of correlated ε-equilibria if the agents follow the proposed algorithm. Thus simple behavior by individual agents can result in sophisticated global behavior.

Index Terms— Adaptive filtering, correlated equilibrium, differential inclusions, stochastic approximation.

1. INTRODUCTION

Consider a network of agents forming multiple non-overlapping clusters. Each cluster is characterized by a subset of agents which perform a particular localized task and share information of their actions with each other. However, the action profile of each cluster cannot be observed by other clusters and agents outside the cluster. Agents repeatedly take actions to which there corresponds two payoffs: i) local payoff, due to performing tasks allocated to the cluster, ii) global payoff, due to global interaction with agents outside the cluster. Agents continuously update their strategies – via a non-linear adaptive filtering algorithm – to maximize their expected payoff based on the realized payoffs in the past and observations of the action profile of cluster members. The question we pose in this paper is: Given this simple local behavior of individual agents, can the clustered network of agents achieve sophisticated global behavior? Similar problem have been studied in the Economics literature. For seminal works, the interested reader is referred to [1, 2].

Main Results: In [2], the authors consider a network model where no information about the action profiles are disclosed and agents only realize their payoffs once they take action. They propose a regret-based reinforcement learning algorithm whereby agents build statistics of their past experience and infer how their payoff would have improved based *only* on the realized payoffs so far. The model in this paper differs as it incorporates cluster structure where actions are *only* locally shared. That is, some agents have no information about the actions and payoffs of other agents, who in fact may be oblivious to their existence, and other agents form clusters where their actions are revealed to the cluster members. The main result of this paper is that if every agent follows the proposed adaptive filtering algorithm, the global behavior of the network converges to the set of correlated ε-equilibria [3]. In addition, we show (via simulations) that, taking advantage of the excess information shared within clusters, faster convergence to the set of correlated ε-equilibria can be achieved.

Correlated equilibrium is a generalization of the Nash equilibrium and describes a condition of competitive optimality. It is arguably best suited for decentralized adaptive learning in multi-agent systems as coordination among agents is directly taken into account. In fact, the common history of actions (partially observed by the agents) serves as the coordination device. Each agent's payoff is a function of others' action profile whether or not the agent is capable of observing it. Therefore, agents indirectly acquire the coordination signal through the realized payoffs. This coordination leads to potentially higher payoffs than if agents take actions independently (as required by Nash equilibrium).

Context: The motivation for such formulation stems from sensor networks. Consider a multiple target tracking scenario in an unattended ground sensor network [4]. Depending on their locations, sensors form clusters each responsible for tracking a particular target. Sensors receive two payoffs: i) local payoff, based on the importance and accuracy of the information provided about the local phenomena, ii) global payoff, for communicating the collected data to the sink through the channel, which is globally shared amongst all sensors. Consideration of the potential local interaction among sensors leads to a more realistic modeling, hence, more sophisticated design of reconfigurable networked sensors.

2. ADAPTIVE FILTERING ALGORITHM FOR CONSENSUS FORMATION IN ACTIONS

2.1. Multi-agent Network Model

Let $\mathcal{L} = \{1, 2, \ldots, L\}$ denote the set of agents. Each agent l is characterized by a set of actions $\mathcal{A}^l = \{1, \ldots, A^l\}$ and a payoff function $U^l : \mathcal{A} \to \mathbb{R}$, where $\mathcal{A} = \times_{l\in\mathcal{L}}\mathcal{A}^l$ represents the set of L-tuple of joint *action profiles*. A generic element of $\mathcal{A}$ is denoted by $\mathbf{a} = (a^1, \ldots, a^L)$ and, for any agent l, can

be rearranged as $(a^l, \mathbf{a}^{-l})$, where $\mathbf{a}^{-l} \in \times_{l' \neq l} \mathcal{A}^{l'}$.

Agents are partitioned into non-overlapping clusters $\mathcal{C}_k \subset \mathcal{L}$, $k = 1, \ldots, K$. Isolated agents will be formulated as singleton clusters. Time is discrete $n = 1, 2, \ldots$. Each agent l takes an action a_n^l at time instant n and receives a payoff U_n^l. We make the *cluster monitoring* assumption, that is,

$$l, l' \in \mathcal{C}_k \quad \textit{iff} \quad l \text{ knows } a_n^{l'} \text{ and } l' \text{ knows } a_n^l. \tag{1}$$

The payoff function for each agent l is formulated as

$$U^l\left(a^l, \mathbf{a}^{-l}\right) = U_{\text{loc}}^l(a^l, \mathbf{a}^{\mathcal{C}_k^{-l}}) + U_{\text{glob}}^l(a^l, \mathbf{a}^{-\mathcal{C}_k}). \tag{2}$$

Here, $\mathbf{a}^{\mathcal{C}_k^{-l}}$ and $\mathbf{a}^{-\mathcal{C}_k}$ denote the action profile of cluster $\mathcal{C}_k$ (to which agent l belongs) excluding agent l and the action profile of all agents excluding cluster $\mathcal{C}_k$, respectively. In addition, $U_{\text{loc}}^l(a^l, \mathbf{a}^{\mathcal{C}_k^{-l}}) = 0$ if cluster $\mathcal{C}_k$ is singleton, i.e., $\mathcal{C}_k^{-l} = \emptyset$. Agents know $U_{\text{loc}}(\cdot)$ and, knowing $\mathbf{a}_n^{\mathcal{C}_k^{-l}}$ and their chosen action a_n^l, are capable of computing their local stage payoff. Knowing the overall realized payoff U_n^l, each agent can then compute its *realized* global payoff:

$$U_{\text{glob},n}^l = U_n^l - U_{\text{loc}}^l(a_n^l, \mathbf{a}_n^{\mathcal{C}_k^{-l}}), \tag{3}$$

Note that agents cannot compute $U_{\text{glob}}^l(a_n^l, \mathbf{a}_n^{-\mathcal{C}_k})$ as they do not acquire $\mathbf{a}_n^{-\mathcal{C}_k}$.

2.2. Regret-based Adaptive Filtering with Partial Local Information

Each agent l generates two average regret matrices and updates them over time: (i) $\bar{\boldsymbol{\alpha}}_{A^l \times A^l}^l$, which records *average local-regrets*, and (ii) $\bar{\boldsymbol{\beta}}_{A^l \times A^l}^l$, which is an unbiased estimator of the *average global-regrets*. Each element $\bar{\alpha}_n^l(i,j)$, $i, j \in \mathcal{A}^l$, gives the time-average regret (in terms of gains and losses in local payoff values) had the agent selected action j every time it took action i in the past. Formally,

$$\bar{\alpha}_n^l(i,j) = \frac{1}{n} \times \sum_{\tau=1}^{n} \left[U_{\text{loc}}^l\left(j, \mathbf{a}_\tau^{\mathcal{C}_k^{-l}}\right) - U_{\text{loc}}^l\left(a_\tau^l, \mathbf{a}_\tau^{\mathcal{C}_k^{-l}}\right) \right] \mathbb{I}\{a_\tau^l = i\}, \tag{4}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function.

Agents, however, do not observe/receive the action profile of agents outside cluster, hence, are unable to perform the thought experiment to compute $U_{\text{glob}}^l\left(j, \mathbf{a}_\tau^{-\mathcal{C}_k}\right)$ as in (4). This paper adopts the approach in [2] to formulate an unbiased estimator of the average global-regrets. Formally, each element $\bar{\beta}_n^l(i,j)$, $i, j \in \mathcal{A}^l$, will be defined as:

$$\bar{\beta}_n^l(i,j) = \frac{1}{n} \times \sum_{\tau=1}^{n} \frac{\sigma_\tau^l(i)}{\sigma_\tau^l(j)} U_{\text{glob},\tau}^l \mathbb{I}\{a_\tau^l = j\} - U_{\text{glob},\tau}^l \mathbb{I}\{a_\tau^l = i\}, \tag{5}$$

where $\boldsymbol{\sigma}_\tau^l = \left(\sigma_\tau^l(i)\right)_{i \in \mathcal{A}^l}$ denotes the randomized strategy according to which agent l picked action at period τ. Intuitively speaking, the normalization factor $\sigma_\tau^l(i)/\sigma_\tau^l(j)$ in (5) makes the length of periods, when actions i and j have been selected, comparable.

Each agent then combines these measures to update its randomized strategy for the following period. Positive overall-regrets $\alpha_n^l(i,j) + \beta_n^l(i,j)$ imply the opportunity to achieve higher payoffs by switching from action i to action j. The more positive the regret for not choosing an action, the higher is the probability that the agent picks that action.

Define $|x|^+ = \max\{0, x\}$ and let $0 < \delta < 1$. At each period, with probability $1 - \delta$, agent l chooses its consecutive action according to $|\bar{\alpha}_n^l(i,j) + \bar{\beta}_n^l(i,j)|^+$. With the remaining probability δ, it randomizes over the action set $\mathcal{A}^l$ according to a uniform distribution. This can be interpreted as "exploration" which is essential as agents continuously learn their global payoff functions. Exploration forces all actions to be chosen with a minimum frequency, hence, rules out actions being rarely chosen.

The regret-based adaptive filtering algorithm can then be summarized as follows:

Algorithm 1:
0) **Initialization:** Set $0 < \delta < 1$. Initialize $\psi_0^l(i) = 1/|A^l|$, for all $i \in \mathcal{A}^l$, $\bar{\boldsymbol{\alpha}}_0^l = \mathbf{0}_{A^l \times A^l}$ and $\bar{\boldsymbol{\beta}}_0^l = \mathbf{0}_{A^l \times A^l}$.
For $n = 1, 2, \ldots$ repeat the following steps:
1) **Strategy Update and Action Selection:** Select action $a_{n+1}^l = i$ according to the following distribution

$$\sigma_{n+1}^l(i) = (1 - \delta)\, \mu_n^l(i) + \delta / A^l, \tag{6}$$

where $\boldsymbol{\mu}_n^l$ denotes the stationary distribution of the following transition probability matrix

$$\psi_n^l(i) = \begin{cases} \frac{1}{\xi^l} |\bar{\alpha}_n^l\left(a_{n-1}^l, i\right) + \bar{\beta}_n^l\left(a_{n-1}^l, i\right)|^+, & i \neq a_{n-1}^l, \\ 1 - \sum_{\substack{j \in \mathcal{A}^l \\ j \neq i}} \psi_n^l(j), & i = a_{n-1}^l. \end{cases} \tag{7}$$

Here, ξ^l is chosen such that $\xi^l > \sum_{k \in A^l \setminus \{a_{n-1}^l\}} \psi_n^l(k)$.
2) **Local Information Exchange:** Agent l: i) broadcasts a_{n+1}^l to the cluster members, ii) receives actions of cluster members and forms the profile $\mathbf{a}_{n+1}^{\mathcal{C}_k^{-l}}$.
3) **Regret Update:**
Step 1: Local Regret

$$\bar{\alpha}_{n+1}^l(i,j) = \bar{\alpha}_n^l(i,j) + \epsilon_{n+1} \times \left(\left[U_{\text{loc}}^l\left(j, \mathbf{a}_\tau^{\mathcal{C}_k^{-l}}\right) - U_{\text{loc}}^l\left(a_\tau^l, \mathbf{a}_\tau^{\mathcal{C}_k^{-l}}\right) \right] \mathbb{I}\{a_\tau^l = i\} - \bar{\alpha}_n^l(i,j) \right). \tag{8}$$

Step 2: Global Regret

$$\bar{\beta}_{n+1}^l(i,j) = \bar{\beta}_n^l(i,j) + \epsilon_{n+1} \times \left(\left[\frac{\sigma_\tau^l(i)}{\sigma_\tau^l(j)} U_{\text{glob},\tau}^l \mathbb{I}\{a_\tau^l = j\} - U_{\text{glob},\tau}^l \mathbb{I}\{a_\tau^l = i\} \right] - \bar{\beta}_n^l(i,j) \right). \tag{9}$$

Here, the step size is selected as $\epsilon_n = 1/(n+1)$ (in static multi-agent system models) or $\epsilon_n = \bar{\epsilon}$, $0 < \bar{\epsilon} \ll 1$, (in slowly time-varying multi-agent networks).
4) **Recursion:** Set $n \leftarrow n + 1$ and go to Step 1.

Remarks:

1) Interpretation of ξ^l in (7): The normalization factor ξ^l is to guarantee that probability distribution $\boldsymbol{\psi}_n^l$ remains valid. Higher ξ^l lowers the probability of switching actions, hence, can be viewed as an *inertia* parameter [1]. To avoid computing ξ^l in each decision period, one can fix $\xi^l > 2A^l \cdot \max_{a^l,\mathbf{a}^{-l}} |U^l(a^l,\mathbf{a}^{-l})|$. We should emphasize that the rate of convergence is closely related to ξ^l.

2) Slowly Time-variant Multi-agent Network: The multi-agent system model may evolve over time due to: i) changes of agents' incentives (payoff functions), ii) changes in cluster membership, and iii) agents joining/disjoining the network. To keep agents responsive to these changes, a constant step-size $\epsilon_n = \bar{\varepsilon}$ is required in (8) and (9). Algorithm 1 cannot respond to multiple successive changes as agents' strategies are functions of the *time-average* regrets.

3) Better-reply Adaptive Procedure: The strategy in Step 1 reinforces all plausible actions with positive probabilities. Hence, the behavior of the agents is non-fully rational (better-reply strategy) as compared to a sophisticated decision-maker who takes the most plausible action (best-reply) given its limited conception of the outside world.

3. GLOBAL BEHAVIOR AND CONVERGENCE ANALYSIS

Assume each agent employs Algorithm 1 to select action for the next period. The *global behavior* for the network of agents is defines as the (discounted) *empirical frequency* of the joint action profiles:

$$\bar{\mathbf{z}}_n = \begin{cases} \frac{1}{n}\sum_{\tau\le n} \mathbf{e}_{\mathbf{a}_\tau}, & \text{if} \quad \epsilon_n = \frac{1}{n}, \\ \bar{\varepsilon}\sum_{\tau\le n}(1-\bar{\varepsilon})^{n-\tau}\,\mathbf{e}_{\mathbf{a}_\tau}, & \text{if} \quad \epsilon_n = \bar{\varepsilon}. \end{cases} \tag{10}$$

Here, $\mathbf{e}_{\mathbf{a}_\tau}$ denotes the $\prod_{l\in\mathcal{L}}|\mathcal{A}^l|$ dimensional unit vector in the set of Cartesian product of agents' joint actions with the element corresponding to $\mathbf{a}_\tau$ being equal to one. The second line in (10) is a *discounted* version of the first line and will be used in slowly evolving multi-agent systems. Given $\bar{\mathbf{z}}_n$, the total expected payoff accrued by the network can be straightforwardly computed. Note that $\bar{\mathbf{z}}_n$ is a network "diagnostic" and is only used for the convergence analysis of Algorithm 1– it does not need to be computed by the network. That being said, a network controller can monitor $\bar{\mathbf{z}}_n$, e.g., in a sensor network, and use it to adjust the price of sensors, thereby changing the equilibrium set in novel ways.

The main result of this paper is that $\bar{\mathbf{z}}_n$ converges to the set of correlated ε-equilibrium. Before proceeding with the formal statement, we define the set of correlated ε-equilibrium.

Definition 3.1 *Let $\boldsymbol{\pi}$ denote a joint distribution on $\mathcal{A}$, where $\pi(\mathbf{a}) \ge 0$ for all $\mathbf{a}\in\mathcal{A}$ and $\sum_{\mathbf{a}\in\mathcal{A}}\pi(\mathbf{a}) = 1$. The set of* correlated ε-equilibrium*, denoted by $\mathcal{C}_\varepsilon$, is the convex set [3]*

$$\mathcal{C}_\varepsilon = \Big\{\boldsymbol{\pi} : \sum_{\mathbf{a}^{-l}} \pi^l(i,\mathbf{a}^{-l})\times \big[U^l(j,\mathbf{a}^{-l}) - U^l(i,\mathbf{a}^{-l})\big] \le \varepsilon, \quad \forall i,j\in\mathcal{A}^l, \forall l\in\mathcal{L}\Big\}. \tag{11}$$

For $\varepsilon = 0$ in (11), $\mathcal{C}_0$ is called the set of correlated equilibria*.*

In (11), $\pi^l\left(i,\mathbf{a}^{-l}\right)$ denotes the probability of agent l choosing action i and the rest playing $\mathbf{a}^{-l}$. Dividing (11) by $\sum_{\mathbf{a}^{-l}\in\mathcal{A}^{-l}}\pi\left(i,\mathbf{a}^{-l}\right)$, agent l can compute *a posteriori* distribution $\pi\left(\mathbf{a}^{-l}|i\right)$, hence, evaluate an expected payoff for each action $i\in\mathcal{A}^l$. Considering $\varepsilon \ll 1$, Definition 3.1 simply states that no agent is better off by unilaterally deviating the recommended signal $\mathbf{a}$ chosen randomly according to distribution $\boldsymbol{\pi}$. Hence, reaching a correlated equilibrium can be viewed as consensus formation in strategy amongst agents.

We now proceed to state the main theorem of this paper.

Theorem 3.1 *Suppose each agent $l\in\mathcal{L}$ deploys the adaptive filter in Algorithm 1 and updates strategy according to (6), where $\boldsymbol{\mu}_n^l$ represents the stationary distribution of (7), i.e.,*

$$\sum_{j\in\mathcal{A}^l\setminus\{i\}} \mu_n^l(j)\cdot\left|\bar{\alpha}_n^l(j,i)+\bar{\beta}_n^l(j,i)\right|^+ = \mu_n^l(i)\cdot\sum_{j\in\mathcal{A}^l\setminus\{i\}}\left|\bar{\alpha}_n^l(i,j)+\bar{\beta}_n^l(i,j)\right|^+. \tag{12}$$

Then, for each $\varepsilon > 0$, there exists $\hat{\delta}(\varepsilon)$ such that if $\delta < \hat{\delta}(\varepsilon)$ in Algorithm 1, the global behavior $\bar{\mathbf{z}}_n$ converges almost surely (for $\epsilon_n = 1/n$) to the set of correlated ε-equilibria in the following sense:

$$\bar{\mathbf{z}}_n \xrightarrow{\text{a.s.}} \mathcal{C}_\varepsilon \ \text{ as } \ n\to\infty \qquad \text{iff} \qquad d(\bar{\mathbf{z}}_n,\mathcal{C}_e) = \inf_{\mathbf{z}\in\mathcal{C}_e}|\bar{\mathbf{z}}_n - \mathbf{z}| \xrightarrow{\text{a.s.}} 0 \ \text{ as } \ n\to\infty. \tag{13}$$

For constant step-size $\epsilon_n = \varepsilon$, $\bar{\mathbf{z}}_n$ converges weakly to $\mathcal{C}_\varepsilon$.

The above theorem implies that the stochastic process $\bar{\mathbf{z}}_n$ enters and stays in the ε-neighborhood of $\mathcal{C}_\varepsilon$ forever with probability one. In other words, for any $\varepsilon > 0$, there exists $N(\varepsilon) > 0$ with probability one such that for $n > N(\varepsilon)$, one can find a correlated equilibrium $\boldsymbol{\pi}$ at the most ε-distance of $\bar{\mathbf{z}}_n$.

Sketch of the Proof: The proof follows from averaging theory [5] and Lyapunov stability of differential inclusions [6]. The main steps to prove Theorem 3.1 are as follows: 1) Trajectories of the piecewise constant continuous-time interpolations of the stochastic processes $\bar{\boldsymbol{\alpha}}_n^k$ and $\bar{\boldsymbol{\beta}}_n^k$ converges almost surely (for $\epsilon_n = 1/n$) or weakly (for $\epsilon_n = \varepsilon$) to the global attractors of the associated continuous-time system of inter-connected differential inclusions: [See (14), shown at the bottom of the page]. In (14),

$$U_{\text{loc}}^l(a^l,\boldsymbol{\nu}^{\mathcal{C}_k^{-l}}) = \int_{\mathcal{A}^{\mathcal{C}_k^{-l}}} U_{\text{loc}}^l(a^l,\mathbf{a}^{\mathcal{C}_k^{-l}})d\nu^{\mathcal{C}_k^{-l}}(\mathbf{a}^{\mathcal{C}_k^{-l}}), \tag{15}$$

$$\begin{cases} \frac{d\bar{\alpha}^l(i,j)}{dt} \in \left\{\left[U_{\text{loc}}^l\left(j,\boldsymbol{\nu}^{\mathcal{C}_k^{-l}}\right) - U_{\text{loc}}^l\left(i,\boldsymbol{\nu}^{\mathcal{C}_k^{-l}}\right)\right]\sigma^l(i)\,;\boldsymbol{\nu}^{\mathcal{C}_k^{-l}}\in\Delta\mathcal{A}^{\mathcal{C}_k^{-l}}\right\} - \bar{\alpha}^l(i,j), \\ \frac{d\bar{\beta}^l(i,j)}{dt} \in \left\{\left[U_{\text{glob}}^l\left(j,\boldsymbol{\nu}^{-\mathcal{C}_k}\right) - U_{\text{glob}}^l\left(i,\boldsymbol{\nu}^{-\mathcal{C}_k}\right)\right]\sigma^l(i)\,;\boldsymbol{\nu}^{-\mathcal{C}_k}\in\Delta\mathcal{A}^{-\mathcal{C}_k}\right\} - \bar{\beta}^l(i,j), \end{cases} \tag{14}$$

$$U^l_{\text{glob}}(a^l, \boldsymbol{\nu}^{-\mathcal{C}_k}) = \int_{\mathcal{A}^{-\mathcal{C}_k}} U^l_{\text{glob}}(a^l, \mathbf{a}^{-\mathcal{C}_k}) d\nu^{-\mathcal{C}_k}(\mathbf{a}^{-\mathcal{C}_k}). \tag{16}$$

In addition, $\Delta\mathcal{A}^{\mathcal{C}_k^{-l}}$ and $\Delta\mathcal{A}^{-\mathcal{C}_k}$ denote simplexes of probability measures over $\mathcal{A}^{\mathcal{C}_k^{-l}}$ and $\mathcal{A}^{-\mathcal{C}_k}$, respectively. 2) The system of inter-connected differential inclusions (14) is Lyapunov stable and the set of global attractors is characterized by positive combined-regrets $\left|\alpha^l(i,j) + \beta^l(i,j)\right|^+$ being confined to an ε-distance of $\mathbb{R}^-$, for all $i, j \in \mathcal{A}^l$. Hence, if every agent employs the regret-based adaptive filtering (Algorithm 1), $\forall \varepsilon \geq 0$, there exists $\hat{\delta}(\varepsilon) \geq 0$ such that if $\delta \leq \hat{\delta}(\varepsilon)$, for all agents $l \in \mathcal{L}$:

$$\limsup_{n\to\infty} \left|\bar{\alpha}^l_n(i,j) + \bar{\beta}^l_n(i,j)\right|^+ \leq \varepsilon \text{ w.p.1, } \forall i, j \in \mathcal{A}^l. \tag{17}$$

3) The global behavior $\bar{\mathbf{z}}_n$ converges to $\mathcal{C}_\varepsilon$ *if and only if* (17) is satisfied.

4. NUMERICAL EXAMPLE

In this section, we study a small hypothetical multi-agent system comprising three agents $\mathcal{L} = \{l_1, l_2, l_3\}$. Agents l_1 and l_2 are allocated the same task, hence, form cluster $\mathcal{C} = \{l_1, l_2\}$. Agent l_3 forms a singleton cluster, hence, neither observes the action profile of $\mathcal{C}$, nor does it disclose its action information to l_1 and l_2. Agents l_1 and l_2 take action from the same action set $\mathcal{A}^{l_1} = \mathcal{A}^{l_2} = \{a_1, a_2\}$. Agent l_3, due to performing a different task, chooses from a different action set $\mathcal{A}^{l_3} = \{b_1, b_2\}$. Table 1 gives the payoffs to the agents for taking a particular action. Each element (x, y, z) in the table represents the payoff to l_1, l_2 and l_3, respectively, corresponding to the particular action profile taken by agents. The set of correlated equilibrium is singleton (a pure strategy), where probability *one* is assigned to (a_2, a_2, b_1) and *zero* to others.

Here, we set $\epsilon_n = 1/n$ and $\delta = 0.1$. Figure 1 illustrates the global behavior $\bar{\mathbf{z}}_n$ averaged over 50 different runs of Algorithm 1 and compares its performance with the reinforcement learning algorithm in [2]. Note that Theorem 3.1 proves convergence to the set of correlated ε-equilibrium. Therefore, although $\bar{z}_n(a_2, a_2, b_1)$ increases with the number of iterations, it could only reach an ε-distance of one depending on the value of δ in Algorithm 1. Similarly, $\bar{z}_n$ decreases for other combination of action profiles. However, they are always played with small positive frequencies. Figure 1 numerically verifies that Algorithm 1 converges faster to the set of correlated ε-equilibrium. As can be seen, in a certain number of iterations, $\bar{z}_n$ gets closer to the the probability values allocated to joint action profiles in correlated equilibrium.

5. CONCLUSION

we proposed a simple decentralized adaptive filtering algorithm to form consensus in actions in a clustered network of agents. Agents form clusters to perform localized tasks and share their action information with cluster members. It is proved that simple non-fully rational local behavior by individual agents can lead to rational global behavior

Table 1: Agents' Payoff Matrix

Local:

	$l_2 : a_1$	$l_2 : a_2$
$l_1 : a_1$	$(3,5)$	$(2,3)$
$l_1 : a_2$	$(3,3)$	$(5,4)$

Global:

	$l_2 : a_1$	$l_2 : a_2$	$l_2 : a_1$	$l_2 : a_2$
$l_1 : a_1$	$(1,4,1)$	$(2,1,3)$	$(2,2,3)$	$(1,3,1)$
$l_1 : a_2$	$(2,1,3)$	$(3,4,4)$	$(3,3,1)$	$(2,2,3)$
	$l_3 : b_1$		$l_3 : b_2$	

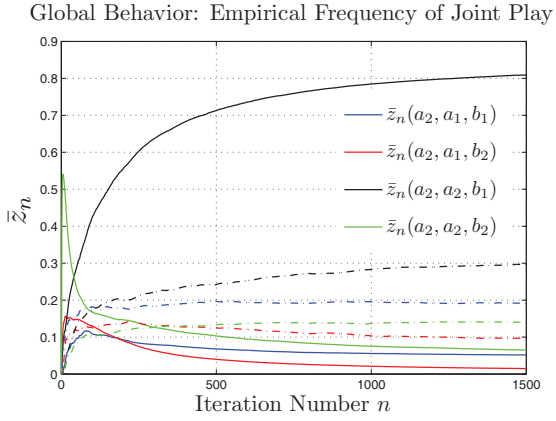

Fig. 1: Global behavior $\bar{\mathbf{z}}_n$: i) Solid lines represent results from Algorithm 1, ii) dashed lines represent results from the reinforcement learning algorithm in [2].

in the network. Agents in clusters can utilize the excess shared/observed information to achieve faster coordination.

6. REFERENCES

[1] S. Hart and A. Mas-Colell, "A simple adaptive procedure leading to correlated equilibrium," *Econometrica*, vol. 68, no. 5, pp. 1127–1150, Sept. 2000.

[2] S. Hart and A. Mas-Colell, "A reinforcement procedure leading to correlated equilibrium," *Economic Essays: A Festschrift for Werner Hildenbrand*, pp. 181–200, 2001.

[3] R. J. Aumann, "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica: Journal of the Econometric Society*, vol. 55, no. 1, pp. 1–18, Jan. 1987.

[4] V. Krishnamurthy, M. Maskery, and G. Yin, "Decentralized adaptive filtering algorithms for sensor activation in an unattended ground sensor network," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6086–6101, Dec. 2008.

[5] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 2nd edition, 2003.

[6] M. Benaïm, J. Hofbauer, and S. Sorin, "Stochastic approximations and differential inclusions; Part II: Applications," *Mathematics of Operations Research*, vol. 31, no. 3, pp. 673–695, Nov. 2006.