

# Connections Between Cooperative Control and Potential Games Illustrated on the Consensus Problem

Jason R. Marden, Gürdal Arslan and Jeff S. Shamma

**Abstract**—This paper presents a view of cooperative control using the language of learning in games. We review the game theoretic concepts of potential games and weakly acyclic games and demonstrate how the specific cooperative control problem of consensus can be formulated in these settings. Motivated by this connection, we build upon game theoretic concepts to better accommodate a broader class of cooperative control problems. In particular, we introduce sometimes weakly acyclic games for time-varying objective functions and action sets, and provide distributed algorithms for convergence to an equilibrium. Finally, we illustrate how to implement these algorithms for the consensus problem in a variety of settings, most notably, in an environment with non-convex obstructions.

## I. INTRODUCTION

Our goals in this paper are to establish a relationship between cooperative control problems, such as the consensus problem, and game theoretic methods, and to demonstrate the effectiveness of utilizing game theoretic approaches for controlling multi-agent systems. The results presented here are of independent interest in terms of their applicability to a large class of games. However, we will focus on the consensus problem as the main illustration of the approach.

We consider a discrete time version of the consensus problem initiated in [1] in which a group of players  $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_n\}$  seek to come to an agreement, or consensus, upon a common scalar value<sup>1</sup> by repeatedly interacting with one another. By reaching consensus, we mean converging to the agreement space characterized by

$$a_1 = a_2 = \dots = a_n,$$

where  $a_i$  is referred to as the state of player  $\mathcal{P}_i$ . Several papers study different interaction models and analyze the conditions under whether these interactions lead to consensus [2], [3], [4], [5], [6], [7], [8], [9].

A well studied protocol, referred to here as the “consensus algorithm”, can be described as follows. At each time step  $t \in \{0, 1, \dots\}$ , each player  $\mathcal{P}_i$  is allowed to interact with a group of other players, who are referred to as the *neighbors* of player  $\mathcal{P}_i$  and denoted as  $N_i(t)$ . During an interaction,

Research supported by ARO grant #W911NF0410316, AFOSR grant #FA9550-05-1-0239, and NSF grants #ECS-0501394 and #ECCS-0547692.

J. R. Marden (marden@seas.ucla.edu) and Jeff S. Shamma (shamma@ucla.edu) are with the Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095-1597.

G. Arslan (gurdal@hawaii.edu) is with the Department of Electrical Engineering, University of Hawaii at Manoa, 440 Holmes Hall, 2540 Dole Street, Honolulu, HI 96822.

<sup>1</sup>The forthcoming results will hold for multi-dimensional consensus as well.

each player  $\mathcal{P}_i$  is informed of the current (or possibly delayed) state of all his neighbors. Player  $\mathcal{P}_i$  then updates his state by forming a convex combination of his state along with the state of all his neighbors. The consensus algorithm takes on the general form

$$a_i(t+1) = \sum_{\mathcal{P}_j \in N_i(t)} \omega_{ij}(t) a_j(t), \quad (1)$$

where  $\omega_{ij}(t)$  is the relative weight that player  $\mathcal{P}_i$  places on the state of player  $\mathcal{P}_j$  at time  $t$ . The interaction topology is described in terms of a time varying directed graph  $G(V, E(t))$  with the set of nodes  $V = \mathcal{P}$  and the set of edges  $E(t) \subset \mathcal{P} \times \mathcal{P}$  at time  $t$ . The set of edges is directly related to the neighbor sets as follows:  $(\mathcal{P}_i, \mathcal{P}_j) \in E(t)$  if and only if  $\mathcal{P}_j \in N_i(t)$ . We will refer to  $G(V, E(t))$  as the interaction graph at time  $t$ .

There has been extensive research centered on understanding the conditions necessary for guaranteeing the convergence of all states, i.e.  $\lim_{t \rightarrow \infty} a_i(t) \rightarrow a^*$ , for all players  $\mathcal{P}_i \in \mathcal{P}$ . The convergence properties of the consensus algorithm have been studied under several interaction models encompassing delays in information exchange, connectivity issues, varying topologies and noisy measurements.

Surprisingly, there has been relatively little research that links cooperative control problems to a branch of the learning in games literature [10] that emphasizes coordination games. The goal of this paper is to better establish this link and to develop new algorithms for broader classes of cooperative control problems as well as games.

In section 2 we establish a connection between cooperative control problems and potential games. In section 3 we model the consensus problem as a potential game and present suitable learning algorithms that guarantee that players will come to a consensus even in an environment filled with non-convex obstructions. In section 4 we introduce a new class of games called sometimes weakly acyclic games, which generalize potential games, and present simple learning dynamics with desirable convergence properties. In section 5, we show that the consensus problem can be modeled as a sometime weakly acyclic game. Section 6 presents some final remarks.

## II. COOPERATIVE CONTROL PROBLEMS AND POTENTIAL GAMES

Cooperative control problems entail several autonomous players seeking to collectively accomplish a global objective. The consensus problem is one example of a cooperative control problem, where the global objective is for all players to reach consensus upon a given state. The challenge in

cooperative control problems is designing local control laws and/or local objective functions for each of the individual players so that collectively they accomplish the desired global objective.

One approach for cooperative control problems is to assign each individual player a fixed protocol or policy. This protocol specifies precisely what each player should do under any environmental condition. The consensus algorithm set forth in equation (1) is an example of such a policy based approach. One challenge in this approach is to incorporate dynamic or evolving constraints on player policies. For example, suppose a global planner desires a group of autonomous players to physically converge to a central location in an environment containing obstructions. The standard consensus algorithm may not be applicable to this problem since limitations in control capabilities caused by environmental obstructions are not considered. Variations of the consensus algorithm could possibly be designed to accommodate obstructions, but the analysis and control design would be more challenging.

An alternative, *game theoretic*, approach to cooperative control problems, and our main interest in this paper, is to assign each individual player a local objective function. In this setting, each player  $\mathcal{P}_i \in \mathcal{P}$  is assigned an action set  $\mathcal{A}_i$  and a local objective function  $U_i : \mathcal{A} \rightarrow \mathbb{R}$ , where  $\mathcal{A} = \prod_{\mathcal{P}_i \in \mathcal{P}} \mathcal{A}_i$  is the set of joint actions. An example of an objective function that will be studied in the following section is

$$U_i(a_i, a_{-i}) := - \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\|,$$

where  $\|\cdot\|$  is any norm,  $N_i$  is the neighbor set of player  $\mathcal{P}_i$ , and  $a_{-i} = \{a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n\}$  denotes the collection of actions of players other than player  $\mathcal{P}_i$ . With this notation, we will frequently express the joint action  $a$  as  $(a_i, a_{-i})$ .

We are interested in analyzing the long term behavior when players are repeatedly allowed to interact with one another in a competitive environment where each player seeks to selfishly maximize his own objective function. These interactions will be modeled as a repeated game, in which a one stage game is repeated each time step  $t \in \{0, 1, 2, \dots\}$ . At every time step  $t > 0$ , each player  $\mathcal{P}_i \in \mathcal{P}$  selects an action  $a_i \in \mathcal{A}_i$  seeking to myopically maximize his expected utility. Since a player's utility may be adversely affected by the actions of other players, the player can use his observations from the games played at times  $\{0, 1, \dots, t-1\}$  to develop a behavioral model of the other players.

At any time  $t > 0$ , the learning dynamics specify how any player  $\mathcal{P}_i$  processes past observations from the interactions at times  $\{0, 1, \dots, t-1\}$  to generate a model of the behavior of the other players. The learning dynamics that will be used throughout this paper are referred to as single stage memory dynamics which have a structural form similar to that of the consensus algorithm; namely, that the decision of any player  $\mathcal{P}_i$  at time  $t$  is made using only observations from the game played at time  $t-1$ . The learning dynamics need not be restricted to single stage memory. A follow up study could

analyze the benefit of using additional memory in learning dynamics for the consensus problem.

The challenge of the design in the game theoretic approach lies in designing the objective functions *and* the learning dynamics such that, when players selfishly pursue their own objectives, they also collectively accomplish the objective of the global planner. Suppose that the objective of the global planner is captured by a potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$ . In any successful multi-agent system each player's objective function should be appropriately "aligned" with the objective of the global planner. This notion of utility alignment in multi-agent systems has a strong connection to potential games [11].

*Definition 2.1 (Potential Games):* Player action sets  $\{\mathcal{A}_i\}_{i=1}^n$  together with player objective functions  $\{U_i : \mathcal{A} \rightarrow \mathbb{R}\}_{i=1}^n$  constitute a **potential game** if, for some potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$ ,

$$U_i(a_i'', a_{-i}) - U_i(a_i', a_{-i}) = \phi(a_i'', a_{-i}) - \phi(a_i', a_{-i}),$$

for every player  $\mathcal{P}_i \in \mathcal{P}$ , for every  $a_i', a_i'' \in \mathcal{A}_i$ , and for every  $a_{-i} \in \times_{j \neq i} \mathcal{A}_j$ .

A potential game, as defined above, requires perfect alignment between the global objective and the players' local objective functions, meaning that if a player unilaterally changed his action, the change in his objective function would be equal to the change in the potential function. There are weaker notions of potential games, called weakly acyclic games, which will be discussed later. The connection between cooperative control problems and potential games is important because learning algorithms for potential games have been studied extensively in the game theory literature [12], [11], [13], [14], [15]. Accordingly, if it is shown that a cooperative control problem can be modeled as a potential game, established learning algorithms with guaranteed asymptotic results could be used to tackle the cooperative control problem at hand.

In the following section we will illustrate this opportunity by showing that the consensus problem can be modeled as a potential game by defining players' utilities appropriately.

### III. CONSENSUS MODELED AS A POTENTIAL GAME

In this section we will formulate the consensus problem as a potential game. First, we establish a global objective function that captures the notion of consensus. Next, we show that local objective functions can be assigned to each player so that the resulting game is in fact a potential game. Finally, we present a learning algorithm that guarantees consensus even in an environment containing non-convex obstructions.

It turns out that the potential game formulation of the consensus problem discussed in this section requires the interaction graph to be time-invariant and undirected. In Section V we relax these requirements by formulating the consensus problem as a sometimes weakly acyclic game.

### A. Setup: Consensus Problem with a Time-Invariant and Undirected Interaction Graph

Consider a consensus problem with  $n$ -player set  $\mathcal{P}$  where each player  $\mathcal{P}_i \in \mathcal{P}$  has a finite action set  $\mathcal{A}_i$ . A player's action set could represent the finite set of locations that a player could select.

We will consider the following potential function for the consensus problem

$$\phi(a) := - \sum_{\mathcal{P}_i \in \mathcal{P}} \sum_{\mathcal{P}_j \in N_i} \frac{\|a_i - a_j\|}{2}, \quad (2)$$

where  $N_i \subset \mathcal{P}$  is player  $\mathcal{P}_i$ 's time-invariant neighbor set. In the case where the interaction graph induced by the neighbor sets  $\{N_i\}_{i=1}^n$  is connected<sup>2</sup>, the potential function above achieves the value of 0 if and only if the action profile  $a \in \mathcal{A}$  constitutes a *consensus*, i.e.,

$$\phi(a) = 0 \Leftrightarrow a_1 = \dots = a_n.$$

The goal is to assign each player an objective function that it is perfectly aligned with the global objective in (2). One approach would be to assign each player the following objective function:

$$U_i(a) = \phi(a).$$

This assignment means that a player's objective function is dependent on the actions of all players. This structure relies on global information which may be unavailable to the player. An alternative approach would be to assign each player an objective function that captures the player's marginal contribution to the potential function. For the consensus problem, this translates to each player being assigned the objective function

$$U_i(a_i, a_{-i}) := - \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\|. \quad (3)$$

Now, each player's objective function is *only* dependent on the actions of his neighbors. These objective functions are referred to as a Wonderful Life Utility; see [16], [17]. It is known that assigning each agent a Wonderful Life Utility leads to a potential game [16], [17]; however, we will explicitly show this for the consensus problem in the following claim.

*Claim 3.1: Player objective functions (3) constitute a potential game with the potential function (2) provided that the time-invariant interaction graph induced by the neighbor sets  $\{N_i\}_{i=1}^n$  is undirected, i.e.,*

$$\mathcal{P}_j \in N_i \Leftrightarrow \mathcal{P}_i \in N_j.$$

*Proof:* Since the interaction graph is time-invariant and undirected, the potential function can be expressed as

$$\phi(a) = - \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\| - \sum_{\mathcal{P}_j \neq \mathcal{P}_i} \sum_{\mathcal{P}_k \in N_j \setminus \mathcal{P}_i} \frac{\|a_j - a_k\|}{2}.$$

<sup>2</sup>A graph is connected if there exists a path from any node to any other node.

The change in objective of player  $\mathcal{P}_i$  by switching from action  $a_i^1$  to action  $a_i^2$  provided that all other players collectively play  $a_{-i}$  is

$$\begin{aligned} U_i(a_i^2, a_{-i}) - U_i(a_i^1, a_{-i}) &= \sum_{\mathcal{P}_j \in N_i} -\|a_i^2 - a_j\| + \|a_i^1 - a_j\|, \\ &= \phi(a_i^2, a_{-i}) - \phi(a_i^1, a_{-i}). \end{aligned}$$

Note that the above claim does not require the interaction graph to be connected. There may exist other potential functions and subsequent player objective functions that can accommodate more general setups. For a detailed discussion on possible player objective functions derived from a given potential function, see [16].

We now assume that the above game is repeatedly played at discrete time steps  $t \in \{0, 1, 2, \dots\}$ . We are interested in determining the limiting behavior of the players, in particular whether or not they reach a consensus, under various interaction models. Since the consensus problem is modeled as a potential game, there are a large number of learning algorithms available with guaranteed results [10], [18], [16], [11], [14], [15]. Most of the learning algorithms for potential games guarantee that the player behavior converge to a Nash equilibrium, defined as follows:

*Definition 3.1 (Nash Equilibrium): An action profile  $a^* \in \mathcal{A}$  is a Nash equilibrium if for any player  $\mathcal{P}_i \in \mathcal{P}$  and any action  $a_i \in \mathcal{A}_i$ ,  $U_i(a_i^*, a_{-i}^*) \geq U_i(a_i, a_{-i}^*)$ .*

It is straightforward to see that any consensus point is a Nash equilibrium of the game characterized by the player objective functions (3). This is because a consensus point maximizes the potential function as well as the player objective functions (3). However, the converse statement is not true. Let  $\mathcal{A}^*$  denote the set of Nash equilibria and  $\mathcal{A}^c$  denote the set of consensus points. We know that  $\mathcal{A}^c \subset \mathcal{A}^*$  where the inclusion can be proper. In other words, a Nash equilibrium, say  $a^* \in \mathcal{A}^*$ , can be suboptimal, i.e.,  $\phi(a^*) < 0$ , and hence fail to be a consensus point.

### B. A Learning Algorithm for Potential Games with Suboptimal Nash Equilibria

Before stating the learning algorithm, we start with some notation. Let the strategy of player  $\mathcal{P}_i$  at time  $t$  be denoted by the probability distribution  $p_i(t) \in \Delta(\mathcal{A}_i)$  where  $\Delta(\mathcal{A}_i)$  denotes the set of probability distributions over the set  $\mathcal{A}_i$ . Using this strategy, player  $\mathcal{P}_i$  randomly selects an action from  $\mathcal{A}_i$  at time  $t$  according to  $p_i(t)$ .

Consider the following learning algorithm known as spatial adaptive play (SAP) [10]. At each time  $t > 0$ , one player  $\mathcal{P}_i \in \mathcal{P}$  is randomly chosen (with equal probability for each player) and allowed to update his action. All other players must repeat their actions, i.e.  $a_{-i}(t) = a_{-i}(t-1)$ . At time  $t$ , the updating player  $\mathcal{P}_i$  randomly selects an action from  $\mathcal{A}_i$  according to his strategy  $p_i(t) \in \Delta(\mathcal{A}_i)$  where the  $a_i$ -th component  $p_i^{a_i}(t)$  of his strategy is given as

$$p_i^{a_i}(t) = \frac{\exp\{\beta U_i(a_i, a_{-i}(t-1))\}}{\sum_{\bar{a}_i \in \mathcal{A}_i} \exp\{\beta U_i(\bar{a}_i, a_{-i}(t-1))\}},$$

for some exploration parameter  $\beta \geq 0$ . The constant  $\beta$  determines how likely player  $\mathcal{P}_i$  is to select a suboptimal action. If  $\beta = 0$ , player  $\mathcal{P}_i$  will select any action  $a_i \in \mathcal{A}_i$  with equal probability. As  $\beta \rightarrow \infty$ , player  $\mathcal{P}_i$  will select an action from his best response set

$$\{a_i \in \mathcal{A}_i : U_i(a_i, a_{-i}(t-1)) = \max_{a'_i \in \mathcal{A}_i} U_i(a'_i, a_{-i}(t-1))\}$$

with arbitrarily high probability.

In a repeated potential game in which all players adhere to SAP, the stationary distribution  $\mu \in \Delta(\mathcal{A})$  of the joint action profiles is given in [10] as

$$\mu(a) = \frac{\exp\{\beta \phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta \phi(\bar{a})\}}.$$

One can interpret the stationary distribution  $\mu$  as follows: for sufficiently large times  $t > 0$ ,  $\mu(a)$  equals the probability that  $a(t) = a$ . As  $\beta \uparrow \infty$ , all the weight of the stationary distribution  $\mu$  is on the joint actions that maximize the potential function. In the potential game formulation of the consensus problem, the joint actions that maximize the potential function (2) are precisely the consensus points provided that the interaction graph is connected. Therefore, if all players update their actions using the learning algorithm SAP with sufficiently large  $\beta$ , then the players will reach a consensus asymptotically with arbitrarily high probability.

### C. A Learning Algorithm for Potential Games with Suboptimal Nash Equilibria and Restricted Action Sets

One issue with the applicability of the learning algorithm SAP for the consensus problem is that it permits any player to select any action in his action set. Because of player mobility limitations, this may not be possible. For example, a player may only be able to move to a position within a fixed radius of his current position. Therefore, we seek to modify SAP by conditioning a player's action set on his previous action. Let  $a(t-1)$  be the joint action at time  $t-1$ . With restricted action sets, the set of actions available to player  $\mathcal{P}_i$  at time  $t$  is a function of his action at time  $t-1$  and will be denoted as  $R_i(a_i(t-1)) \subset \mathcal{A}_i$ . We will adopt the convention that  $a_i \in R_i(a_i)$  for any action  $a_i \in \mathcal{A}_i$ , i.e., a player is always allowed to stay with his previous action.

We will introduce a variant of SAP called binary restrictive spatial adaptive play (RSAP) to accommodate the notion of restricted action sets. RSAP can be described as follows: At each time step  $t > 0$ , one player  $\mathcal{P}_i \in \mathcal{P}$  is randomly chosen (with equal probability for each player) and allowed to update his action. All other players must repeat their actions, i.e.  $a_{-i}(t) = a_{-i}(t-1)$ . At time  $t$ , the updating player  $\mathcal{P}_i$  selects one trial action  $\hat{a}_i$  randomly from his allowable set  $R_i(a_i(t-1))$  with the following probability:

- $\Pr[\hat{a}_i = a_i] = \frac{1}{N_i}$  for any  $a_i \in R_i(a_i(t-1)) \setminus a_i(t-1)$ ,
- $\Pr[\hat{a}_i = a_i(t-1)] = 1 - \frac{|R_i(a_i(t-1))| - 1}{N_i}$ ,

where  $N_i$  denotes the maximum number of actions in any restricted action set for player  $\mathcal{P}_i$ , i.e.,  $N_i := \max_{a_i \in \mathcal{A}_i} |R_i(a_i)|$ . After player  $\mathcal{P}_i$  selects a trial action  $\hat{a}_i$ ,

he updates his action to  $\hat{a}_i$ , i.e.,  $a_i(t) = \hat{a}_i$ , with probability

$$\frac{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\}}{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\} + \exp\{\beta U_i(a_i(t-1))\}},$$

and he stays with his previous action, i.e.,  $a_i(t) = a_i(t-1)$ , with probability

$$\frac{\exp\{\beta U_i(a_i(t-1))\}}{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\} + \exp\{\beta U_i(a_i(t-1))\}},$$

where  $\beta \geq 0$  is an exploration parameter. Note that if  $\hat{a}_i$  is selected as  $a_i(t-1)$  then  $\Pr[a_i(t) = a_i(t-1)] = 1$ .

We make the following assumptions regarding the restricted action sets.

*Assumption 3.1 (Reversibility):* For any player  $\mathcal{P}_i \in \mathcal{P}$  and any action pair  $a_i^1, a_i^2 \in \mathcal{A}_i$ ,

$$a_i^2 \in R_i(a_i^1) \Leftrightarrow a_i^1 \in R_i(a_i^2).$$

*Assumption 3.2 (Feasibility):* For any player  $\mathcal{P}_i \in \mathcal{P}$  and any action pair  $a_i^0, a_i^n \in \mathcal{A}_i$ , there exists a sequence of actions  $a_i^0 \rightarrow a_i^1 \rightarrow \dots \rightarrow a_i^n$  satisfying  $a_i^k \in R_i(a_i^{k-1})$  for all  $k \in \{1, 2, \dots, n\}$ .

*Theorem 3.1:* Consider a finite  $n$ -player potential game with potential function  $\phi(\cdot)$ . If the restricted action sets satisfy Assumptions 3.1 and 3.2, then RSAP induces a Markov process over the state space  $\mathcal{A}$  where the unique stationary distribution  $\mu \in \Delta(\mathcal{A})$  is given as

$$\mu(a) = \frac{\exp\{\beta \phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta \phi(\bar{a})\}}, \text{ for any } a \in \mathcal{A}. \quad (4)$$

*Proof:* The proof follows along the lines of the proof of Theorem 6.2 in [10]. By Assumptions 3.1 and 3.2 we know that the Markov process induced by RSAP is irreducible and aperiodic; therefore, the process has a unique stationary distribution. Below, we show that this unique distribution must be (4) by verifying that the distribution (4) satisfies the detailed balanced equations

$$\mu(a)P_{ab} = \mu(b)P_{ba},$$

for any  $a, b \in \mathcal{A}$ , where

$$P_{ab} := \Pr[a(t) = b | a(t-1) = a].$$

Note that the only nontrivial case is the one where  $a$  and  $b$  differ by exactly one player  $\mathcal{P}_i$ , that is,  $a_{-i} = b_{-i}$  but  $a_i \neq b_i$  where  $a_i \in R_i(b_i)$  which also implies that  $b_i \in R_i(a_i)$ . Since player  $\mathcal{P}_i$  has probability  $1/n$  of being chosen in any given period and any trial action  $b_i \in R_i(a_i)$ ,  $b_i \neq a_i$ , has probability of  $1/N_i$  of being chosen, it follows that

$$\mu(a)P_{ab} = \left[ \frac{\exp\{\beta \phi(a)\}}{\sum_{z \in \mathcal{A}} \exp\{\beta \phi(z)\}} \right] \times \left[ (1/n)(1/N_i) \frac{\exp\{\beta U_i(b)\}}{\exp\{\beta U_i(a)\} + \exp\{\beta U_i(b)\}} \right].$$

Letting

$$\lambda = \left( \frac{1}{\sum_{z \in \mathcal{A}} \exp\{\beta \phi(z)\}} \right) \times \left( \frac{(1/n)(1/N_i)}{\exp\{\beta U_i(a)\} + \exp\{\beta U_i(b)\}} \right),$$

we obtain

$$\mu(a)P_{ab} = \lambda \exp\{\beta\phi(a) + \beta U_i(b)\}.$$

Since  $U_i(b) - U_i(a) = \phi(b) - \phi(a)$ , we have

$$\mu(a)P_{ab} = \lambda \exp\{\beta\phi(b) + \beta U_i(a)\},$$

which leads us to

$$\mu(a)P_{ab} = \mu(b)P_{ba}.$$

Note that if all players adhere to the learning dynamics RSAP in a consensus problem where the interaction graph is time-invariant and undirected, the restricted action sets satisfy Assumptions 3.1 and 3.2, and players are assigned the utilities (3), then, at sufficiently large times  $t$ , the players' collective behavior will maximize the potential function (2) with arbitrarily high probability provided that  $\beta$  is sufficiently large. Furthermore, if the interaction graph is connected and consensus is possible, meaning  $(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_n) \neq \emptyset$ , then, at sufficiently large times  $t > 0$ , the players' actions will constitute a consensus with arbitrarily high probability even in an environment filled with non-convex obstructions.

#### D. Example: Consensus in an Environment with Non-Convex Obstructions

Consider the 2-D consensus problem with player set  $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4\}$ . Each player  $\mathcal{P}_i$  has an action set  $\mathcal{A}_i = \{1, 2, \dots, 10\} \times \{1, 2, \dots, 10\}$  as illustrated in Figure 1. The

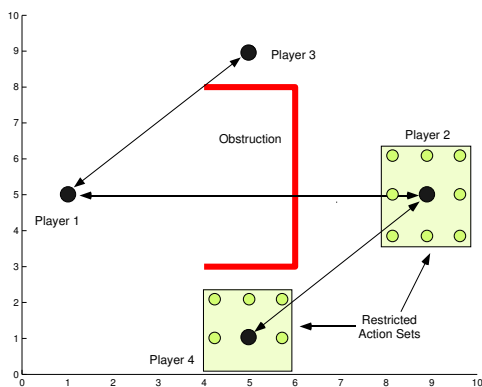


Fig. 1. A consensus problem setup.

arrows represent the time-invariant and undirected edges of the connected interaction graph. The restricted action sets are highlighted for players  $\mathcal{P}_2$  and  $\mathcal{P}_4$ . At any given time, any player can have at most 9 possible actions; therefore,  $N_i = 9$  for all players  $\mathcal{P}_i \in \mathcal{P}$ .

We simulated RSAP on the consensus problem with the interaction graph, environmental obstruction, and the initial conditions shown in Figure 1. We increase the exploration parameter  $\beta$  as  $t/200$  during player interactions. The complete action path of all players reaching a consensus is shown in Figure 2.

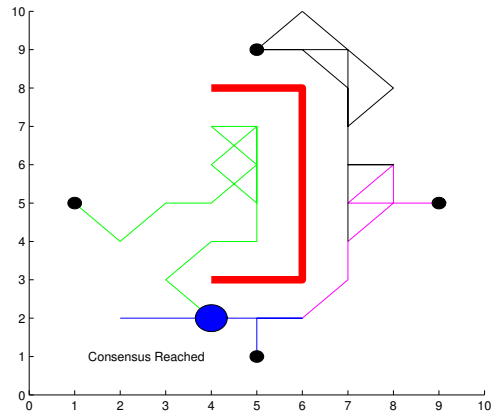


Fig. 2. Evolution of player actions during player interactions.

## IV. WEAKLY ACYCLIC AND SOMETIMES WEAKLY ACYCLIC GAMES

In potential games, players' objective functions must be perfectly aligned with the potential of the game. In the potential game formulation of the consensus problem, this alignment condition required that the interaction graph be time-invariant and undirected. In this section we will seek to relax this alignment requirement by allowing player objective functions to be "somewhat" aligned with the potential of the game. We will review a weaker form of potential games called weakly acyclic games and introduce a new class of games called *sometimes weakly acyclic games*. We will also present simple learning dynamics that guarantee convergence to a *universal Nash equilibrium*, to be defined later, in any sometimes weakly acyclic game.

### A. Weakly Acyclic Games

The main difference between potential games and weakly acyclic games lies in the degree to which player objective functions are aligned with the potential function. Consider any game with a finite set  $\mathcal{A}$  of action profiles. A *better reply path* is a sequence of action profiles  $a^1, a^2, \dots, a^L$  such that, for every  $1 \leq \ell \leq L-1$ , there is exactly one player  $\mathcal{P}_{i_\ell}$  such that i)  $a_{i_\ell}^\ell \neq a_{i_\ell}^{\ell+1}$ , ii)  $a_{-i_\ell}^\ell = a_{-i_\ell}^{\ell+1}$ , and iii)  $U_{i_\ell}(a^\ell) < U_{i_\ell}(a^{\ell+1})$ . In other words, one player moves at a time, and each time a player moves he increases his own objective function.

Consider now a potential game with potential function  $\phi$ . Starting from an arbitrary action profile  $a \in \mathcal{A}$ , construct a better reply path  $a = a^1, a^2, \dots, a^L$  until it can no longer be extended. Note first that such a path cannot cycle back on itself, because  $\phi$  is strictly increasing along the path. Since  $\mathcal{A}$  is finite, the path cannot be extended indefinitely. Hence, the last element in a maximal better reply path from any joint action,  $a$ , must be a Nash equilibrium.

This idea may be generalized as follows. A game is *weakly acyclic* if, for any  $a \in \mathcal{A}$ , there exists a better reply path starting at  $a$  and ending at some Nash equilibrium [10], [18]. Potential games are special cases of weakly acyclic games.

A weakly acyclic game can also be defined in terms of a potential function.

*Lemma 4.1:* A game is weakly acyclic if and only if there exists a potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$  such that for any action  $a \in \mathcal{A}$  that is *not* a Nash equilibrium, there exists a player  $\mathcal{P}_i \in \mathcal{P}$  with an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i(a_i^*, a_{-i}) > U_i(a_i, a_{-i})$  and  $\phi(a_i^*, a_{-i}) > \phi(a_i, a_{-i})$ .

*Proof:* ( $\Leftarrow$ ) Select any action  $a^0 \in \mathcal{A}$ . If  $a^0$  is not a Nash equilibrium, there exists a player  $\mathcal{P}_i \in \mathcal{P}$  with an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i(a_i^*) > U_i(a^0)$  and  $\phi(a_i^*) > \phi(a^0)$  where  $a^1 = (a_i^*, a_{-i}^0)$ .

Repeat this process and construct a path  $a^0, a^1, \dots, a^n$  until it can no longer be extended. Note first that such a path cannot cycle back on itself, because  $\phi$  is strictly increasing along the path. Since  $\mathcal{A}$  is finite, the path cannot be extended indefinitely. Hence, the last element in this path must be a Nash equilibrium.

( $\Rightarrow$ ) We will construct a potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$  recursively. Select any action  $a^0 \in \mathcal{A}$ . Since the game is weakly acyclic, there exists a better reply path  $a^0, a^1, \dots, a^n$  where  $a^n$  is a Nash equilibrium. Let  $\mathcal{A}^0 = \{a^0, a^1, \dots, a^n\}$ . Define the (finite) potential function  $\phi$  over the set  $\mathcal{A}^0$  satisfying the following conditions:

$$\phi(a^0) < \phi(a^1) < \dots < \phi(a^n).$$

Now select any action  $\tilde{a}^0 \in \mathcal{A} \setminus \mathcal{A}^0$ . There exists a better reply path  $\tilde{a}^0, \tilde{a}^1, \dots, \tilde{a}^m$  where  $\tilde{a}^m$  is a Nash equilibrium. Let  $\mathcal{A}^1 = \{\tilde{a}^0, \tilde{a}^1, \dots, \tilde{a}^m\}$ . If  $\mathcal{A}^1 \cap \mathcal{A}^0 = \emptyset$  then define the potential function  $\phi$  over the set  $\mathcal{A}^1$  satisfying the following conditions:

$$\phi(\tilde{a}^0) < \phi(\tilde{a}^1) < \dots < \phi(\tilde{a}^m).$$

If  $\mathcal{A}^1 \cap \mathcal{A}^0 \neq \emptyset$ , then let  $k^* = \min\{k \in \{1, 2, \dots, m\} : \tilde{a}^k \in \mathcal{A}^0\}$ . Define the potential function  $\phi$  over the truncated (redefined) set  $\mathcal{A}^1 = \{\tilde{a}^0, \tilde{a}^1, \dots, \tilde{a}^{k^*-1}\}$  satisfying the following conditions:

$$\phi(\tilde{a}^0) < \phi(\tilde{a}^1) < \dots < \phi(\tilde{a}^{k^*}).$$

Now select any action  $\hat{a}^0 \in \mathcal{A} \setminus (\mathcal{A}^0 \cup \mathcal{A}^1)$  and repeat until no such action exists.

The construction of the potential function  $\phi$  guarantees that for any action  $a \in \mathcal{A}$  that is not a Nash equilibrium, there exists a player  $\mathcal{P}_i \in \mathcal{P}$  with an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i(a_i^*, a_{-i}) > U_i(a_i, a_{-i})$  and  $\phi(a_i^*, a_{-i}) > \phi(a_i, a_{-i})$ . ■

## B. Learning Dynamics for Weakly Acyclic Games

We will consider the *better reply with inertia dynamics* for weakly acyclic games analyzed in [19], [18]. Before stating the learning dynamics, we define a player's *strict better reply set* for any action profile  $a^0 \in \mathcal{A}$  as

$$B_i(a^0) := \{a_i \in \mathcal{A}_i : U_i(a_i, a_{-i}^0) > U_i(a^0)\}.$$

The better reply with inertia dynamics can be described as follows. At each time  $t > 0$ , each player  $\mathcal{P}_i$  presumes that all other players will continue to play their previous actions  $a_{-i}(t-1)$ . Under this presumption, each player  $\mathcal{P}_i \in \mathcal{P}$

selects an action according to the following strategy at time  $t$ :

$$\begin{aligned} B_i(a(t-1)) = \emptyset &\Rightarrow a_i(t) = a_i(t-1), \\ B_i(a(t-1)) \neq \emptyset &\Rightarrow \Pr[a_i(t) = a_i(t-1)] = \alpha(t), \\ &\Pr[a_i(t) = a_i^*] = \frac{1-\alpha(t)}{|B_i(a(t-1))|}, \end{aligned}$$

for any action  $a_i^* \in B_i(a(t-1))$  where  $\alpha(t) \in (0, 1)$  is referred to as the player's inertia at time  $t$ . According to these rules, player  $\mathcal{P}_i$  will stay with the previous action  $a_i(t-1)$  with probability  $\alpha(t)$  even when there is a perceived opportunity for improvement. We make the following standing assumption on the players' willingness to optimize.

*Assumption 4.1:* There exist constants  $\underline{\epsilon}$  and  $\bar{\epsilon}$  such that for all time  $t \geq 0$  and for all players  $\mathcal{P}_i \in \mathcal{P}$ ,

$$0 < \underline{\epsilon} < \alpha_i(t) < \bar{\epsilon} < 1.$$

This assumption implies that players are always willing to optimize with some nonzero inertia.

If all players adhere to the better reply with inertia dynamics satisfying Assumption 4.1, then the joint action profiles will converge to a Nash equilibrium almost surely in any weakly acyclic game [19], [18].

## C. Sometimes Weakly Acyclic Games

In the potential game formulation of the consensus problem, each player was assigned a time-invariant objective function of the form (3). However, in the case of a time-varying interaction topology, we would like to allow player objective function to be time-varying. In this framework, each player  $\mathcal{P}_i$  is now assigned a local objective function  $U_i : \mathcal{A} \times \{0, 1, 2, \dots\} \rightarrow \mathbb{R}$ . We will denote the objective function of player  $\mathcal{P}_i$  at time  $t$  as  $U_i(a(t), t)$  where  $a(t)$  is the action profile at time  $t$ .

We will call an action profile  $a^*$  a *universal Nash equilibrium* if

$$U_i(a^*, t) = \max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}^*, t)$$

for all times  $t \geq 0$ .

We will call a game *sometimes weakly acyclic* if there exists a potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$  and a finite time constant  $T$  such that for any time  $t_0 > 0$  and any action profile  $a^0$  that is *not* a universal Nash equilibrium, there exists a time  $t_1 \in [t_0, t_0 + T]$ , a player  $\mathcal{P}_i \in \mathcal{P}$ , and an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i(a_i^*, a_{-i}^0, t_1) > U_i(a^0, t_1)$  and  $\phi(a_i^*, a_{-i}^0) > \phi(a^0)$ .

## D. Learning Dynamics for Sometimes Weakly Acyclic Games

We will consider the better reply with inertia dynamics for games involving time-varying objective functions. Before stating the learning dynamics, we redefine a player's *strict better reply set* for any action profile  $a^0 \in \mathcal{A}$  and time  $t > 0$  as

$$B_i(a^0, t) := \{a_i \in \mathcal{A}_i : U_i(a_i, a_{-i}^0, t) > U_i(a^0, t)\}.$$

The better reply with inertia dynamics can be described as follows. At each time  $t > 0$ , each player  $\mathcal{P}_i$  presumes that all other players will continue to play their previous actions  $a_{-i}(t-1)$ . Under this presumption, each player  $\mathcal{P}_i \in \mathcal{P}$

selects an action according to the following strategy at time  $t$ :

$$\begin{aligned} B_i(a(t-1), t) = \emptyset &\Rightarrow a_i(t) = a_i(t-1), \\ B_i(a(t-1), t) \neq \emptyset &\Rightarrow \Pr[a_i(t) = a_i(t-1)] = \alpha(t), \\ &\Pr[a_i(t) = a_i^*] = \frac{(1-\alpha(t))}{|B_i(a(t-1), t)|}, \end{aligned}$$

for any action  $a_i^* \in B_i(a(t-1), t)$  where  $\alpha(t) \in (0, 1)$  is the player's inertia at time  $t$ .

*Theorem 4.1:* Consider an  $n$ -player sometimes weakly acyclic game with finite action sets. If all players adhere to the better reply with inertia dynamics satisfying Assumption 4.1, then the joint action profiles will converge to a universal Nash equilibrium almost surely.

*Proof:* Let  $\phi : \mathcal{A} \rightarrow \mathbb{R}$  and  $T$  be the potential function and time constant for the sometimes weakly acyclic game. Let  $a(t_0) = a^0$  be the action profile at time  $t_0$ . If  $a^0$  is a universal Nash equilibrium, then  $a(t) = a^0$  for all times  $t \geq t_0$  and we are done. Otherwise, there exists a time  $t_1$  satisfying  $(t_0 + T) \geq t_1 > t_0$ , a player  $\mathcal{P}_i \in \mathcal{P}$ , and an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i((a_i^*, a_{-i}^0), t_1) > U_i(a^0, t_1)$  and  $\phi(a_i^*, a_{-i}^0) > \phi(a^0)$ . Because of players' inertia, the action  $a^1 = (a_i^*, a_{-i}^0)$  will be played at time  $t_1$  with at least probability  $\underline{\epsilon}^{n-1} \frac{(1-\bar{\epsilon})}{|\mathcal{A}|} \underline{\epsilon}^{nT}$ .

One can repeat this argument to show that for any time  $t_0 > 0$  and any action profile  $a(t_0)$  there exists a joint action  $a^*$  such that

$$\Pr[a(t) = a^*, \forall t \geq t^*] \geq \epsilon^*$$

where

$$\begin{aligned} t^* &= t_0 + T |\mathcal{A}|, \\ \epsilon^* &= \left( \underline{\epsilon}^{n-1} \frac{(1-\bar{\epsilon})}{|\mathcal{A}|} \underline{\epsilon}^{nT} \right)^{|\mathcal{A}|}. \end{aligned}$$

## V. CONSENSUS MODELED AS A SOMETIMES WEAKLY ACYCLIC GAME

Two main problems arose in the potential game formulation of the consensus problem. The first problem was that a Nash equilibrium was not necessarily a consensus point even when the interaction graph was connected and the environment was obstruction free. Therefore, we needed to employ a stochastic learning algorithm like SAP or RSAP to guarantee that the collective behavior of the players would be a consensus point with arbitrarily high probability. SAP or RSAP led to consensus by introducing noise into the decision making process, meaning that a player would occasionally make a suboptimal choice. The second problem was that the interaction graph needed to be time-invariant, undirected, and connected in order to guarantee consensus.

In this section, we will illustrate that, by modeling the consensus problem as a sometimes weakly acyclic game, one can effectively alleviate both problems. For brevity, we will show that the 1-dimensional consensus problem with appropriately designed player objective functions is a sometimes weakly acyclic game. However, one can easily extend this to the multi-dimensional case.

### A. Setup: Consensus Problem with a Time-Varying and Directed Interaction Graph

Consider a consensus problem with  $n$ -player set  $\mathcal{P}$  and a time-varying and directed interaction graph. Each player has a finite action set  $\mathcal{A}_i \subset \mathbb{R}$  and without loss of generalities, we will assume that  $\mathcal{A}_1 = \mathcal{A}_2 = \dots = \mathcal{A}_n$ . Each player  $\mathcal{P}_i \in \mathcal{P}$  is assigned an objective function  $U_i : \mathcal{A} \times \{0, 1, 2, \dots\} \rightarrow \mathbb{R}$ .

Before introducing the player objective functions, we define the following measure and extreme player sets:

$$\begin{aligned} D(a, \mathcal{P}') &:= \max_{\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}'} (a_i - a_j), \\ \mathcal{P}^u(a) &:= \{\mathcal{P}_i \in \mathcal{P} : a_i = \max_{\mathcal{P}_j \in \mathcal{P}} a_j\}, \\ \mathcal{P}^l(a) &:= \{\mathcal{P}_i \in \mathcal{P} : a_i = \min_{\mathcal{P}_j \in \mathcal{P}} a_j\}, \\ \underline{n}(a) &:= \min\{|\mathcal{P}^u(a)|, |\mathcal{P}^l(a)|\}. \end{aligned} \quad (5)$$

We also define the constant  $\delta_A > 0$  as follows. For any  $a^1, a^2 \in \mathcal{A}$  and any player sets  $\mathcal{P}^1, \mathcal{P}^2 \subset \mathcal{P}$  such that  $D(a^1, \mathcal{P}^1) \neq D(a^2, \mathcal{P}^2)$ , the following inequality is satisfied:

$$|D(a^1, \mathcal{P}^1) - D(a^2, \mathcal{P}^2)| > \delta_A.$$

Consider now the following potential function  $\phi : \mathcal{A} \rightarrow \mathbb{R}$

$$\phi(a) = -D(a, \mathcal{P}) + \delta_A(1 - \underline{n}(a)/n). \quad (6)$$

Note that the potential function is a non-positive function that achieves the value of 0 if and only if the action profile constitutes a consensus. Furthermore, note that the potential function is independent of the interaction topology.

Rather than specifying a particular objective functions as in (3), we will introduce a class of admissible objective functions. To that end, we define the set of *reasonable actions* for player  $\mathcal{P}_i$  at time  $t$  given any action profile  $a^0 \in \mathcal{A}$  as

$$S_i(a^0, t) := \{a_i \in \mathcal{A}_i : \max_{\mathcal{P}_j \in N_i(t)} a_j^0 \geq a_i \geq \min_{\mathcal{P}_k \in N_i(t)} a_k^0\}.$$

Note that

$$a_i \in S_i(a^0, t) \Rightarrow D(a_i, a_{-i}^0, N_i(t)) \leq D(a^0, N_i(t)).$$

We will define a general class of *reasonable objective functions*. An objective function for player  $\mathcal{P}_i$  is called a reasonable objective function if, for any time  $t > 0$ , and any action profile  $a \in \mathcal{A}$ , the better response set satisfies the following two conditions:

- 1)  $B_i(a, t) \subset \{S_i(a, t), \emptyset\}$ ,
- 2)  $|S_i(a, t)| > 1 \Rightarrow B_i(a, t) \neq \emptyset$ .

Roughly speaking, these conditions ensure that a player will not value moving further away from his belief about the location of his neighbors.

We will now relax our requirements on the connectivity and time-invariance of the interaction graph in the consensus problem. We make the following assumption on graph connectivity:

*Assumption 5.1 (Connectedness Over Intervals):* There exists a constant  $T > 0$  such that for any time

$t > 0$ , the interaction graph with nodes  $\mathcal{P}$  and edges  $E = E(t) \cup \dots \cup E(t+T)$  is connected.

*Claim 5.1: Reasonable objective functions introduced above constitute a sometimes weakly acyclic game with the potential function (6) provided that the interaction graph satisfies Assumption 5.1. Furthermore, every universal Nash equilibrium constitutes consensus.*

*Proof:* It is easy to see that any consensus point is a universal Nash equilibrium. We will show that if an action profile is not a consensus point, then there exists a player who can increase his objective function as well as the potential function at some time in a fixed time window. This implies that every universal Nash equilibrium is a consensus point and furthermore that the game is sometime weakly acyclic.

Let  $a^0 = a(t_0)$  be any joint action that is not a consensus point. We will show that for some time  $t_1 \in [t_0, t_0+T]$  there exists a player  $\mathcal{P}_i \in \mathcal{P}$  with an action  $a_i^* \in \mathcal{A}_i$  such that  $U_i(a_i^*, a_{-i}^0, t_1) > U_i(a^0, t_1)$  and  $\phi(a_i^*, a_{-i}^0) > \phi(a^0)$ . To see this let  $\mathcal{P}^*$  be the extreme player set with the least number of players, i.e.,  $\mathcal{P}^* = \mathcal{P}^u(a^0)$  if  $|\mathcal{P}^u(a^0)| \leq |\mathcal{P}^l(a^0)|$  or  $\mathcal{P}^* = \mathcal{P}^l(a^0)$  if  $|\mathcal{P}^u(a^0)| > |\mathcal{P}^l(a^0)|$ . Since the interaction graph is connected over intervals, for some  $t_1 \in [t_0, t_0+T]$  there exists at least one player  $\mathcal{P}_i \in \mathcal{P}^*$  with a neighbor  $\mathcal{P}_j \in N_i(t_1)$  such that  $\mathcal{P}_j \notin \mathcal{P}^*$ . Therefore,

$$|S_i(a^0, t_1)| > 1 \Rightarrow |B_i(a^0, t_1)| \neq \emptyset.$$

Let  $a_i^* \in B_i(a^0, t_1)$  and for notional convenience let  $a^1 = (a_i^*, a_{-i}^0)$ . We know that  $D(a^1, \mathcal{P}) \leq D(a^0, \mathcal{P})$ . If  $D(a^1, \mathcal{P}) < D(a^0, \mathcal{P})$ , then

$$\begin{aligned} \phi(a^1) &= -D(a^1, \mathcal{P}) + \delta_A(1 - \underline{n}(a^1)/n), \\ &> -D(a^0, \mathcal{P}) + \delta_A + \delta_A(1 - \underline{n}(a^1)/n), \\ &> -D(a^0, \mathcal{P}) + \delta_A + \delta_A(1 - (\underline{n}(a^0) + n)/n), \\ &= \phi(a^0). \end{aligned}$$

If  $D(a^1, \mathcal{P}) = D(a^0, \mathcal{P})$ , then

$$\begin{aligned} \phi(a^1) &= -D(a^0, \mathcal{P}) + \delta_A(1 - \underline{n}(a^1)/n), \\ &> -D(a^0, \mathcal{P}) + \delta_A(1 - (\underline{n}(a^1) + 1)/n), \\ &\geq -D(a^0, \mathcal{P}) + \delta_A(1 - \underline{n}(a^0)/n), \\ &= \phi(a^0). \end{aligned}$$

Therefore,  $a^0$  is not a universal Nash equilibrium. ■

If all players adhere to the better reply with inertia dynamics in a consensus problem where the interaction graph satisfies Assumption 5.1 and the players are assigned reasonable objective functions then the joint action profile will converge almost surely to a consensus.

These results can easily be extended to multi-dimensional consensus problem with bounded observational delays.

## VI. CONCLUSION

We have proposed a game theoretic approach to cooperative control by highlighting a connection between cooperative control problems and potential games. We introduced

a new class of games and enhanced existing learning algorithms to broaden the applicability of game theoretic methods in cooperative control setting. We demonstrated that one could successfully implement game theoretic methods on the cooperative control problem of consensus in a variety of settings. While the main example used was the consensus problem, the results in Theorems 3.1 and 4.1 and the notion of a sometimes weakly acyclic game is applicable to a broader class of games as well as other cooperative control problems.

## REFERENCES

- [1] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 35, no. 9, pp. 803–812, 1986.
- [2] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *IEEE Conf. on Decision and Control*, 2005.
- [3] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, 2004.
- [4] —, "A scheme for robust distributed sensor fusion based on average consensus," in *Information processing in sensor networks*, 2005.
- [5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. on Automatic Control*, vol. 49, no. 6, June 2003.
- [6] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," in *Proceedings of the IEEE*, January 2007, to appear.
- [7] L. Moreau, "Stability of continuous-time distributed consensus algorithms," in *43rd IEEE Conference on Decision and Control*, 2004, pp. 3998–4003.
- [8] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. on Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.
- [9] A. Kashyap, T. Basar, and R. Srikant, "Consensus with quantized information updates," in *45th IEEE Conference on Decision and Control*, 2006, pp. 2728–2733.
- [10] H. P. Young, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton University Press, 1998.
- [11] D. Monderer and L. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, pp. 124–143, 1996.
- [12] —, "Fictitious play property for games with identical interests," *Games and Economic Theory*, vol. 68, pp. 258–265, 1996.
- [13] D. Monderer and A. Sela, "Fictitious play and no-cycling conditions," 1997, <http://www.sfb504.uni-mannheim.de/publications/dp97-12.pdf>.
- [14] J. R. Marden, G. Arslan, and J. S. Shamma, "Regret based dynamics: Convergence in weakly acyclic games," in *Proceedings of the 2007 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, Honolulu, Hawaii, May 2007, to appear.
- [15] —, "Joint strategy fictitious play with inertia for potential games," in *Proceedings of the 44th IEEE Conference on Decision and Control*, December 2005, pp. 6692–6697, submitted to *IEEE Transactions on Automatic Control*.
- [16] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: a game theoretical formulation," *ASME Journal of Dynamic Systems, Measurement and Control*, 2007, to appear.
- [17] D. Wolpert and K. Tumor, "An overview of collective intelligence," in *Handbook of Agent Technology*, J. M. Bradshaw, Ed. AAAI Press/MIT Press, 1999.
- [18] H. P. Young, *Strategic Learning And Its Limits*. Oxford University Press, 2005.
- [19] —, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, 1993.