2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Bayesian Graphical Games for Synchronization in Dynamical Systems

Victor G. Lopez, Yan Wan, *Senior Member, IEEE*, and Frank L. Lewis, *Fellow, IEEE*

***Abstract*— In this paper, differential games with incomplete information, or Bayesian games, are formulated for a set of continuous-time dynamical systems linked together by a communication graph. These new Bayesian graphical games for dynamical systems represent the situation where the agents are uncertain about their actual payoff and must collect additional information to improve their estimation of the real setting of their environment. Furthermore, the agents play their best response strategies with respect to the policies of their neighbors. A tight relationship between the beliefs of an agent and his distributed best response policy is obtained. Conditions for the so-called Bayes-Nash equilibrium are provided. A distributed belief update algorithm is developed that does not require the full knowledge of graph topology.**

I. Introduction

Game theory has become one of the most useful tools in multiagent systems analysis due to their rigorous mathematical representation of optimal decision making [1]. Differential games have been studied with increasing interest to encompass the need of the players to consider the evolution of their payoff functions along time rather than static, immediate costs per action [2]-[6]. It is proven in [2] that if the agents use the solutions of the Hamilton-Jacobi-Isaacs (HJI) equations in their feedback control policies, then Nash equilibrium is achieved in the game and no agent can unilaterally change his control policy without negatively affecting his performance.

A downside of these standard differential games solutions is the assumption that all agents are fully aware of all the aspects of the game being played. In complex practical applications, the agents operate in fast-evolving and uncertain environments. A more general case has been described with the study of graphical games [6]-[10], in which the agents are taken as nodes in a communication graph, such that each agent can only observe the state of a subset of all other agents, regarded as his neighbors. In graphical games the agents must only use their partial knowledge of the game to achieve desirable outcomes in cooperative and adversarial networks. The objectives of every player in the game, however, are still assumed to be known by every agent.

Bayesian games [1], [11]-[14], or games with incomplete information, describe the situation on which the agents participate in an unspecified game. The true intentions of the other players may be unknown, and each agent must adjust his objectives accordingly. The initial information of each agent about the game, and the personal experience gained during his interaction with other agents, form the basis for the epistemic analysis of the dynamical systems. Each agent must employ the evidence that his environment provides to update his beliefs about the game. Thus, the aim is to develop belief assurance protocols, distributed control protocols and distributed learning mechanisms to induce optimal behaviors with respect to a cost function. Different learning algorithms for static agents in Bayesian games have been studied [15]-[18], but not for differential graphical games per knowledge of the authors.

The main contributions of this paper are the following. A novel description of Bayesian games for continuous-time dynamical systems, which requires an adequate definition of the expected cost that is to be minimized by each agent is proposed. This leads to the definition of the Bayes-Nash equilibrium for dynamical systems, which is obtained by solving a set of HJI equations that include the epistemic beliefs of the agents as a parameter. We call these partial differential equations the Bayes-Hamilton-Jacobi-Isaacs (BHJI) equations. We reveal for the first time the tight relationship between the beliefs of an agent and his distributed best response control policy. The beliefs of the agents are constantly updated throughout the game using the Bayesian rule to incorporate new evidence to the individual current estimates of the game.

The paper is structured as follows. Section 2 presents the formal mathematical definitions of Bayesian games and graphical games. Section 3 presents the formulation of Bayesian games for dynamical systems in a graph topology; the best response strategies for the minimization of the expected costs of every agent are obtained. Section 4 is focused on the Bayesian algorithm for the belief updates. Finally, a simulation of the proposed control scheme is presented in Section 5.

II. Preliminaries

This section states the formal definition of Bayesian games and graphical games, which will be used extensively in this paper for the formulation and analysis of Bayesian graphical games for dynamical systems.

This work is partially supported by ONR grant N00014-17-1-2239, NSF grants under numbers 1714519 and 1730675, and China NSFC grant number 61633007.

V. G. Lopez is with the UTA Research Institute, University of Texas at Arlington, Fort Worth, TX 76118, USA (phone: 817-262-7520, e-mail: victor.lopezmejia@mavs.uta.edu).

Y. Wan is with the Electrical Engineering, University of Texas at Arlington, TX 76010, UTA (email: yan.wan@uta.edu).

F. L. Lewis is with the UTA Research Institute, University of Texas at Arlington, Fort Worth, TX 76118, USA and Qian Ren Consulting Professor, State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang China (e-mail: lewis@uta.edu).

978-1-5386-5427-9/$31.00 ©2018 AACC

A. *Bayesian games*

Many practical applications of game-theoretic models require considering players with incomplete knowledge about their environment. The number of players, the set of possible actions and the cost paid for each action are aspects of the games that can be unknown to the agents. The category of games that considers incomplete information environments is regarded as Bayesian games [1], [11].

The information that is unknown by the agents in a Bayesian game can be often formulated as an uncertainty about the payoff corresponding to each possible action [1]. Thus, the players are presented with a set of possible games, one of which is being played. Being aware of their lack of knowledge, the agents must define a probability distribution over the set of all possible games they may be engaged on. We call these probabilities the *beliefs* of an agent.

Each agent has two types of knowledge. First, a *common prior* is assumed to be known by all the agents, and is taken as the starting point for them to make rational inferences about the game. In repeated games, the common prior is updated individually based on the information collected by the agent. Second, the agents start with some personal information, only known by themselves, and regarded as their *epistemic type*. The goals of an agent during the game depend on his current type and the types of the other agents.

For each of the N agents, define the *epistemic type space* that represents the private information available to the agent. The epistemic type space for agent i is defined as $\Theta_i = \{\theta_i^1, \ldots, \theta_i^{M_i}\}$, where θ_i^k, $k = 1, \ldots, M_i$, represent the different possible epistemic types for agent i. When there is no risk of ambiguity, we ease our notation representing the current type of agent i as θ_i.

Formally, a Bayesian game for N players is defined as a tuple (N, A, Θ, P, J), where N is the set of agents in the game, $A = A_1 \times \cdots \times A_N$, with A_i the set of possible actions of agent i, $\Theta = \Theta_1 \times \cdots \times \Theta_N$ with Θ_i the type space of player i, $P : \Theta \to [0,1]$ expresses the probability of finding every agent i in type θ_i^k, $k=1,\ldots,M_i$, and the payoff function of the agents are $J = (J_1, \ldots, J_N)$.

B. *Differential games on graphs*

Graphical games capture the dynamics of a multiagent system with limited sensing capabilities; that is, every player in the game can only interact with a subset of the other players, regarded as his neighbors. Consider a set of N agents connected by a communication graph $G = (V, E)$. The edge weights of the graph are represented as a_{ij}, with $a_{ij} > 0$ if $(v_j, v_i) \in E$ and $a_{ij} = 0$ otherwise. By assumption, there are no self-loops in the graph, i.e., $a_{ii} = 0$ for all players i. The weighted in-degree of node i is defined as $d_i = \sum_{j=1}^{N} a_{ij}$.

Consider a canonical leader-follower synchronization game. Each node of the graph G represents a player of the game, consisting on a system with linear dynamics as

$$\dot{x}_i = Ax_i + Bu_i, \quad i = 1, \ldots, N, \tag{1}$$

where $x_i(t) \in \mathbb{R}^n$ is the vector of state variables and $u_i \in \mathbb{R}^m$ is the control input vector of agent i. Consider an extra node, regarded as the leader node, with state dynamics

$$\dot{x}_0 = Ax_0. \tag{2}$$

The leader is connected to the other nodes by means of the pinning gains $g_i \geq 0$. This paper studies the general objective of achieving synchronization with the leader node.

The local synchronization error for agent i is defined as

$$\delta_i = \sum_{j=1}^{N} a_{ij}(x_i - x_j) + g_i(x_i - x_0). \tag{3}$$

Using the expressions (1) - (2), the local error dynamics are

$$\dot{\delta}_i = A\delta_i + (d_i + g_i)Bu_i - \sum_{j=1}^{N} a_{ij}Bu_j. \tag{4}$$

Each agent i expresses his objective in the game by defining a performance index as

$$J_i(\delta_i, \delta_{-i}, u_i, u_{-i}) = \int_0^\infty r_i(\delta_i, \delta_{-i}, u_i, u_{-i})dt, \tag{5}$$

where $r_i(\delta_i, \delta_{-i}, u_i, u_{-i})$ is selected as a positive definite scalar function of the variables expected to be minimized by agent i, with δ_{-i} and u_{-i} the local errors and control inputs of the neighbors of agent i, respectively. For synchronization games, r_i can be selected as

$$r_i(\delta_i, \delta_{-i}, u_i, u_{-i}) = \sum_{j=1}^{N} a_{ij}\bar{\delta}_{ij}^T Q_{ij}\bar{\delta}_{ij} + u_i^T R_{ii}u_i + \sum_{j=1}^{N} a_{ij}u_j^T R_{ij}u_j, \tag{6}$$

where $\bar{\delta}_{ij} = \begin{bmatrix} \delta_i^T & \delta_j^T \end{bmatrix}^T$, $Q_{ij} = Q_{ij}^T > 0$ and $R_{ii} = R_{ii}^T > 0$. It is also presented in a simplified form,

$$r_i(\delta_i, u_i, u_{-i}) = \delta_i^T Q_i \delta_i + u_i^T R_{ii}u_i + \sum_{j=1}^{N} a_{ij}u_j^T R_{ij}u_j, \tag{7}$$

which is widely employed in the differential graphical games literature [6], [9].

The best response of agent i for fixed neighbor policies u_{-i} is defined as the control policy u_i^* such that the inequality $J_i(u_i^*, u_{-i}) \leq J_i(u_i, u_{-i})$ holds for all policies u_i. Nash equilibrium is achieved if every agent plays his best response with respect to all his neighbors, that is,

$$J_i(\delta, u_i^*, u_{-i}^*) \leq J_i(\delta, u_i, u_{-i}^*), \qquad i = 1, \ldots, N. \tag{8}$$

It is proven in [6] that the best response of agent i with cost function defined by (5) and (7) is given by

$$u_i^* = -\frac{1}{2}(d_i + g_i)R_{ii}^{-1}B^T \nabla V_i(\delta_i), \tag{9}$$

where the functions $V_i(\delta_i)$ are the solutions of a set of the Hamilton-Jacobi-Isaacs (HJI) equations,

$$r_i(\delta,u_i,u_{-i})+\nabla V_i^T\left(A\delta_i+(d_i+g_i)Bu_i^*-\sum_{j=1}^{N}a_{ij}Bu_j^*\right)=0. \quad (10)$$

III. Bayesian Graphical Games for Dynamic Systems

This section presents our main results on the formulation of Bayesian games for dynamical systems connected by a communication graph and the analysis of the conditions to achieve Bayes-Nash equilibrium in the game.

A. *Bayesian graphical game formulation*

Consider a system of *N* agents with linear dynamics (1), distributed on a communication graph *G*, with leader state dynamics (2) and local synchronization errors (3). The desired objective of an agent depend on his current type and those of his neighbors. This condition can be expressed by defining the performance index of agent *i* as

$$J_i^\theta(\delta_i,u_i,u_{-i})=\int_0^\infty r_i^\theta(\delta_i,u_i,u_{-i})dt, \quad (11)$$

where θ refers to the set of current types of all the agents in the game, $\theta=\theta_1\times\cdots\times\theta_N$, as defined in Section 2-A, and each function r_i^θ is defined for that particular combination of types. We define a new category of game as follows.

Definition 1. A Bayesian graphical game for dynamical systems is defined as a tuple (N,X,U,Θ,P,J) where *N* is the set of agents in the game, $X=X_1\times\cdots\times X_N$ is a set of states with X_i the set of reachable states of agent *i*, $U=U_1\times\cdots\times U_N$ with U_i the set of admissible controllers for agent *i* and $\Theta=\Theta_1\times\cdots\times\Theta_N$ with Θ_i the type space of player *i*. The common prior over types $P:\Theta\rightarrow[0,1]$ describes the probability of finding every agent *i* in type $\theta_i^k\in\Theta_i$, $k=1,\ldots,M_i$, at the beginning of the game. The performance indices $J=(J_1,\ldots,J_N)$, with $J_i:X\times U\times\Theta\rightarrow\mathbb{R}$, are the costs of every agent for the use of a given control policy in a state value and a particular combination of types.

Define the set $\Delta_i=X_1^i\times\cdots\times X_N^i$, where X_j^i is the set of possible states of the *j*th neighbor of agent *i*; that is, Δ_i represents the set of states that agent *i* can observe from the graph topology.

It is assumed that the sets *N*, *X*, *U*, *P* and *J* are of common prior for all the agents before the game starts. However, the set of states Δ_i and the actual type θ_i are known only by agent *i*. The objective of every agent in the game is now to use their (limited) knowledge about δ_i and θ to determine the control policies $u_i^*(\delta_i,\theta)$, such that every agent expects to minimize the cost he pays during the game according to the cost functions (11).

To fulfill this objective, a different cost index formulation is required to allow the agents to determine their optimal policies according to their current beliefs. This requirement is addressed by defining the *expected cost* of agent *i*.

Definition 2. Given a Bayesian game (N,X,U,Θ,P,J), where the agents play with policies u_i and the type of agent *i* is θ_i, the *ex interim* expected cost is

$$EJ_i(\delta_i,u_i,u_{-i},\theta_i)=\sum_{\theta\in\Theta}p(\theta\,|\,\delta_i,\theta_i)J_i^\theta(\delta_i,u_i,u_{-i}), \quad (12)$$

where $p(\theta\,|\,\delta_i,\theta_i)$ is the probability of having global type θ, given the information that agent *i* possesses, and the summation index $\theta\in\Theta$ indicates that all possible combination of types in the game must be considered.

B. *Best response policy and Bayes-Nash equilibrium*

The best response of an agent in a Bayesian game for given fixed neighbor strategies u_{-i}, is defined as the control policy that minimizes the expected cost (12). Formally, agent *i*'s best response to control policies u_{-i} is given by

$$u_i^*=\arg\min_{u_i}EJ_i(\delta_i,u_i,u_{-i},\theta) \quad (13)$$

Now, it is said that a Bayes-Nash equilibrium is reached in the game if each agent plays a best response to the strategies of the other players during a Bayesian game. The Bayes-Nash equilibrium is the most important solution concept in Bayesian graphical games for dynamical systems. Definition 3 formalizes this idea.

Definition 3. A Bayes-Nash equilibrium is a set of control policies $u=u_1\times\cdots\times u_N$ that satisfies $u_i=u_i^*$, as in (13), for all agents *i*, such that

$$EJ_i(\delta_i,u_i^*,u_{-i}^*)\le EJ_i(\delta_i,u_i,u_{-i}^*) \quad (14)$$

for any control policy u_i.

Following an analogous procedure to single-agent optimal control, define the value function of agent *i*, given the types of all agents θ, as

$$V_i^\theta(\delta_i,u_i,u_{-i})=\int_t^\infty r_i^\theta(\delta_i,u_i,u_{-i})d\tau, \quad (15)$$

with r_i^θ as defined in (11). The *expected value* function for a control policy u_i is defined as

$$EV_i(\delta_i,u_i,u_{-i},\theta)=\sum_{\theta\in\Theta}p(\theta\,|\,\delta_i,\theta_i)V_i^\theta(\delta_i,u_i,u_{-i}). \quad (16)$$

Function (16) can be used to define the *expected Hamiltonian* of agent *i* as

$$EH_i(\delta_i,u,\theta)=\sum_{\theta\in\Theta}p(\theta\,|\,\delta_i,\theta_i)\times \left[r_i^\theta(\delta_i,u)+\nabla V_i^{\theta T}\left(A\delta_i+(d_i+g_i)Bu_i-\sum\nolimits_{j=1}^{N}a_{ij}Bu_j\right)\right]. \quad (17)$$

The expected Hamiltonian (17) is now employed to determine the best response control policy of agent *i*, by computing its derivative with respect to u_i and equating it to zero. This procedure yields the optimal policy

$$u_i^* = -\frac{1}{2}(d_i + g_i)\left[\sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) R_{ii}^{\theta}\right]^{-1} \times \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) B^T \nabla V_i^{\theta} \qquad (18)$$

As in the deterministic multiplayer nonzero-sum games [2], the functions $V_i^{\theta}(\delta_i)$ are the solutions of a set of coupled partial differential equations which, for the setting of Bayesian games, we refer to as Bayes-Hamilton-Jacobi-Isaacs (BHJI) equations, and are given by

$$\sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\Big[r_i^{\theta}(\delta_i, u) + \nabla V_i^{\theta T} \times\Big(A\delta_i + (d_i + g_i)Bu_i - \textstyle\sum_{j=1}^{N} a_{ij} B u_j\Big)\Big] = 0 \qquad (19)$$

Remark 1. The optimal control policy (18) establishes for the first time, the relation between belief and distributed control in multi-agent systems with unawareness. Each agent should compute his best response by observing only his immediate neighbors. This is distributed computation with bounded rationality imposed by the communication network.

The next lemma shows that the Hamiltonian function for general policies u_i, u_{-i} can be expressed as a quadratic form of the optimal policies u_i^* and u_{-i}^* defined in (18).

Lemma 1. Given the expected Hamiltonian function (17) for agent *i* and the optimal control policy (18), then

$$EH_i(\delta_i, u_i, u_{-i}) = EH_i(\delta_i, u_i^*, u_{-i}) + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)(u_i - u_i^*)^T R_{ii}^{\theta}(u_i - u_i^*) \qquad (20)$$

Proof. The proof is similar to the proof of Lemma 10.1-1 in [2], performed by completing the squares in (17) to obtain

$$\begin{aligned} EH_i(\delta_i, u, \theta) = \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) \times \\ \Bigg[\delta_i^T Q_i^{\theta} \delta_i + u_i^T R_{ii}^{\theta} u_i + \sum_{j=1}^{N} a_{ij} u_j^T R_{ij}^{\theta} u_j + u_i^{*T} R_{ii}^{\theta} u_i^* \\ - u_i^{*T} R_{ii}^{\theta} u_i^* + (d_i + g_i)\nabla V_i^{\theta T} B u_i^* - (d_i + g_i)\nabla V_i^{\theta T} B u_i^* \\ + \nabla V_i^{\theta T}\left(A\delta_i + (d_i + g_i)Bu_i - \sum_{j=1}^{N} a_{ij} B u_j \right)\Bigg] \end{aligned}$$

and conducting algebraic operations to obtain (20). □

The following theorem shows that Bayes-Nash equilibrium is achieved by means of control policies (18). The proof is performed using quadratic cost functions as in (7), but it can easily be extended to other functions as (6).

Theorem 1. **Bayes-Nash Equilibrium**. Consider a multiagent system on a communication graph, with agents' dynamics (1) and target node dynamics (2). Let $V_i^{*\theta}(\delta_i)$, $i = 1, \ldots, N$, be the solutions of the BHJI equations (19). Define the control policy u_i^* as in (18). Then, control inputs u_i^* make the dynamics (4) asymptotically stable for all agents. Moreover, all agents are in Bayes-Nash equilibrium as defined in Definition 2, and the corresponding expected costs of the game are

$$EJ_i^* = V_i^{\theta *}(\delta_i(0)).$$

Proof. (Stability) Take the expected value function (16) as a Lyapunov function candidate. Its derivative, taking $p(\theta|\delta_i,\theta_i)$ constant in a time interval (see Section 4), is given by

$$E\dot{V}_i = \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\dot{V}_i^{\theta} = \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\nabla V_i^{\theta T}\dot{\delta}_i.$$

As V_i^{θ} satisfies Equation (19), then

$$E\dot{V}_i = -\sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\left(\delta_i^T Q_i^{\theta}\delta_i + u_i^T R_{ii}^{\theta} u_i + \sum_{j=1}^{N} a_{ij} u_j^T R_{ij}^{\theta} u_j \right) \le 0$$

and the dynamics (4) are stable.

(Bayes-Nash equilibrium) Notice that $V_i^{\theta}(\delta_i(\infty)) = 0$ because of the stability of the system. Now, the expected cost of the game for agent *i* is expressed as

$$\begin{aligned} EJ_i &= \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\int_0^{\infty}\left(\delta_i^T Q_i^{\theta}\delta_i + u_i^T R_{ii}^{\theta} u_i + \sum_{j=1}^{N} a_{ij} u_j^T R_{ij}^{\theta} u_j \right) dt \\ &\quad + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\int_0^{\infty}\dot{V}_i^{\theta} dt + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) V_i^{\theta}(\delta_i(0)) \\ &= EH_i(\delta_i, u_i^*, u_{-i}) + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) V_i^{\theta}(\delta_i(0)). \end{aligned}$$

By Lemma 1, this expression becomes

$$\begin{aligned} EJ_i &= H(\delta_i, u_i^*, u_{-i}) + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)(u_i - u_i^*)^T R_{ii}(u_i - u_i^*) \\ &\quad + \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i) V_i^{\theta}(\delta_i(0)) \end{aligned}$$

for all u_i and u_{-i}. Assume all the neighbors of agent *i* are using their best response strategies u_{-i}^*. Then, as the BHJI equations (19) holds, we have

$$EJ_i = \sum_{\theta\in\Theta} p(\theta \mid \delta_i, \theta_i)\left[(u_i - u_i^*)^T R_{ii}(u_i - u_i^*) + V_i^{\theta}(\delta_i(0)) \right]$$

We conclude that u_i^* minimizes the expected cost of agent *i* and the value of the game is $EV_i^{\theta}(\delta_i(0))$. □

The probabilities $p(\theta \mid \delta_i, \theta_i)$ in the control policies (18) have an initial value given by the common prior of the agents, expressed by *P* in Definition 1. However, as the system dynamics (1) - (2) evolve through time, all agents are able to collect new evidence that can be used to update their estimates of the probabilities of the types θ. This *belief update* scheme is studied in the next section.

IV. Bayesian Belief Updates

Let every agent in the game to revise his beliefs every *T* units of time. Then, using his knowledge about his type θ_i, the previous states of his neighbors $x_{-i}(t)$, and the current states of the neighbors $x_{-i}(t+T)$, agent *i* can perform his belief update at time *t*+*T* using the Bayesian rule as

$$p(\theta \mid x_{-i}(t+T), x_{-i}(t), \theta_i) = \frac{p(x_{-i}(t+T) \mid x_{-i}(t), \theta) p(\theta \mid x_{-i}(t), \theta_i)}{p(x_{-i}(t+T) \mid x_{-i}(t), \theta_i)} \quad (21)$$

where $p(\theta \mid x_{-i}(t), \theta_i)$ is agent *i*'s belief at time *t* about the types θ, $p(x_{-i}(t+T) \mid x_{-i}(t), \theta)$ is the likelihood of the neighbors reaching the states $x_{-i}(t+T)$ T time units after being in states $x_{-i}(t)$ given that the global type is θ, and $p(x_{-i}(t+T) \mid x_{-i}(t), \theta_i)$ is the overall probability of the neighbors reaching $x_{-i}(t+T)$ from $x_{-i}(t)$ regardless of every other agent's type.

Remark 2. Notice that $p(\theta \mid \delta_i(t), \theta_i) = p(\theta \mid x_{-i}(t), \theta_i)$ because, in the game here defined, an agent cannot use his own state as evidence for the global type θ.

The term $p(\theta \mid x_{-i}(t), \theta_i)$ in (21) expresses the joint probability of the types of each individual agent, that is, $p(\theta \mid x_{-i}(t), \theta_i) = p(\theta_1, \ldots, \theta_N \mid x_{-i}(t), \theta_i)$. In some applications, the Bayesian graphical game is defined such that the types of the agents do not depend on each other. Thus, the knowledge of an agent about one type does not affect his belief in the others. In this case we can write the expression $p(\theta_1, \theta_2, \ldots, \theta_N \mid x_{-i}(t)) = p(\theta_1 \mid x_{-i}(t)) p(\theta_2 \mid x_{-i}(t)) \cdots p(\theta_N \mid x_{-i}(t))$.

Following a similar procedure, the likelihood function $p(x_{-i}(t+T) \mid x_{-i}(t), \theta)$ can be expressed in terms of the individual positions of agent *i*'s neighbors as the probability $p(x_{-i}(t+T) \mid x_{-i}(t), \theta) = p(x_1^i(t+T), \ldots, x_{N_i}^i(t+T) \mid x_{-i}(t), \theta)$, where $x_j^i(t)$ is the state of the *j*th neighbor of *i*. Notice that $x_i(t+T)$ is dependent of $x_i(t)$ and of $x_{-i}(t)$ by means of the control input u_i, for all agents *i*. However, the current state value of agent *i*, $x_i(t+T)$, is independent of the current state value of his neighbors, $x_{-i}(t+T)$, for there has been no time for the values $x_{-i}(t+T)$ to affect the policy u_i. Independence of the state variables at time $t+T$ allows writing $p(x_{-i}(t+T) \mid x_{-i}(t), \theta) = \prod_{j \in N_i} p(x_j(t+T) \mid x_{-i}(t), \theta)$.

Similarly, the denominator of (21) can be expressed as the product $p(x_{-i}(t+T) \mid x_{-i}(t), \theta_i) = \prod_{j \in N_i} p(x_j(t+T) \mid x_{-i}(t), \theta_i)$.

Using these expressions, the belief update (21) can be written as

$$p(\theta \mid x_{-i}(t+T), x_{-i}(t), \theta_i) = \prod_{j \in N_i} \frac{p(x_j(t+T) \mid x_{-i}(t), \theta) p(\theta_j \mid x_{-i}(t))}{p(x_j(t+T) \mid x_{-i}(t), \theta_i)} \times \prod_{k \notin N_i} p(\theta_k \mid x_{-i}(t)), \quad (22)$$

where the set of factors $\prod_{k \notin N_i} p(\theta_k \mid x_{-i}(t))$ consists on the types of the non-neighbors of agent *i*.

V. Simulation Results

The following simulation is performed to show the behavior of the agents during a Bayesian graphical game. The solution of the BHJI equations is given and every agent uses his optimal policy corresponding to the actual combination of types θ, if he knows it.

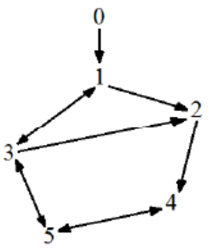

Figure 1. Graph topology employed in simulation.

Consider a multiagent system with 5 agents and one leader, connected in a directed graph as shown in Fig. 1. All agents are taken with single integrator dynamics, as

$$\dot{x}_i = \begin{bmatrix} \dot{x}_{i,1} \\ \dot{x}_{i,2} \end{bmatrix} = \begin{bmatrix} u_{i,1} \\ u_{i,2} \end{bmatrix}$$

In this game, only agent 1 has two possible types, and all other agents start with a prior knowledge of the probabilities of each type. Let agent 1 have type 1 40% of the cases, and type 2 60% of the cases.

The cost functions of the agents are taken in the form (6), considering the same weighting matrices for all agents; that is, $Q_{ij}^{\theta_1} = Q_{kl}^{\theta_1}$, $R_{ij}^{\theta_1} = R_{kl}^{\theta_1}$, $Q_{ij}^{\theta_2} = Q_{kl}^{\theta_2}$ and $R_{ij}^{\theta_2} = R_{kl}^{\theta_2}$ for all $i, j, k, l \in \{1,2,3,4,5\}$. For type θ_1, the matrices are taken as

$$Q_{ij}^{\theta_1} = \begin{bmatrix} \frac{1}{10} I & -\frac{1}{10} I \\ -\frac{1}{10} I & \frac{2}{10} I \end{bmatrix},$$

$R_{ii}^{\theta_1} = 10$ and $R_{ij}^{\theta_1} = -20$ for $i \neq j$, where I is the identity matrix. The solutions of the corresponding HJI equations are

$$P_i^{\theta_2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

for all agents. For the cost functions of type θ_2, define

$$Q_{ij}^{\theta_2} = \begin{bmatrix} 4I & -4I \\ -4I & 8I \end{bmatrix},$$

$R_{ii}^{\theta_2} = 1$ and $R_{ij}^{\theta_2} = -2$ for $i \neq j$. This yields the solutions

$$P_i^{\theta_2} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

for all agents.

With the exception of agent 1, all players update their beliefs about the type θ every 0.1 seconds, using a Bayesian belief update as described in Section 4. During this simulation, agent 1 is in type 1.

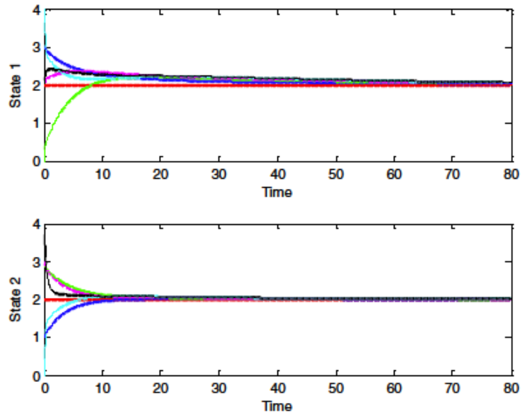

Figure 2. Trajectories for both states of the five agents.

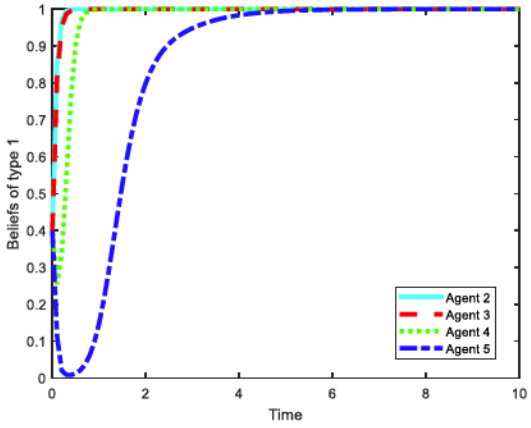

Figure 3. Convergence of the beliefs of the five agents about type 1.

The state dynamics of the agents are shown in Fig. 2. In Fig. 3, the evolution of the beliefs of every agent is displayed. Note that all beliefs approach probability one for type θ_1, and all agents end up playing the same game.

VI. Conclusion

Multiagent systems analysis was performed for dynamical agents engaged on interactions with uncertain objectives. We reveal for the first time the tight relationship between the beliefs of an agent and his distributed best response control policy. The Bayes-Nash equilibrium were proved for the best response control policy to achieve under general conditions. The proposed Bayesian belief update scheme works appropriately provided that the information available to the agents is not excessively restricted. A more practical method to compute the likelihood function required for the Bayesian update is under development by the authors.

Acknowledgment

Author V. G. Lopez thanks the Mexican Council of Science and Technology (Conacyt) for its support.

References

[1] Y. Shoham and K. Leyton-Brown, *Multiagent systems. Algorithmic, Game-Theoretic and Logical Foundations.* New York, NY: Cambridge University Press, 2008.

[2] F. L. Lewis, D. Vrabie and V. L. Syrmos, *Optimal Control,* 2nd ed. New Jersey: John Wiley & Sons, inc., 2012.

[3] H. Li, D. Liu and D. Wang, "Integral reinforcement learning for linear continuous-time zero-sum games with completely unknown dynamics," *IEEE Transactions on Automation Science and Engineering*, vol. 11, No. 3, pp. 706-714, 2014.

[4] P. Kumar and J. Van Schuppen, "On Nash equilibrium solutions in stochastic dynamic games," *IEEE Transactions on Automatic Control*, vol. 25, No. 6, pp. 1146-1149, 1980.

[5] W. Lin, Z. Qu and M. A. Simaan, "Nash strategies for pursuit-evasion differential games involving limited observations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, No. 2, pp. 1347-1356, 2015.

[6] K. G. Vamvoudakis, F. L. Lewis and G. R. Hudas, "Multiagent differential graphical games: Online adaptive learning solution for synchronization with optimality," *Automatica*, vol. 48, pp. 1598-1611, 2012.

[7] Z. Li, Z. Duan, G. Chen and L. Huang, "Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint," *IEEE Trans. Circuits and Systems*, vol. 57, No. 1, pp. 213–224, 2010.

[8] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, No. 1, pp. 215–233, 2007.

[9] M. I. Abouheaf and M. S. Mahmoud, "Online policy iteration solution for dynamic graphical games," presented at the 13th International Multi-Conference on Systems, Signals & Devices, Leipzig, Germany, 2016, pp. 787–797.

[10] R. Kamalapurkar, J. R. Klotz, P. Walters and W. E. Dixon, "Model-based reinforcement learning in differential graphical games," *IEEE Trans. Control of Network Systems*, to be published.

[11] J. C. Harsanyi, "Games with incomplete information played by Bayesian players, I-III," *Management Science Theory*, vol. 14, No. 3, pp. 159–182, 1967.

[12] E. Einy, O. Haimanko, D. Moreno and B. Shitovitz, "On the existence of Bayesian Cournot equilibrium," *Games and Economic Behavior*, vol. 68, pp. 77-94, 2010.

[13] G. Carmona and K. Podczeck, "Ex-post stability of Bayes-Nash equilibria of large games," *Games and Economic Behavior*, vol. 74, pp. 418-430, 2012.

[14] E. Cartwright and M. Wooders, "On purification of equilibrium in Bayesian games and expost Nash equilibrium," *International Journal of Game Theory*, vol. 38, pp. 127-136, 2009.

[15] A. Jadbabaie, P. Molavi, A. Sandroni and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, pp. 210-225, 2012.

[16] Q. Zhu, H. Tebine and T. Basar, "Heterogeneus learning in zero-sum stochastic games with incomplete information," presented at the 49th IEEE Conference on Decision and Control, Atlanta, USA, 2010.

[17] P. S. Sastry, V.V. Phansalkar and M. A. L. Thathachar, "Decentralized learning of Nash equilibria in multi-person stochastic games with incomplete information," *IEEE Transactions on Systems, Man and Cybernetycs*, vol. 24, No. 5, pp. 769-777, 1994.

[18] P. M. Djuric and Y. Wang, "Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations," *IEEE Signal Processing Magazine*, vol. 29, No. 2, pp. 65-76, 2012.

[19] M. Caramia and P. Dell'Olmo, "Multi-objective optimization," in *Multi-objective Management in Freight Logistics. Increasing capacity, service level and safety with optimization algorithms.* London: Springer-Verlag, 2008.

[20] R. Song, W. Xiao and H. Zhang, "Multi-objective optimal control for a class of unknown nonlinear systems based on finite-approximation-error ADP algorithm," *Neurocomputing*, vol. 119, pp. 212–221, 2013.