Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Learning in Linear Games over Networks

Ceyhun Eksin, Pooya Molavi, Alejandro Ribeiro, and Ali Jadbabaie
Dept. of Electrical and Systems Engineering, University of Pennsylvania
200 South 33rd Street, Philadelphia, PA 19104
Email: {ceksin, pooya, aribeiro, jadbabai}@seas.upenn.edu

Abstract— **We consider a dynamic game over a network with information externalities. Agents' payoffs depend on an unknown true state of the world and actions of everyone else in the network; therefore, the interactions between agents are strategic. Each agent has a private initial piece of information about the underlying state and repeatedly observes actions of her neighbors. We consider strictly concave and supermodular utility functions that exhibit a quadratic form. We analyze the asymptotic behavior of agents' expected utilities in a connected network when it is common knowledge that the agents are myopic and rational. When utility functions are symmetric and adhere to the *diagonal dominance* criterion, each agent believes that the limit strategies of her neighbors yield the same payoff as her own limit strategy. Given a connected network, this yields a consensus in the actions of agents in the limit. We demonstrate our results using examples from technological and social settings.**

I. Introduction

Players behaving strategically in uncertain environments make decisions that are optimal given the information available to them while often trying to learn more about the environment by observing the actions of the others. While the trade decisions of players in a stock market depend on their belief about the true value of the stock, traders also tend to consider how the other traders will act as others' actions could directly affect the gains from trade. When buying certain products, consumers have the incentive to act in coordination with the population while trying to get the best product. In a case of cooperative robotic movement, robots try to rendezvous at a point using only noisy private observations of the coordinates of the target. In all of these environments, interactions are strategic in the sense that agents make decisions by considering the possible choices made by the other agents in addition to trying to estimate the value of an unknown (stock value, product quality, or the location of a goal); furthermore, players oftentimes only observe the actions of a handful of other players while trying to coordinate with and learn from everybody else. In this paper, we investigate the question of whether, in such scenarios, players who act selfishly and myopically would be able to aggregate the dispersed pieces of information and coordinate on the optimal action.

We consider games with strategic complementarities that have a linear best response and where players only have access to local information. Such games could be used to model financial markets [1], consumption [2], cooperative robotics [3], or organizational coordination [4]. The games in this class have a common feature; that is, agents' best replies are linear functions of other agents' play and the true state of the world. Having strategic complementarity among agents' actions implies that an increase in an agent's action provides an incentive for the others to increase their actions respectively, reminiscently of potential games [3]. A network is a natural way to model agents' access to information. The neighborhood of a player could signify the subset of the traders she interacts with in the stock market, the friends who she observes the consumption behavior of, or the robots that are moving in close proximity.

This paper analyzes the asymptotic behavior of agents' actions in the class of repetitive linear games with strategic complementarities given a connected network (Section II). At each stage, agents observe actions of their neighbors from the previous stage, infer about the true state of the world in a rational way, and take an action synchronously with everyone else. Rational inference corresponds to belief updates according to the Bayes' rule. Bayesian learning is often computationally intractable due to the requirement that agents infer about the inferences of other agents that are all interconnected via the network. Consequently, our results are asymptotic (Section III). We show that when agents play according to a Bayesian Nash equilibrium strategy, their expected utilities conditioned on their private observations are equal in the limit (Theorem 1). This implies that eventually agents also reach consensus in the actions they choose (Corollary 1). We provide examples of linear games based on investment choice, multi-robot cooperative movement, and product selection (Section IV). We further provide a numerical example of the multi-robot movement given line and star graph structures (Section V). The numerical results suggest that agents' actions converge in the number steps equal to the diameter of the network and that the eventual actions are optimal given agents' private observations. Finally, we conclude with a summary of our results in Section VI.

This paper is related to two lines of research. The first is the economics literature that focuses on one-shot equilibrium analysis of linear games [4], [5]. While some of these papers characterize the equilibrium based on agents' information structure [4], [5], the others provide a characterization based on network properties [2], [6]. The second line of research relates to the social learning/distributed estimation literature

Research supported in parts by ARO P-57920-NS, NSF CAREER CCF-0952867, NSF CCF-1017454, ONR MURI "Next Generation Network Science", and AFOSR MURI FA9550-10-1-0567.

where a canonical model consists of a set of connected agents exchanging their estimates of an unknown state and using this local information to update their estimates. Such papers can be divided into two categories based on their modeling approach: In rational learning models, agents incorporate the information about the unknown state using the Bayes' rule and are able to discard redundant information [7], [8]. However, due to the intractability of rational learning, the results are asymptotic [8]. In bounded rational models, there is a heuristic update rule that the agents follow [9]–[13]. These rules often help provide a complete characterization of convergence properties based on the network structure and information distribution. While the bounded rational approach provides tangible results, the update rules used are harder to justify. In this paper, we extend the rational learning view of social learning with pure informational externalities to the case of linear games where each agent's payoff is directly affected by the actions of other agents.

II. The Model

We consider linear games with incomplete information, in which identical agents in a network repeatedly choose an action and receive a payoff that depends on their own action, the underlying parameter, and actions of everyone else.

We use an undirected connected graph $G = (V, E)$ with node set V and edge set E to denote the network. Agents are placed at the nodes of the graph belonging to the finite node set $V = 1, \dots, n$. Set of links between agents is denoted by E. Agent i can only observe neighboring agents $n(i) = \{j : (j, i) \in E\}$ that form an edge with it. We use $-i$ to denote the set of agents in V except i; i.e., $-i := \{j : j \neq i, j \in V\}$.

There exists an underlying parameter θ that belongs to the measurable space $\Theta \subseteq \mathbb{R}$. At the beginning of the game, agent i receives a noisy private signal x_i about underlying state coming from some measurable signal set X_i. We use $X := X_1 \times \cdots \times X_n$ to denote the space of all agents' signals. Let $\mathbf{P}$ be agents' common prior over $\Theta \times X$.

At each stage of the game, agents simultaneously take actions. Time t ticks with each stage of the game and hence it is discrete; i.e., $t \in \mathbb{N}$. At each stage t, each player i takes an action $a_{i,t}$ from its compact measurable action space $A_i \subseteq \mathbb{R}$ and observes actions of its neighbors $a_{n(i),t}$. We assume that the agents' action spaces are the same; that is, $A_i = A_j$ for all agent pairs $i \in V$ and $j \in V$. Furthermore, we assume that agents' payoffs could be represented by the following quadratic function

$$u_i(a_i, a_{-i}, \theta) = \alpha a_i - \frac{1}{2} \sum_{j \in V} a_j^2 + \sum_{j \in V \setminus \{i\}} \beta_{ij} a_i a_j + \delta a_i \theta + c\theta^2, \tag{1}$$

where α, β_{ij}, δ and c are real valued constants. Note that the payoff function is strictly concave in agents' self-action; that is, $\partial^2 u_i(\cdot)/\partial a_i^2 < 0$.

In this paper, we consider supermodular games in which agents' strategies are complementary to each other. Strategic complementarity means that the marginal utility of an agent's action increases with an increase in other agents' actions. More formally, the utility function $u_i(a, \theta)$ is supermodular with respect to $a \in A := \bigtimes_{i \in V} A_i$ if for $a_i' \geq a_i$ and $a_{-i}' \geq a_{-i}$, and for any $\theta \in \Theta$,

$$u_i(a_i', a_{-i}', \theta) - u_i(a_i, a_{-i}', \theta) \geq u_i(a_i', a_{-i}, \theta) - u_i(a_i, a_{-i}, \theta) \tag{2}$$

for all $i \in V$. For the utility function in (1), the influence of agent j's action on i's payoff is captured by the cross-derivatives $\partial^2 u_i(\cdot)/\partial a_i \partial a_j = \beta_{ij}$ for all $i \in V$ and $j \in V \setminus \{i\}$. When $\beta_{ij} \geq 0$, the actions of i and j are strategic complements. Hence, we assume that $\beta_{ij} \geq 0$ for all $i \in V$ and $j \in V \setminus \{i\}$.

The past history at stage t is a complete list of the parameter, agents' private signals, and actions of all players in all previous stages; i.e, it lies in $H_t := \Theta \times X \times A^{t-1}$. The space of plays is the measurable space $(\Omega, \mathcal{F})$, where $\Omega := H_\infty = \Theta \times X \times A^{\mathbb{N}}$ and $\mathcal{F}$ is the Borel σ-algebra. We use $h_{i,t}(\omega)$ to denote the history observed by agent i up to time t given that $\omega \in \Omega$ is realized. We let $H_{i,t}$ to be the set of all possible histories observed by agent i up to time t and $\mathcal{H}_{i,t}$ the corresponding σ-algebra over Ω. Let $\mathcal{H}_{i,\infty}$ be the σ-field generated by the union of all $\mathcal{H}_{i,t}$ for all t. This represents agent i's information at the end of the game.

A (pure behavior) strategy σ_i is the sequence of functions $(\sigma_{i,\tau})_{\tau=1,\dots,\infty}$ such that $\sigma_{i,t} : H_{i,t} \mapsto A_i$. We use σ to denote the strategy profile of all agents $\{\sigma_i\}_{i \in V}$.

Any strategy profile σ, together with the common prior $\mathbf{P}$ induces a probability distribution over Ω. We denote the probability distribution induced by σ by $\mathbf{P}_\sigma$, and let $\mathbf{E}_\sigma$ be the corresponding expectation operator. For any σ, let $\mathcal{A}_i$ be the space consisting of $\mathbf{P}_\sigma$-a.e. equivalence classes of (bounded) $\mathcal{H}_{i,\infty}$-measurable functions $f_i : \Omega \mapsto A_i$ with the norm[1]

$$\|f_i\|_{\sigma,2} = \left(\int_\Omega f_i^2 d\mathbf{P}_\sigma \right)^{\frac{1}{2}}. \tag{3}$$

By the Riesz-Fischer theorem, $\mathcal{A}_i$ is complete. We also let $\mathcal{A} := \bigtimes_{i \in V} \mathcal{A}_i$ and $\mathcal{A}_{-i} := \bigtimes_{j \in V \setminus \{i\}} \mathcal{A}_j$. With slight abuse of notation, we use $\|\cdot\|_{\sigma,2}$ to denote the norm on $\mathcal{A}$ defined as

$$\|f\|_{\sigma,2} = \max_{i \in V} \|f_i\|_{\sigma,2}. \tag{4}$$

Note that for any strategy profile σ, since $\sigma_{i,t}(h_{i,t}(\omega))$ is $\mathcal{H}_{i,t}$-measurable and $\mathcal{H}_{i,t} \subset \mathcal{H}_{i,\infty}$, $\sigma_{i,t}(h_{i,t}(\omega)) \in \mathcal{A}_i$ for all $t \in \mathbb{N}$. Whenever there is no risk of confusion, we use $\sigma_{i,t}$ to mean both agent i's strategy at time t and the $\mathcal{H}_{i,t}$-measurable random variable $\sigma_{i,t}(h_{i,t}(\omega))$.

A belief μ is a probability distribution over the space of plays Ω. In particular, we let $\mu_{i,t}(\sigma)$ denote the belief of agent i at time t given strategy profile σ and her information $\mathcal{H}_{i,t}$; that is, $\mu_{i,t}(\sigma)[B]$ is the random variable that satisfies

$$\mu_{i,t}(\sigma)[B] = \mathbf{P}_\sigma(B | \mathcal{H}_{i,t}), \tag{5}$$

for any measurable event $B \in \mathcal{F}$.[2] We let $\mathbf{E}_\mu$ denote the expectation operator over Ω given belief μ.

[1] $\mathcal{A}_i$ is the $L^2(\Omega)$ space of A_i valued functions over the probability triple $(\Omega, \mathcal{H}_{i,\infty}, \mathbf{P}_\sigma)$.

[2] More formally, $\mu_{i,t}(\sigma) : \Omega \times \mathcal{F} \mapsto [0, 1]$ is a regular conditional distribution of $\mathbf{P}_\sigma$ given $\mathcal{H}_{i,t}$.

Given strategy profile σ, the best response of agent i at time t to the strategies of other agents $\mathbf{s}_{-i} \in \mathcal{A}_{-i}$ is a random function $\mathrm{BR}_{i,t}(\sigma) : \mathcal{A}_{-i} \mapsto \mathcal{A}_i$ defined as

$$\begin{aligned}\mathrm{BR}_{i,t}(\sigma; \mathbf{s}_{-i}) &= \underset{a_i \in A_i}{\operatorname{argmax}}\ \mathbf{E}_{\mu_{i,t}(\sigma)}[u_i(a_i, \mathbf{s}_{-i}, \theta)] \\ &= \underset{a_i \in A_i}{\operatorname{argmax}}\ \mathbf{E}_\sigma[u_i(a_i, \mathbf{s}_{-i}, \theta)|\mathcal{H}_{i,t}]. \quad (6)\end{aligned}$$

Note that since $u_i(\cdot)$ is continuous and A_i is compact, the above function is always well-defined. Likewise, define

$$\mathrm{BR}_{i,\infty}(\sigma; \mathbf{s}_{-i}) = \underset{a_i \in A_i}{\operatorname{argmax}}\ \mathbf{E}_\sigma[u_i(a_i, \mathbf{s}_{-i}, \theta)|\mathcal{H}_{i,\infty}] \quad (7)$$

We also use $\mathrm{BR}_t(\sigma) : \mathcal{A} \mapsto \mathcal{A}$ to denote the vector function whose ith component is given by $\mathrm{BR}_{i,t}(\sigma)$.

Assuming an interior solution, the best response function for the payoff in (1) is obtained by taking the derivative with respect to a_i, equating the result to zero and solving for a_i. Agent i's best response to any strategy $\mathbf{s}_{-i}$ is a linear function of strategies of other agents and the underlying parameter; that is,

$$\mathrm{BR}_{i,t}(\sigma; \mathbf{s}_{-i}) = \alpha + \sum_{j \in V \setminus \{i\}} \beta_{ij} \mathbf{E}_\sigma[s_j|\mathcal{H}_{i,t}] + \delta\, \mathbf{E}_\sigma[\theta|\mathcal{H}_{i,t}]. \quad (8)$$

We consider games in which the network structure is common knowledge. It is also common knowledge that agents are rational and myopic. Common knowledge of rational and myopic behavior means that everybody knows that [everybody knows that...] (repeated infinite times) agents are rational and myopic. Rational agents follow Bayes' rule in updating their beliefs over Ω. Myopic behavior implies that agents do not account for the changes in future payoffs when making decisions; rather, they only maximize the expected utility for the current stage of the game. Given that agents are myopic and rational, a Bayesian Nash Equilibrium (henceforth, BNE) is a strategy profile σ that satisfies

$$\sigma_{i,t} = \mathrm{BR}_{i,t}(\sigma; \sigma_{-i,t}) \quad \text{for all} \quad i \in V,\, t \in \mathbb{N}, \quad (9)$$

or equivalently

$$\sigma_t = \mathrm{BR}_t(\sigma; \sigma_t) \quad \text{for all} \ \ t \in \mathbb{N}. \quad (10)$$

A BNE strategy is such that there is no other strategy that agent i could unilaterally deviate to, that will provide a higher payoff; i.e., it is the best that agent i can do given other agents' strategies and her own information.

III. Asymptotic Properties of Rational Learning

According to the setup in Section II, there is a new game played at each stage t based on the new information available. However, since agents accumulate information about the unknown state over time, it is possible to show that beliefs of agents converge asymptotically. Given this fact, under Bayesian-Nash equilibrium concept, agents' expected utilities converge. By the same token, one expects that agents' equilibrium strategies converge as well, as we prove in Lemma 1. Existence of limit strategies implies that the agents can learn their neighbors' limit strategies (Lemma 2). We use these results to prove our main result: the conditional expected utilities are equal for neighboring agents, if agents' utility functions are symmetric (Theorem 1). This result, in turn, implies that agents eventually play the same strategy. Before we state and prove these results, we specify our assumptions.

(A1) Symmetry. Utility functions are symmetric; that is,

$$u_i(a_i, a_j, a_{V\setminus\{i,j\}}, \theta) = u_j(a_i, a_j, a_{V\setminus\{i,j\}}, \theta) \text{ for all } i, j \in V, \quad (11)$$

for any $a_i, a_j \in A_i = A_j$, where the first element of the utility function u_i is always the action of agent i. For the utility function in (1), this is equivalent to requiring that $\beta_{ij} = \beta_{ji}$ for all $i \in V$ and $j \in V \setminus \{i\}$.

(A2) Diagonal dominance. The Hessian matrices of agents' utility functions are strictly diagonally dominant. For the utility function in (1), this is equivalent to requiring that there exists $\rho < 1$ such that

$$\sum_{j \in V\setminus\{i\}} \beta_{ij} \le \rho, \quad (12)$$

for all $i \in V$.

(A1) implies that the utility of agent i, given that i plays a_i and j plays a_j is equal to utility of agent j, given that j plays a_i and i plays a_j.

(A2) implies that an agent's utility is more sensitive to changes in her own actions than to changes in the actions of other agents.

Our first result shows that if the agents play according to BNE strategies at all times, then agents' strategies converge as t goes to infinity.

Lemma 1: Let σ be a BNE. If (A1) and (A2) hold, then $\sigma_t \to \sigma_\infty$ with $\mathbf{P}_\sigma$-probability one as t goes to infinity.

Proof: Let $\mathbf{s}_{i,t} = \mathrm{BR}_{i,t}(\sigma; \sigma_{-i,\infty})$; that is,

$$\mathbf{s}_{i,t} = \alpha + \sum_{j \in V\setminus\{i\}} \beta_{ij} \mathbf{E}_\sigma[\sigma_{j,\infty}|\mathcal{H}_{i,t}] + \delta\, \mathbf{E}_\sigma[\theta|\mathcal{H}_{i,t}]. \quad (13)$$

Taking limit of the above equation as t goes to infinity and using Levy's zero-one law implies that $\mathbf{s}_{i,t} \to \mathbf{s}_{i,\infty}$ with $\mathbf{P}_\sigma$-probability one, where

$$\begin{aligned}\mathbf{s}_{i,\infty} &= \mathrm{BR}_{i,\infty}(\sigma; \sigma_{-i,\infty}) \\ &= \alpha + \sum_{j \in V\setminus\{i\}} \beta_{ij} \mathbf{E}_\sigma[\sigma_{j,\infty}|\mathcal{H}_{i,\infty}] + \delta\, \mathbf{E}_\sigma[\theta|\mathcal{H}_{i,\infty}]. \quad (14)\end{aligned}$$

By the triangle inequality,

$$\|\sigma_t - \sigma_\infty\|_{\sigma,2} \le \|\sigma_t - \mathbf{s}_t\|_{\sigma,2} + \|\mathbf{s}_t - \sigma_\infty\|_{\sigma,2}. \quad (15)$$

Note that since σ is a BNE, $\sigma_{i,t} = \mathrm{BR}_{i,t}(\sigma; \sigma_{-i,t})$ and using the definition for $\mathbf{s}_{i,t}$ in (13), we have

$$\begin{aligned}\|\sigma_{i,t} - \mathbf{s}_{i,t}\|_{\sigma,2} &= \left\| \sum_{j \in V\setminus\{i\}} \beta_{ij} \mathbf{E}_\sigma[\sigma_{j,t} - \sigma_{j,\infty}|\mathcal{H}_{i,t}] \right\|_{\sigma,2} \\ &\le \sum_{j \in V\setminus\{i\}} \beta_{ij} \left\| \mathbf{E}_\sigma[\sigma_{j,t} - \sigma_{j,\infty}|\mathcal{H}_{i,t}] \right\|_{\sigma,2} \\ &\le \sum_{j \in V\setminus\{i\}} \beta_{ij} \left\| \sigma_{j,t} - \sigma_{j,\infty} \right\|_{\sigma,2}. \quad (16)\end{aligned}$$

The first inequality is by the triangle inequality and the second one holds since the conditional expectation is a projection in $L^2(\Omega)$. Now, note that we can further upper bound the left hand side of (16), when we replace, each normed difference inside the sum, $\|\sigma_{j,t} - \sigma_{j,\infty}\|_{\sigma,2}$, by the maximum difference,

$$\|\sigma_{i,t} - \mathbf{s}_{i,t}\|_{\sigma,2} \leq \left(\sum_{j \in V \setminus \{i\}} \beta_{ij} \right) \left(\max_{j \in V \setminus \{i\}} \|\sigma_{j,t} - \sigma_{j,\infty}\|_{\sigma,2} \right). \tag{17}$$

The sum in parentheses in (17) can be bounded using (A2). Furthermore, the right hand side only increases by including i when taking the maximum. As a result,

$$\begin{aligned} \|\sigma_{i,t} - \mathbf{s}_{i,t}\|_{\sigma,2} &\leq \rho \left(\max_{j \in V} \|\sigma_{j,t} - \sigma_{j,\infty}\|_{\sigma,2} \right) \\ &= \rho \|\sigma_t - \sigma_\infty\|_{\sigma,2}, \end{aligned} \tag{18}$$

for all $i \in V$. The equality follows directly by using the definition in (4). Now, consider the first term in (15). By the definition in (4) and the upper bound in (18), we obtain the following bound

$$\|\sigma_t - \mathbf{s}_t\|_{\sigma,2} = \max_{i \in V} \|\sigma_{i,t} - \mathbf{s}_{i,t}\|_{\sigma,2} \leq \rho \|\sigma_t - \sigma_\infty\|_{\sigma,2}. \tag{19}$$

On the other hand by (14), for any $\epsilon > 0$, there exists t_0 such that for all $t > t_0$,

$$\|\mathbf{s}_t - \sigma_\infty\|_{\sigma,2} \leq \epsilon (1 - \rho). \tag{20}$$

When we substitute the bounds in (19) and (20) for terms in (15), we can conclude that $\|\sigma_t - \sigma_\infty\|_{\sigma,2} \leq \epsilon$ for $t > t_0$. Since $\epsilon > 0$ can be chosen arbitrarily small, this implies that each agent's equilibrium strategy converges to a limit strategy $\mathbf{P}_\sigma$-almost surely. ∎

This result implies that there is a limit strategy that the agents play at infinity. The following lemma shows that the limit strategies of agents can be identified by their neighbors.

Lemma 2: If i is a neighbor of j, then $\sigma_{i,\infty}$ is measurable with respect to $\mathcal{H}_{j,\infty}$.

Proof: Since j observes i, for all $\omega \in \Omega$,

$$\sigma_{i,t-1}(h_{i,t-1}(\omega)) \in h_{j,t}(\omega). \tag{21}$$

This implies that $\sigma_{i,t-1}$ is measurable with respect to $\mathcal{H}_{j,t}$. Therefore, since $\sigma_{i,t} \to \sigma_{i,\infty}$ with $\mathbf{P}_\sigma$-probability one and $\mathcal{H}_{j,t} \uparrow \mathcal{H}_{j,\infty}$ as t goes to infinity, $\sigma_{i,\infty}$ is measurable with respect to $\mathcal{H}_{j,\infty}$. ∎

According to Lemma (2), agent j can imitate agent i's limit strategy and vice versa. We use this observation to prove that if i and j are neighbors, then the limit strategy of agent j is a best response strategy for agent i as well.

Theorem 1: Let σ be a BNE strategy. If assumptions (A1) and (A2) are satisfied, then

$$\mathbf{E}_\sigma[u_i(\sigma_{i,\infty}, \sigma_{-i}, \theta) | \mathcal{H}_{i,\infty}] = \mathbf{E}_\sigma[u_i(\sigma_{j,\infty}, \sigma_{-i}, \theta) | \mathcal{H}_{i,\infty}] \tag{22}$$

with $\mathbf{P}_\sigma$-probability one for any $i \in V$, $j \in n(i)$.

Proof: For any $i \in V$, σ_i is measurable with respect to $\mathcal{H}_{i,\infty}$. We define σ_∞ as the vector of limit equilibrium strategy. Let $\sigma^i_\infty = (\sigma_{j,\infty}, \sigma_{-i,\infty})$ be the limit strategy when $i \in V$ unilaterally deviates from equilibrium strategy to play the limit strategy of $j \in n(i)$. This is possible since $\sigma_{j,\infty}$ is measurable with respect to $\mathcal{H}_{i,\infty}$ by Lemma 2. Since network G is undirected, j can also deviate to limit the strategy of $i \in n(j)$. Similarly, define $\sigma^{i,j}_\infty = (\sigma_{j,\infty}, \sigma_{i,\infty}, \sigma_{-i \setminus j,\infty})$ as the limit strategy when connected pairs $i \in V$ and $j \in n(i)$ deviate from their equilibrium strategies by swapping their strategies. By definition of BNE (9),

$$\mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta) | \mathcal{H}_{i,\infty}] \leq \mathbf{E}_\sigma[u_i(\sigma_\infty, \theta) | \mathcal{H}_{i,\infty}] \text{ for all } i \in V. \tag{23}$$

When we take expectation of both sides of (23),

$$\mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta)] \leq \mathbf{E}_\sigma[u_i(\sigma_\infty, \theta)] \quad \text{for all} \quad i \in V. \tag{24}$$

Let $\omega \in \bar{\Omega}$ be the set over which limit action of agent i, $\sigma_{i,\infty}(h_{i,\infty}(w)) \in A_i$, is smaller than $\sigma_{j,\infty}(h_{j,\infty}(w)) \in A_j$; i.e. $\sigma_{i,\infty} \leq \sigma_{j,\infty}$. Since $u_i(\cdot, \theta)$ is supermodular, the conditional expected utility function $\mathbf{E}_\sigma[u(\cdot, \theta) | \mathcal{H}_{i,\infty}]$ is also supermodular. Using this fact, applying definition of supermodularity (2) to limit strategies σ_∞ and $\sigma^{i,j}_\infty$ for any $\omega \in \bar{\Omega}$, and rearranging terms,

$$\begin{aligned} &\mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta) | \mathcal{H}_{i,\infty}] + \mathbf{E}_\sigma[u_i(\sigma^j_\infty, \theta) | \mathcal{H}_{i,\infty}] \geq \\ &\qquad \mathbf{E}_\sigma[u_i(\sigma_\infty, \theta) | \mathcal{H}_{i,\infty}] + \mathbf{E}_\sigma[u_i(\sigma^{i,j}_\infty, \theta) | \mathcal{H}_{i,\infty}]. \end{aligned} \tag{25}$$

We obtain the same relation as in (25) for $\omega \in \Omega \setminus \bar{\Omega}$ in which $\sigma_{i,\infty} > \sigma_{j,\infty}$. Hence, (25) holds for all $\omega \in \Omega$. We take the expectation of both sides of (25) to get

$$\begin{aligned} &\mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta)] + \mathbf{E}_\sigma[u_i(\sigma^j_\infty, \theta)] \geq \\ &\qquad \mathbf{E}_\sigma[u_i(\sigma_\infty, \theta)] + \mathbf{E}_\sigma[u_i(\sigma^{i,j}_\infty, \theta)]. \end{aligned} \tag{26}$$

Notice that the second term on the right hand side is exactly equal to the utility of agent j playing the BNE strategy because the utility functions are symmetric; that is,

$$\mathbf{E}_\sigma[u_i(\sigma^{i,j}_\infty, \theta)] = \mathbf{E}_\sigma[u_j(\sigma_\infty, \theta)]. \tag{27}$$

In the second term on the left hand side of (26), agent j is deviating to agent i's action $\mathbf{E}_\sigma[u_i(\sigma^j_\infty, \theta)]$; hence, both agents are playing with the same strategy. By symmetry of the utility functions,

$$\mathbf{E}_\sigma[u_i(\sigma^j_\infty, \theta)] = \mathbf{E}_\sigma[u_j(\sigma^j_\infty, \theta)]. \tag{28}$$

Substituting (27) and (28) in (26), we can conclude that the individual deviations of i and j provide higher payoff than equilibrium actions; i.e.,

$$\begin{aligned} &\mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta)] + \mathbf{E}_\sigma[u_j(\sigma^j_\infty, \theta)] \geq \\ &\qquad \mathbf{E}_\sigma[u_i(\sigma_\infty, \theta)] + \mathbf{E}_\sigma[u_j(\sigma_\infty, \theta)]. \end{aligned} \tag{29}$$

On the other hand by inequality (24), the first and second terms on the left hand side are less than or equal to the first and second terms on the right hand side, respectively. Hence, the inequalities in (24) must be equalities,

$$\mathbf{E}_\sigma[u_i(\sigma_{j,\infty}, \sigma_{-i,\infty}, \theta)] = \mathbf{E}_\sigma[u_i(\sigma_{i,\infty}, \sigma_{-i,\infty}, \theta)], \tag{30}$$

for all $i \in V$ and $j \in n(i)$. By (23), we have that $\mathbf{E}_\sigma[u_i(\sigma_\infty, \theta) | \mathcal{H}_{i,\infty}] - \mathbf{E}_\sigma[u_i(\sigma^i_\infty, \theta) | \mathcal{H}_{i,\infty}] \geq 0$ with $\mathbf{P}_\sigma$-probability one. This proves (22). ∎

According to Theorem 1, each agent expects that her limit strategy results in a payoff no worse in expectation than if she was to play according to the limit strategy of one of her neighbors. The intuition behind this result is that by the symmetry property of the utility function, strategic complementarity yields that unilateral deviations by i and j to each other's actions are at least as good as playing according to their limit strategy in expectation. However, by the definition of BNE in (9), this behavior can actually never yield strictly better payoffs. Hence, it must be that deviations to neighbors limit strategy are result in equal payoffs in expectation. More importantly, this is also true given the information available to agents, i.e,. from the perspective of agent i, j's limit strategy is just as good as i's limit strategy. An immediate corollary of Theorem 1 is that for a connected network agents reach consensus in their strategies. We formally state and prove this result next.

Corollary 1: Let σ be a BNE strategy. If (A1) and (A2) hold, then

$$\sigma_{i,\infty} = \sigma_{j,\infty} \quad \text{for all} \quad i, j \in V \qquad \mathbf{P}_\sigma\text{-a.s.} \tag{31}$$

Proof: By Theorem 1, the conditional expectations are equal when agents deviate to their neighbors' limit actions. This means that the limit action of agent j is a maximizer of the expected utility of agent i; i.e., $\sigma_{j,\infty} = \mathrm{BR}_{i,\infty}(\sigma, \sigma_{-i,\infty})$. Since agents' best responses are essentially unique, it must be the case that $\sigma_{i,\infty} = \sigma_{j,\infty}$ for all $i \in V$ and $j \in n(i)$ with $\mathbf{P}_\sigma$-probability one. Given that the network is connected, (31) is proven. ■

This results shows that actions of all agents are the same in the limit given a connected network for the type of games we consider.

IV. Linear Games

We now present two examples in which best replies are linear and assumptions (A1) and (A2) hold.

A. Learning in coordination problems

Consider a network of mobile agents that want to align themselves so that they move toward a finish line on a straight path i.e. there exists a correct heading angle $\theta \in [0, 2\pi]$ that is common for all agents. Each agent collects an initial noisy measurement of θ that is denoted by x_i. Mobile agents also have the goal of maintaining the starting formation while moving at equal speed by coordinating their angle of movement with other agents. Agents need to coordinate with the entire population while communication is restricted to neighboring agents whose decisions they can observe. In this context, agent i's decision a_i represents the heading angle. It is possible to formulate the goals of agent i as a maximization of the payoff given by

$$u_i(a, \theta) = -\frac{1-\lambda}{2}(a_i - \theta)^2 - \frac{\lambda}{2(n-1)} \sum_{j \in V \setminus \{i\}} (a_i - a_j)^2, \tag{32}$$

where $\lambda \in (0, 1)$ is a constant measuring the importance of estimation vs. coordination. The first term in (32) is the estimation error in the true heading angle. The second term is the coordination component that weights the discrepancy between self heading and headings of other agents. The utility function is reminiscent of the one in a potential game [3]. Note that the utility function is strictly concave since its second derivative with respect to a_i is negative. Furthermore, it is supermodular since cross derivatives are always positive; i.e., $\partial^2 u_i(\cdot)/\partial a_i \partial a_j = \lambda/n{-}1 > 0$. Further, for the payoff in (32) Assumptions (A1) and (A2) hold: First, $\beta_{ij} = \lambda/2(n-1)$ for all $i \in V$ and $j \in V \setminus \{i\}$; hence, the utility function is symmetric. Second, the sum of the constants that weight the disagreement amongst agents is equal to $\lambda/2 < 1$ hence diagonal dominance is satisfied.

Since the agents can only observe the actions of their neighbors, each agent's best response to the strategies of others σ_{-i} given the information available to her at time t is obtained by solving $\partial \mathbf{E}_\sigma[u_i(a_i, \sigma_{-i}, \theta)|h_{i,t}]/\partial a_i = 0$ as postulated in (6),

$$\begin{aligned} \mathrm{BR}_{i,t}(\sigma; \sigma_{-i,t}) = (1-\lambda)\mathbf{E}_\sigma[\theta|\mathcal{H}_{i,t}] \\ + \frac{\lambda}{n-1} \sum_{j \in V \setminus \{i\}} \mathbf{E}_\sigma[\sigma_{j,t}|\mathcal{H}_{i,t}]. \end{aligned} \tag{33}$$

The same payoff formulation can be motivated by looking at learning in organizations [4]. In an organization, individuals share a common task and have the incentive to coordinate with other units. Each individual receives a private piece of information about the task that needs to be performed while only being able to share her information with individuals with whom she has a direct contact in the organization.

B. Bilateral influences

Consider the population of smoking individuals in a certain region. Smoking is a leading risk factor to human health. One could model this by associating a marginal cost $\theta \in [\underline{c}, \bar{c}]$ with finite and positive bounds $\bar{c} > \underline{c} > 0$ to smoking. Each individual has a prior belief x_i about the risk of smoking. Furthermore, smoking habits are very much determined by social interactions; that is, individuals value conforming to society in the amount of cigarettes they smoke per week. In other words, if the society has a lower rate of smoking, individual i feels that if she smokes more, she would be an outcast in the society. Based on the risks and societal pressures, smokers decide on how many cigarettes they smoke in a given period of time, $a_i \in [0, M]$. Furthermore, individuals only observe smoking behavior of their friends. Given this setup, each individual has the following payoff function,

$$u_i(a, \theta) = -\theta a_i - \frac{1}{2}(a_i - \frac{1}{n-1} \sum_{j \in V \setminus \{i\}} a_j)^2. \tag{34}$$

The first term captures the total risk associated with the amount of cigarettes smoked that the individual wants to minimize. It is proportional to the number of cigarettes consumed. The second term is the utility gain by conforming to the society in smoking behavior.

The payoff function (34) is strictly concave. It is also supermodular since cross derivatives are positive. Assumption

(A1) holds since the utility is of the quadratic form postulated in (1) with $\beta_{ij} = 1/2(n-1)$ for all $i \in V$ and $j \in V \setminus \{i\}$. Further, (A2) is satisfied since the sum of bilateral constants is equal to $1/2 < 1$.

Based on this payoff function, we obtain a linear best response for individual i in consumption of other agents and risk factor θ when we take the derivative of (34) with respect to a_i, equate it to zero and solve for a_i:

$$\mathrm{BR}_{i,t}(\sigma; \sigma_{-i,t}) = -\mathbf{E}_\sigma[\theta|\mathcal{H}_{i,t}] + \frac{1}{n-1} \sum_{j \in V\setminus\{i\}} \mathbf{E}_\sigma[\sigma_{j,t}|\mathcal{H}_{i,t}]. \tag{35}$$

V. Numerical Example

We consider the linear coordination game discussed in Section IV-A with private signals that are normally distributed. Rational learning becomes tractable when private signals are normally distributed and best response function is linear.

Agents play according to BNE strategy where the equilibrium is determined by solving the set of equations determined by definition of BNE (9) and the best response for the linear coordination game (33). At time $t = 0$, agent i plays its private signal x_i since there is no other information available to her. We define the vector of private signals as $\mathbf{x} = [x_1, \ldots, x_n]^T$. At time $t = 1$, agents observe the actions; i.e. the private signals, of their neighbors from previous time, update their estimate for the parameter θ. Since private signals are normally distributed, the estimate of θ will be a linear combination of private signals at all times [14]. Furthermore, according to best response in (33), agent i needs to estimate actions of other agents. Given this, if the actions of all agents except i are linear functions of the private signals, then i's best response is also a linear function of private signals. Let agent i's estimate of θ at time t be given by $\mathbf{E}_\sigma[\theta|\mathcal{H}_{i,t}] = \mathbf{d}_i^T(t)\mathbf{x}$, where $\mathbf{d}_i^T(t)$ is the vector that determines the weight i assigns to different private signals at time t. At time $t = 0$, the ith element of $\mathbf{d}_i^T(0)$ is equal to one and the others are equal to zero. Further, define agent i's estimate of $\mathbf{x}$ at time t as $\mathbf{E}_\sigma[\mathbf{x}|\mathcal{H}_{i,t}] = L_i(t)\mathbf{x}$ where $L_i(t)$ is the weight matrix for estimates of private signals. At time $t = 0$, the ith column of $L_i(0)$ is equal to the vector of ones and the rest of the elements are all equal to zero. This is agent i's estimate of j's private signal x_i for $j \in V\setminus\{i\}$ at time $t = 0$. Furthermore, suppose that agent i's action at time t will be a linear combination of private signals $a_i(t) = \mathbf{v}_i^T(t)\mathbf{x}$. Then, by (9), the equilibrium strategy is obtained by solving the following set of linear equations for $\mathbf{v}_i^T(t)$,

$$\mathbf{v}_i^T(t)\mathbf{x} = (1-\lambda)\mathbf{d}_i^T(t)\mathbf{x} + \frac{\lambda}{n-1} \sum_{j \in V\setminus\{i\}} \mathbf{v}_j^T(t) L_i(t)\mathbf{x} \tag{36}$$

for all $i \in V$. Actions of agents are linear combinations of private signals at all times, and hence remain normally distributed. Given common knowledge of rational behavior and network, agent i can efficiently calculate $\mathbf{v}_j(t)$ for $j \in n(i)$, and hence know the weighting of private signals. Using her knowledge of $\mathbf{v}_j(t)$ for $j \in n(i)$, agent i can efficiently update her belief on θ and her response for time $t+1$.

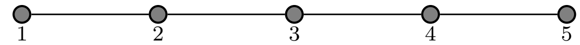

Fig. 1. Line network

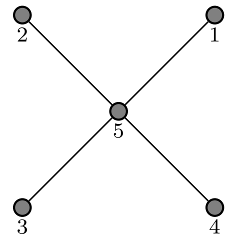

Fig. 2. Star network

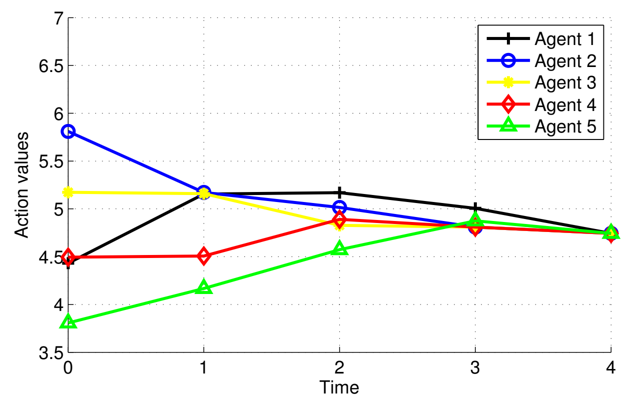

Fig. 3. Values of actions over time for the line network in Fig. 1.

Here, we consider two types of networks, a line network and a star network each with $n = 5$ agents (see Figs. 1 and 2, respectively.) The correct movement angle is chosen to be $\theta = 5°$. Initially, agent i obtains a noisy private measurement about the correct angle

$$x_i = \theta + w_i, \tag{37}$$

where $w_i \sim \mathcal{N}(0, 1)$. We choose $\lambda = 0.5$.

Figs. 3 and 4 show the evolution of actions corresponding to networks in Figs. 1 and 2, respectively. The actions of agents reach consensus at time $t = 4$ for the line network (see Fig. 3.) For the star network, consensus happens in $t = 2$ steps (see Fig. 4.) These results suggest that consensus is achieved in $O(d)$ steps where d is the diameter of the graph. For the distributed estimation problems over tree networks [14] convergence happens in $O(d)$ steps as well. Here, we observe the same result for learning in linear games. Furthermore, the consensus action is optimal; that is, agents converge to the mean of all the private signals which is the optimal estimate of θ given all of the private signals.

VI. Conclusion

We considered learning in repetitive games with quadratic payoff functions in which agents start with noisy private signals. We assumed that the payoff function is symmetric

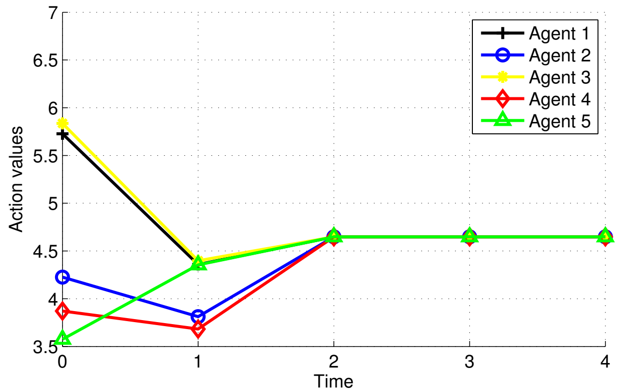

Fig. 4. Values of actions over time for the star network in Fig. 2.

and that agents discount the bilateral effects of actions of other agents in the payoff. We looked at the case where agent behavior is rational and myopic, i.e., when agents are able to incorporate new information according to Bayes' rule and do not consider future plays in their current decision making. Since Bayesian updates are intractable in the general case, we performed an asymptotic analysis. We first showed that there exists a limit strategy for each agent. This result is used to show that the agents' limit strategies are measurable with respect to their neighbors' information. Our main result demonstrates that the agents are indifferent (in expectation) between playing according to their own limit strategies and those of their neighbors. Consequently, consensus is achieved since agents' best responses are essentially unique and the network is connected. Simulations are used to confirm the technical results. They also suggest that the convergence time is equal to the diameter of the network.

References

[1] G.M. Angeletos and A. Pavan. Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142, 2007.

[2] Y. Bramoullé, R. Kranton, and M. D'Amours. Strategic interaction and networks. *Working Paper*, 2009, Available at SSRN: http://ssrn.com/abstract=1612369.

[3] J.R. Marden, G. Arslan, and J.S. Shamma. Cooperative control and potential games. *IEEE Trans. Syst., Man, and Cybern. B, Cybern.*, 39(6):1393–1407, 2009.

[4] A. Calvó-Armengol and J.M. Beltran. Information gathering in organizations: Equilibrium, welfare and optimal network structure. *Journal of the European Economic Association*, 7:116–161, 2009.

[5] S. Morris and H.S. Shin. The social value of public information. *American Economic Review*, 92:1521–1534, 2002.

[6] C. Ballester, A. Calvó-Armengol, and Y. Zenou. Who's who in networks. wanted: The key player. *Econometrica*, 74:1403–1417, 2006.

[7] D. Gale and S. Kariv. Bayesian learning in social networks. *Games Econ. Behav.*, 45:329–346, 2003.

[8] D. Rosenberg, E. Solan, and N. Vieille. Informational externalities and emergence of consensus. *Games Econ. Behav.*, 66:979–994, 2009.

[9] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 2012.

[10] P. M. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118:909–968, 2003.

[11] V. Bala and S. Goyal. Learning from neighbours. *Review of Economic Studies*, 65(3), 1998.

[12] M. Rabbat, R. Nowak, and J. Bucklew. Generalized consensus computation in networked systems with erasure links. In *Proc. of IEEE 6th Workshop on the Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1088–1092, New York, NY, USA., June 2005.

[13] S. Kar and J. M. Moura. Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs. *IEEE J. Sel. Topics Signal Process.*, 5(4):674–690, 2011.

[14] E. Mossel and O. Tamuz. Efficient bayesian learning in social networks with gaussian estimators. Arxiv 1002.0747, 2011.