

Cooperative Control and Potential Games

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

Abstract—We present a view of cooperative control using the language of learning in games. We review the game-theoretic concepts of potential and weakly acyclic games, and demonstrate how several cooperative control problems, such as consensus and dynamic sensor coverage, 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 extend existing learning algorithms to accommodate restricted action sets caused by the limitations of agent capabilities and group-based decision making. Furthermore, we also introduce a new class of games called sometimes weakly acyclic games for time-varying objective functions and action sets, and provide distributed algorithms for convergence to an equilibrium.

Index Terms—Cooperative control, game theory, learning in games, multi-agent systems.

I. INTRODUCTION

THE GOALS of this paper are twofold: 1) to establish a relationship between cooperative control problems and game-theoretic methods, and demonstrate the effectiveness of utilizing game-theoretic approaches for controlling multiagent systems, and 2) motivated by this connection, to build upon existing game-theoretic results to better accommodate a broader class of cooperative control problems.

The results presented here are of independent interest in terms of their applicability to a large class of games. However, we will use the consensus problem as the main illustration of the approach. In a discrete-time version of the consensus problem initiated in [1], a group of players (or agents) $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_n\}$ seeks to come to an agreement, or consensus, upon a common scalar value¹ by repeatedly interacting with

Manuscript received September 16, 2008; revised January 28, 2009. First published April 14, 2009; current version published November 18, 2009. This work was supported in part by the Social and Information Sciences Laboratory, California Institute of Technology, by the Army Research Office under Grant W911NF04316, by the Air Force Office of Scientific Research under Grant FA9550-08-1-0375, and by the National Science Foundation under Grant ECS-0501394 and Grant ECCS-0547692. This paper was recommended by Associate Editor T. Vasilakos.

J. R. Marden is with the Social and Information Sciences Laboratory, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: marden@caltech.edu).

G. Arslan is with the Department of Electrical Engineering, University of Hawaii, Honolulu, HI 96822 USA (e-mail: gurdal@hawaii.edu).

J. S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: shamma@gatech.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCB.2009.2017273

¹The forthcoming results will also hold for multidimensional consensus.

one another. By reaching a 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 that lead to a consensus [2]–[9].

A well-studied protocol, which is 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, each player \mathcal{P}_i is informed of the current (or possibly delayed) state of all its neighbors. Player \mathcal{P}_i then updates its state by forming a convex combination of its state, along with the state of all its 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 induced by 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 that are 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.

There has been considerable recent research in the area of cooperative control (e.g., [10]–[12]). Surprisingly, there has been relatively little research that explicitly links cooperative control problems to the very relevant branches of learning in game literature [13] or multiagent system literature [14], [15] that address coordination problems. The goal of this paper is to better establish this link and develop new algorithms for broader classes of cooperative control problems and games.

In Section II, we establish a connection between cooperative control problems and a particular class of games known as “potential games,” and we model the consensus problem as a potential game. In Section III, we introduce a learning algorithm for potential games with state-dependent action sets. We show that the algorithm, when applied to the consensus problem, guarantees that players will come to a consensus, even in an environment filled with nonconvex obstructions.

In Section IV, 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. We go on to illustrate these methods on the consensus problem modeled as a sometimes weakly acyclic game. In Section V, we develop learning algorithms that can accommodate group-based decisions. In Section VI, we illustrate the connection between cooperative control and potential games on three separate problems, including functional consensus, sensor deployment, and sensor coverage. In Section VII, we present some final remarks.

II. COOPERATIVE CONTROL PROBLEM AND POTENTIAL GAME

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 a 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 they collectively accomplish the desired global objective.

One approach for cooperative control problems is to assign each individual player a fixed protocol or policy. This protocol precisely specifies what each player should do under any environmental condition. The consensus algorithm set forth in (1) is an example of such a policy-based approach. A challenge in this approach is to incorporate dynamic or evolving constraints on player policies. For example, suppose that a global planner desires a group of autonomous agents to physically converge to a central location in an environment that contains obstructions. The standard consensus algorithm may not be applicable to this problem since the limitations of 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, which is our main interest in this paper, is to assign each individual player a local objective (utility) 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. Provided that the assigned objective functions fall under a suitable category of games, one can appeal to algorithms with guaranteed properties for all the games within this category. In terms of the previous discussion, we will see that consensus, with or without obstacles, falls under the same category of games.

The challenge of control design in the game-theoretic approach lies in designing both the player objective functions and the learning dynamics, so that players collectively accomplish the objective of the global planner. Learning dynamics will be formulated as a repeated game, in which a one-stage game is repeated at 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 $a_i \in \mathcal{A}_i$ according to a prescribed learning rule that specifies how the player processes past observations from the interactions

at times $\{0, 1, \dots, t-1\}$ to select an action at time t . The learning dynamics that will be used throughout this paper is referred to as single-stage memory dynamics, which has a structural form that is similar to that of the consensus algorithm, i.e., the decision of any player \mathcal{P}_i at time t is made using only observations from the game played at time $t-1$. Of course, more general learning dynamics need not be restricted to single-stage memory.

A. Potential Games

Suppose that the objective of the global planner is captured by a potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$. We will impose that each player's objective function be appropriately "aligned" with the objective of the global planner. This notion of utility alignment (as presented in [16]) for multiagent systems has a strong connection to potential games [17].

Let

$$a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$$

denote the collection of actions of players other than player \mathcal{P}_i . With this notation, we will frequently express joint action a as (a_i, a_{-i}) .

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 $\phi : \mathcal{A} \rightarrow \mathbb{R}$

$$U_i(a''_i, a_{-i}) - U_i(a'_i, a_{-i}) = \phi(a''_i, a_{-i}) - \phi(a'_i, a_{-i}) \quad (2)$$

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 previously defined, requires perfect alignment between the global objective and the players' local objective functions in the following sense: If a player unilaterally changed its action, the change in its 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.² The connection between cooperative control problems and potential games is important, because learning algorithms for potential games have been extensively studied in the game theory literature [17]–[21]. Accordingly, if it is shown that a cooperative control problem can be formulated as a potential game, established learning algorithms with guaranteed asymptotic results could be used to tackle the cooperative control problem at hand.

Most of the learning algorithms for potential games guarantee convergence to a (pure) Nash equilibrium.

Definition 2.2 (Nash Equilibrium): An action profile $a^* \in \mathcal{A}$ is called a *pure Nash equilibrium* if, for all players $\mathcal{P}_i \in \mathcal{P}$

$$U_i(a_i^*, a_{-i}^*) = \max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}^*). \quad (3)$$

²We will omit mentioning other classes of potential games, such as generalized ordinal or weighted potential games, as they are just special cases of weakly acyclic games.

It is easy to see that, in potential games, any action profile maximizing the potential function is a pure Nash equilibrium. Hence, every potential game possesses at least one such equilibrium. However, there may also exist suboptimal pure Nash equilibria that do not maximize the potential function.

B. Consensus Modeled as a Potential Game

In this section, we will illustrate these concepts by showing that the consensus problem can be modeled as a potential game by appropriately defining the players' utilities. 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. The potential game formulation of the consensus problem discussed in this section requires the interaction graph to be time invariant and undirected. In Section IV-D, we relax these requirements by formulating the consensus problem as a sometimes weakly acyclic game.

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 problem³:

$$\phi(a) := - \sum_{\mathcal{P}_i \in \mathcal{P}} \sum_{\mathcal{P}_j \in N_i} \frac{\|a_i - a_j\|}{2} \quad (4)$$

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 neighbor sets $\{N_i\}_{i=1}^n$ is connected,⁴ the aforementioned potential function achieves the value of 0 if and only if action profile $a \in \mathcal{A}$ constitutes a *consensus*, i.e.,

$$\phi(a) = 0 \Leftrightarrow a_1 = \dots = a_n. \quad (5)$$

The goal is to assign each player an objective function that is perfectly aligned with the global objective in (4). One approach would be to assign each player the following objective function:

$$U_i(a) = \phi(a). \quad (6)$$

This assignment would require each player to observe the decision of all players to evaluate its payoff for a particular action choice, which may be infeasible. 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 with an undirected interaction topology, this translates to each player being assigned the objective function

$$U_i(a_i, a_{-i}) = - \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\|. \quad (7)$$

³This discussion uses a norm as a distance measure. Since we are dealing with finite actions sets, the norm $\|a_i - a_j\|$ could be replaced with a more general symmetric distance function $\delta(a_i, a_j)$, i.e., 1) $\delta(a_i, a_j) > 0 \Leftrightarrow a_i \neq a_j$, 2) $\delta(a_i, a_j) = 0 \Leftrightarrow a_i = a_j$, and 3) $\delta(a_i, a_j) = \delta(a_j, a_i)$ for all a_i, a_j .

⁴A graph is connected if there exists a path from any node to any other node.

Now, each player's objective function is *only* dependent on the actions of its neighbors. An objective function of this form is referred to as wonderful life utility (WLU, see [16] and [22]). It is known that assigning each agent a WLU leads to a potential game [16], [22]; however, we will explicitly show this for the consensus problem in the following claim:

Claim 2.1: Player objective functions (7) constitute a potential game with potential function (4), provided that the time-invariant interaction graph induced by 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

$$\phi(a) = - \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\| - \sum_{\mathcal{P}_j \neq \mathcal{P}_i} \sum_{\mathcal{P}_k \in N_j \setminus \mathcal{P}_i} \frac{\|a_j - a_k\|}{2}. \quad (8)$$

The change in the objective function 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

$$\begin{aligned} U_i(a_i^2, a_{-i}) - U_i(a_i^1, a_{-i}) \\ = \sum_{\mathcal{P}_j \in N_i} -\|a_i^2 - a_j\| + \|a_i^1 - a_j\| \end{aligned} \quad (9)$$

$$= \phi(a_i^2, a_{-i}) - \phi(a_i^1, a_{-i}). \quad (10)$$

■

Note that the preceding 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 [22].

It is straightforward to see that any consensus point is a Nash equilibrium of the game characterized by player objective functions (7). This is because a consensus point maximizes the potential function and the player objective functions (7).⁵ 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, e.g., $a^* \in \mathcal{A}^*$, can be suboptimal, i.e., $\phi(a^*) < 0$, and hence fail to be a consensus point.

With the consensus problem now formulated as a potential game, there are a large number of learning algorithms that are available with guaranteed results [13], [17], [20]–[23]. Most of the learning algorithms for potential games guarantee that the player behavior converges to a (possibly suboptimal) Nash equilibrium. In the ensuing section, we will focus on a particular learning algorithm for potential games that guarantees probabilistic convergence to a pure Nash equilibrium that maximizes the potential function.

⁵Let a be any consensus point. Then, $\phi(a) = 0$ and $U_i(a) = 0$ for all players \mathcal{P}_i . Therefore, a is a Nash equilibrium.

III. POTENTIAL GAME WITH STATE-DEPENDENT ACTION SET

In this section, we analyze potential games with state-dependent action sets. We will consider the special case where the set of actions available for a given player depends on the player's previous action.⁶ We will refer to state-dependent action sets of this form as (range) restricted action sets. We present a learning algorithm for this class of games and demonstrate that, when applied to the consensus problem, this algorithm guarantees consensus, even in an environment that contains arbitrary obstructions.

A. Background: SAP

Before stating the learning algorithm, we start with some notation. Let the strategy for player \mathcal{P}_i at time t be denoted by probability distribution $p_i(t) \in \Delta(\mathcal{A}_i)$, where $\Delta(\mathcal{A}_i)$ denotes the set of probability distributions over 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) [13], [25], [26]: 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 its 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 its strategy $p_i(t) \in \Delta(\mathcal{A}_i)$, where the a_i th component $p_i^{a_i}(t)$ of its strategy is given as

$$p_i^{a_i}(t) = \frac{\exp\{\beta U_i(a_i, a_{-i}(t-1))\}}{\sum_{\bar{a}_i \in \mathcal{A}_i} \exp\{\beta U_i(\bar{a}_i, a_{-i}(t-1))\}} \quad (11)$$

for some exploration parameter $\beta \geq 0$. Constant β 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 its best response set

$$\left\{ a_i \in \mathcal{A}_i : U_i(a_i, a_{-i}(t-1)) = \max_{a'_i \in \mathcal{A}_i} U_i(a'_i, a_{-i}(t-1)) \right\} \quad (12)$$

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 [25] as

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta\phi(\bar{a})\}}. \quad (13)$$

One can interpret stationary distribution μ as follows: For sufficiently large times $t > 0$, $\mu(a)$ is equal to the probability that $a(t) = a$. As $\beta \uparrow \infty$, all the weight of the stationary distribution μ is on the joint actions that maximize the potential function.

⁶We note that this scenario could have been formulated as a stochastic game [24], where the state is defined as the previous action profile, and the state-dependent action sets are defined accordingly. We will avoid formally defining the game as a stochastic game in favor of a direct presentation.

In the potential game formulation of the consensus problem, the joint actions that maximize the potential function (4) 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 β , then the players will asymptotically reach a consensus with arbitrarily high probability.

B. Learning Algorithm for Potential Games With Suboptimal Nash Equilibria and Restricted Action Sets

One issue with the applicability of the learning algorithm SAP to the consensus problem is that it permits any player to select any action in its 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 its current position. Therefore, we seek to modify SAP by conditioning a player's action set on its 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 its 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 its previous action.

We will introduce a variant of SAP called binary restrictive SAP (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 its 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 one trial action \hat{a}_i from its allowable set $R_i(a_i(t-1))$ with the following probabilities, where z_i denotes the maximum number of actions in any restricted action set for player \mathcal{P}_i , i.e., $z_i := \max_{a_i \in \mathcal{A}_i} |R_i(a_i)|$:

- 1) $\Pr[\hat{a}_i = a_i] = (1/z_i)$ for any $a_i \in R_i(a_i(t-1)) \setminus a_i(t-1)$;
- 2) $\Pr[\hat{a}_i = a_i(t-1)] = 1 - ((|R_i(a_i(t-1))| - 1)/z_i)$.

After player \mathcal{P}_i selects a trial action \hat{a}_i , the player chooses its action at time t as follows:

$$\Pr[a_i(t) = \hat{a}_i] = \frac{\exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\}}{D} \quad (14)$$

$$\Pr[a_i(t) = a_i(t-1)] = \frac{\exp\{\beta U_i(a_i(t-1))\}}{D} \quad (15)$$

where

$$D = \exp\{\beta U_i(\hat{a}_i, a_{-i}(t-1))\} + \exp\{\beta U_i(a_i(t-1))\} \quad (16)$$

and $\beta \geq 0$ 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 $a_i^0, a_i^m \in \mathcal{A}_i$, there exists a sequence of actions $a_i^0 \rightarrow a_i^1 \rightarrow \dots \rightarrow a_i^m$ that satisfies $a_i^k \in R_i(a_i^{k-1})$ for all $k \in \{1, 2, \dots, m\}$.

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 state space \mathcal{A} , where the unique stationary distribution $\mu \in \Delta(\mathcal{A})$ is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta\phi(\bar{a})\}} \quad \text{for any } a \in \mathcal{A}. \quad (17)$$

Proof: The proof follows along the lines of the proof of Theorem 6.2 in [13]. 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. Here, we show that this unique distribution must be (17) by verifying that the distribution (17) satisfies the detailed balanced equations

$$\mu(a)P_{ab} = \mu(b)P_{ba} \quad (18)$$

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

$$P_{ab} := \Pr[a(t) = b | a(t-1) = a]. \quad (19)$$

Note that the only nontrivial case is where a and b differ by exactly one player \mathcal{P}_i , i.e., $a_{-i} = b_{-i}$ 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 $1/z_i$ of being chosen, it follows that

$$\begin{aligned} \mu(a)P_{ab} &= \left[\frac{\exp\{\beta\phi(a)\}}{\sum_{z \in \mathcal{A}} \exp\{\beta\phi(z)\}} \right] \\ &\times \left[\left(\frac{1}{n}\right) \left(\frac{1}{z_i}\right) \frac{\exp\{\beta U_i(b)\}}{\exp\{\beta U_i(a)\} + \exp\{\beta U_i(b)\}} \right]. \end{aligned} \quad (20)$$

Letting

$$\lambda = \left(\frac{1}{\sum_{z \in \mathcal{A}} \exp\{\beta\phi(z)\}} \right) \times \left(\frac{(1/n)(1/z_i)}{\exp\{\beta U_i(a)\} + \exp\{\beta U_i(b)\}} \right) \quad (21)$$

we obtain

$$\mu(a)P_{ab} = \lambda \exp\{\beta\phi(a) + \beta U_i(b)\}. \quad (22)$$

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

$$\mu(a)P_{ab} = \lambda \exp\{\beta\phi(b) + \beta U_i(a)\} \quad (23)$$

which leads us to

$$\mu(a)P_{ab} = \mu(b)P_{ba}. \quad (24)$$

■

Note that, if all players adhere to the learning dynamics RSAP in a consensus problem where the interaction graph is

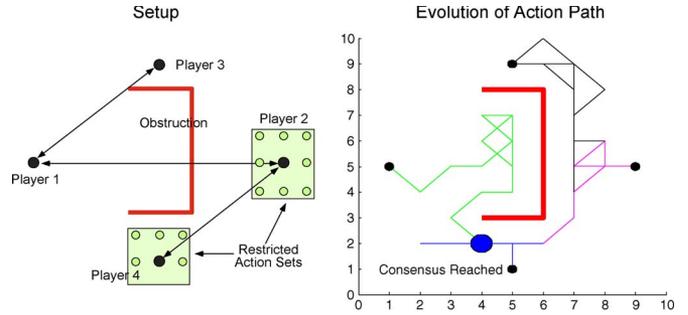


Fig. 1. Consensus problem with restricted action sets and arbitrary (nonconvex) obstructions.

time invariant and undirected, the restricted action sets satisfy Assumptions 3.1 and 3.2, and players are assigned the utilities (7); then, at sufficiently large times t , the players' collective behavior will maximize the potential function (4) with arbitrarily high probability, provided that β 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 nonconvex obstructions.

C. Example: Consensus in an Environment With Arbitrary 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 shown in Fig. 1.

The 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 nine possible actions; therefore, $z_i = 9$ for all players $\mathcal{P}_i \in \mathcal{P}$. The action sets are further restricted by the given obstruction.

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

IV. WEAKLY ACYCLIC AND SOMETIMES WEAKLY ACYCLIC GAMES

In potential games, the player 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 an *invariant Nash equilibrium*, to be defined later, in any sometimes weakly acyclic game.

A. Weakly Acyclic Games

Consider any finite game \mathcal{G} with player set \mathcal{P} , action set \mathcal{A} , and utility functions $\{U_i\}_{i=1}^n$. 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_ℓ} such that $a_{i_\ell}^\ell \neq a_{i_\ell}^{\ell+1}$, $a_{-i_\ell}^\ell = a_{-i_\ell}^{\ell+1}$, and $U_{i_\ell}(a^\ell) < U_{i_\ell}(a^{\ell+1})$. In other words, one player moves at a time, and each time that a player moves, it increases its own utility.

Suppose now that \mathcal{G} is a potential game with potential function ϕ . 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 ϕ is strictly increasing along the path. Since \mathcal{A} is finite, the path cannot indefinitely be extended. Hence, the last element in a maximal better reply path from any joint action a must be a Nash equilibrium of \mathcal{G} .

This idea may be generalized as follows: Game \mathcal{G} is weakly acyclic if, for any $a \in \mathcal{A}$, there exists a better reply path starting at a and ending at some pure Nash equilibrium of \mathcal{G} [13], [23]. Potential games are special cases of weakly acyclic games.

The preceding definition does not clearly identify the similarities between potential games and weakly acyclic games. Furthermore, using this definition to show that a given game \mathcal{G} (i.e., the players, objective functions, and action sets) is weakly acyclic can be problematic in that being weakly acyclic is a pathwise, rather than pointwise, property of the joint actions. With these issues in mind, we will now derive an equivalent definition for weakly acyclic games that utilizes potential functions.

Proposition 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 $U_i(a^1) > U_i(a^0)$ and $\phi(a^1) > \phi(a^0)$, where $a^1 = (a'_i, a^0_{-i})$.

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 ϕ is strictly increasing along the path. Since \mathcal{A} is finite, the path cannot indefinitely be extended. Hence, the last element in this path must be a Nash equilibrium.

(\Rightarrow) We will recursively construct a potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$. 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 ϕ over set \mathcal{A}^0 satisfying the following conditions:

$$\phi(a^0) < \phi(a^1) < \dots < \phi(a^n). \tag{25}$$

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 ϕ over set \mathcal{A}^1 satisfying the following conditions:

$$\phi(\tilde{a}^0) < \phi(\tilde{a}^1) < \dots < \phi(\tilde{a}^m). \tag{26}$$

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 ϕ over the truncated (re-defined) set $\mathcal{A}^1 = \{\tilde{a}^0, \tilde{a}^1, \dots, \tilde{a}^{k'-1}\}$ satisfying the following conditions:

$$\phi(\tilde{a}^0) < \phi(\tilde{a}^1) < \dots < \phi(\tilde{a}^{k'}). \tag{27}$$

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 potential function ϕ 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 $\phi(a'_i, a_{-i}) > \phi(a_i, a_{-i})$. ■

As with potential games, there are several learning algorithms with guaranteed results available for weakly acyclic games [20], [23], [27]. There are both advantages and disadvantages to formulating a cooperative control problem as a weakly acyclic game as opposed to a potential game. One advantage is flexibility in designing the player objective functions. In potential games, the player objective functions must perfectly be aligned with the potential function. In contrast to potential games, weakly acyclic games only require that at least one player's objective function is aligned with the potential function for any action profile. This flexibility in designing objective functions could be exploited in several ways to design more desirable control architectures for distributed systems. One example of this involves relaxing the structural requirements on the player objective functions, such as system requirements, e.g., an invariant interaction graph, or minimizing the degree to which a player's objective function depends on the actions of other players. An alternative example involves equilibrium manipulation, i.e., designing objective functions such that all pure Nash equilibrium are desirable.

When looking at the consensus problem, the potential game formulation required the interaction topology to be undirected as the following example illustrates:

Consider a consensus problem with three players $\mathcal{P}_1, \mathcal{P}_2$, and \mathcal{P}_3 with the following neighbor sets: $N_1 = \{\mathcal{P}_1, \mathcal{P}_2\}$, $N_2 = \{\mathcal{P}_2, \mathcal{P}_3\}$, and $N_3 = \{\mathcal{P}_3, \mathcal{P}_1\}$. Suppose that the action set of each player \mathcal{P}_i is $\mathcal{A}_i = \{a, b\}$ for some $a, b \in \mathbb{R}$. Consider the following better reply path, where the top, middle, and bottom actions are the actions of players $\mathcal{P}_1, \mathcal{P}_2$, and \mathcal{P}_3 , respectively:

$$\begin{aligned} \begin{bmatrix} a \\ b \\ a \end{bmatrix} &\rightarrow \begin{bmatrix} b \\ b \\ a \end{bmatrix} \rightarrow \begin{bmatrix} b \\ a \\ a \end{bmatrix} \rightarrow \begin{bmatrix} b \\ a \\ b \end{bmatrix} \\ &\rightarrow \begin{bmatrix} a \\ a \\ b \end{bmatrix} \rightarrow \begin{bmatrix} a \\ b \\ b \end{bmatrix} \rightarrow \begin{bmatrix} a \\ b \\ a \end{bmatrix}. \end{aligned} \tag{28}$$

If this were a potential game, then each deviation would also increase the potential function. However, this is not possible, because there exists a better reply cycle.

This requirement on the interaction graph is no longer necessary when formulating the consensus problem as a weakly acyclic game. For example, it is easy to see that the example in (28) is a weakly acyclic game, because there exists a better reply path from any action profile to a Nash equilibrium (or, in this case, a consensus point).

A disadvantage of weakly acyclic games when compared with potential games as a paradigm for designing distributed systems is the lack of a systematic procedure for utilizing this flexibility in designing the player objective functions from a given global objective. For example, there are several systematic procedures for designing the player objective functions such that the resulting game is a potential game [22]. One such example is the WLU discussed earlier. An open research question is understanding whether such a procedure for weakly acyclic games can be obtained and quantifying the possible gains by considering weakly acyclic games over potential games.

To avoid redundancy, we will explicitly omit modeling a general consensus problem with a time-invariant directed interaction topology as a weakly acyclic game. Rather, in Section IV-D, we will model the more general consensus problem with a time-varying directed interaction topology as a sometimes weakly acyclic game.

B. Sometimes Weakly Acyclic Games

In this section, we will extend the notion of weakly acyclic games to include state-dependent objective functions. This framework is known as a Markov or stochastic game [24].

In the potential game formulation of the consensus problem, each player was assigned a time-invariant objective function of the form (7). However, in the case of a time-varying interaction topology, we would like to allow player objective functions to be time varying. In this framework, each player \mathcal{P}_i is now assigned a local state-dependent objective function $U_i : \mathcal{A} \times X \rightarrow \mathbb{R}$, where X is the set of states. In the consensus problem, X could represent the set of possible interaction topologies. Denote the objective function of player \mathcal{P}_i at time t as $U_i(a(t), x(t))$, where $a(t)$ and $x(t)$ are the action profile and state at time t . The state dynamics take on the general form

$$x(t) = f(x(t-1), a(t-1), \mathcal{N}(t)) \quad (29)$$

where $\mathcal{N}(t)$ is nature's influence at time t .⁷

An action profile a^* is an *invariant Nash equilibrium* if

$$U_i(a^*, x) = \max_{a_i \in \mathcal{A}_i} U_i((a_i, a_{-i}^*), x) \quad \forall x \in X. \quad (30)$$

A game is *sometimes weakly acyclic* if there exists a potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$ and a finite time constant T such that the following property holds: For any time $t_0 > 0$, if

⁷For example, one can think of $\mathcal{N}(t)$ as time-varying neighborhood sets in the consensus problem.

$a(t_0) = a^0$ is *not* an invariant Nash equilibrium, then there exists a player $\mathcal{P}_i \in \mathcal{P}$, an action $a'_i \in \mathcal{A}_i$, and a time $t \in [t_0, t_0 + T]$, where $U_i((a'_i, a_{-i}^0), x(t)) > U_i(a^0, x(t))$ and $\phi(a'_i, a_{-i}^0) > \phi(a^0)$, provided that $a(t_0) = a(t_0 + 1) = \dots = a(t-1)$. Note that the sometimes weakly acyclic property depends on the objective functions, state dynamics, and nature's influence.

Note that a sometimes weakly acyclic game has at least one invariant Nash equilibrium, i.e., any action profile that maximizes potential function ϕ .

C. Learning Dynamics for Sometimes Weakly Acyclic Games

We will consider the better reply with inertia dynamics for games involving state-dependent objective functions. These dynamics are a slight extension of the finite memory and inertia dynamics in [23] to include state-dependent objective functions. Before stating the learning dynamics, we redefine a player's *better reply set* for any action profile $a \in \mathcal{A}$ and state $x \in X$ as

$$B_i(a, x) := \{a'_i \in \mathcal{A}_i : U_i((a'_i, a_{-i}), x) > U_i(a, x)\}. \quad (31)$$

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 :

$$B_i(a(t-1), x(t)) = \emptyset \Rightarrow a_i(t) = a_i(t-1) \quad (32)$$

$$B_i(a(t-1), x(t)) \neq \emptyset \Rightarrow \begin{cases} \Pr[a_i(t) = a_i(t-1)] = \alpha(t) \\ \Pr[a_i(t) = a'_i] = \frac{(1-\alpha(t))}{|B_i(a(t-1), x(t))|} \end{cases} \quad (33)$$

for any action $a'_i \in B_i(a(t-1), x(t))$, where $\alpha(t) \in (0, 1)$ is 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 exists constants $\underline{\epsilon}$ and $\bar{\epsilon}$ such that, for all times $t \geq 0$ and for all players $\mathcal{P}_i \in \mathcal{P}$

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

Theorem 4.1: Consider any 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 almost surely converge to an invariant Nash equilibrium.

Proof: Let $\phi : \mathcal{A} \rightarrow \mathbb{R}$ and T be the potential function and time constant for the sometimes weakly acyclic game, respectively. Let $a(t_0) = a^0$ be the action profile and $x(t_0)$ be the state at time t_0 . If a^0 is an invariant Nash equilibrium, then $a(t) = a^0$ for all times $t \geq t_0$, and we are done. Otherwise, there exists a time $t_1 \in [t_0, t_0 + T]$, a player $\mathcal{P}_i \in \mathcal{P}$, and an action $a'_i \in \mathcal{A}_i$ such that $U_i(a'_i, a_{-i}^0, x(t_1)) > U_i(a^0, x(t_1))$ and $\phi(a'_i, a_{-i}^0) > \phi(a^0)$, provided that $a(t_0) = a(t_0 + 1) = \dots = a(t_1 - 1)$. Because of the players' inertia, action $a^1 = (a'_i, a_{-i}^0)$

will be played at time t_1 with at least probability $\underline{\epsilon}^{n-1}((1 - \bar{\epsilon})/|\mathcal{A}|)^{\underline{\epsilon}^{nT}}$.

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

$$\Pr [a(t) = a^* \quad \forall t \geq t^*] \geq \epsilon^* \quad (34)$$

where

$$t^* = t_0 + T|\mathcal{A}| \quad (35)$$

$$\epsilon^* = \left(\underline{\epsilon}^{n-1} \frac{(1 - \bar{\epsilon})}{|\mathcal{A}|} \underline{\epsilon}^{nT} \right)^{|\mathcal{A}|}. \quad (36)$$

■

D. Consensus Modeled as a Sometimes Weakly Acyclic Game

Two main drawbacks 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 such as 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 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-D consensus problem with appropriately designed player objective functions is a sometimes weakly acyclic game. One can easily extend this to the multidimensional case.

1) *Setup: Consensus Problem With a Time-Varying and Directed Interaction Graph:* Consider a consensus problem with an 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 $U_i : \mathcal{A} \times X_i \rightarrow \mathbb{R}$, where X_i is the set of states for player \mathcal{P}_i . We define the state of player \mathcal{P}_i at time t as the tuple

$$x_i(t) = \{N_i(t), a_i(t-1)\} \quad (37)$$

where $X_i := 2^{\mathcal{P}} \times \mathcal{A}_i$, $X := \prod_{\mathcal{P}_i} X_i$, and $2^{\mathcal{P}}$ denotes the power sets of \mathcal{P} . We note that there are many alternative possibilities for the state selection. For example, one could alternatively define the state of player \mathcal{P}_i at time t as

$$x_i(t) = \left\{ N_i(t), N_i(t-1), \{a_j(t-1)\}_{j \in N_i(t)}, \{a_j(t-2)\}_{j \in N_i(t-1)} \right\}. \quad (38)$$

This structure allows a player's utility function to depend on information from the last two time periods. Hence, a player's utility function could be designed to depend on how players are changing, as opposed to a static view of the players' actions, which is the structure of the consensus algorithm (1).

In this section, we will focus on the 1-D consensus problem with player states as defined in (37) and a disagreement function of the form

$$D(a, \bar{\mathcal{P}}) := \max_{\mathcal{P}_i, \mathcal{P}_j \in \bar{\mathcal{P}}} (a_i - a_j) \quad (39)$$

for some nonempty player set $\bar{\mathcal{P}} \subseteq \mathcal{P}$. We note that this measure could be generalized for larger dimensional spaces; however, we will focus purely on the state definition (37) and the disagreement measure (39) to highlight the connections between the consensus problem and sometimes weakly acyclic games.

Rather than specifying a particular objective functions as in (7), we will introduce a class of admissible objective functions. An objective function for player \mathcal{P}_i is called a *reasonable objective function* if, for any action profile $a \in \mathcal{A}$ and state $x_i \in X_i$, the better response set satisfies the following two conditions:

$$B_i(a, x_i) \subset \{a'_i \in \mathcal{A}_i : D((a'_i, a_{-i}), N_i) \leq D(a, N_i)\} \quad (40)$$

$$|\{a'_i \in \mathcal{A}_i : D((a'_i, a_{-i}), N_i) \leq D(a, N_i)\}| > 1$$

$$\Rightarrow B_i(a, x_i) \neq \emptyset. \quad (41)$$

Roughly speaking, these conditions ensure that a player will not value moving further away from its belief about the location of its neighbors. An example of a reasonable objective function is

$$U_i(a, \{N_i, \bar{a}_i\}) = -D(a, N_i) - \gamma I\{a_i = \bar{a}_i\} \quad (42)$$

where $\{N_i, \bar{a}_i\} \in X_i$, $I\{\cdot\}$ is the usual indicator function, and γ penalizes players for immobility. If $\gamma > 0$ is sufficiently small, then it is easy to verify that (42) is a reasonable objective function (since action sets \mathcal{A}_i are finite).

We will now relax our requirements on the connectivity and time invariance of the interaction graph in the consensus problem. A common assumption (e.g., [8]) on the interaction graph is connectedness over intervals.

Assumption 4.2 (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.

Proposition 4.2: Under Assumption 4.2, reasonable objective functions satisfying (40) and (41) constitute a sometimes weakly acyclic game. Furthermore, every invariant Nash equilibrium constitutes consensus.

Proof: To prove that the game is sometimes weakly acyclic, we introduce the following potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$, which depends on the disagreement measure (39) and the number of players at the boundaries

$$\phi(a) = -D(a, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a)}{n} \right) \quad (43)$$

where

$$\mathcal{P}_{\min}(a) := \left\{ \mathcal{P}_i \in \mathcal{P} : a_i = \min_j a_j \right\} \quad (44)$$

$$\mathcal{P}_{\max}(a) := \left\{ \mathcal{P}_i \in \mathcal{P} : a_i = \max_j a_j \right\} \quad (45)$$

$$\bar{n}(a) := \min(|\mathcal{P}_{\min}(a)|, |\mathcal{P}_{\max}(a)|) \quad (46)$$

$$\delta_A := \min \left\{ |D(a^1, \mathcal{P}^1) - D(a^2, \mathcal{P}^2)| : a^1, a^2 \in \mathcal{A}, \mathcal{P}^1, \mathcal{P}^2 \in 2^{\mathcal{P}}, D(a^1, \mathcal{P}^1) \neq D(a^2, \mathcal{P}^2) \right\}. \quad (47)$$

Note that the potential function is a nonpositive 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.⁸

To show that the reasonable objective functions constitute a sometimes weakly acyclic game, we need to show that (43) satisfies the conditions set forth in Section IV-B. It is easy to see that any consensus point is an invariant Nash equilibrium. We will show that, if an action profile is not a consensus point, then there exists a player who can increase its objective function and the potential function at some time in a fixed time window. This implies that every invariant Nash equilibrium is a consensus point and, furthermore, that the game is sometimes 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 $U_i((a'_i, a_{-i}^0), x_i(t_1)) > U_i(a^0, x_i(t_1))$ and $\phi(a'_i, a_{-i}^0) > \phi(a^0)$, provided that $a(t_0) = \dots = a(t_1 - 1)$. To see this, let $\mathcal{P}'(a^0)$ be the minimum boundary player set, i.e., $\mathcal{P}'(a^0) = \mathcal{P}_{\min}(a^0)$ if $|\mathcal{P}_{\min}(a^0)| \leq |\mathcal{P}_{\max}(a^0)|$ and $\mathcal{P}'(a^0) = \mathcal{P}_{\max}(a^0)$ otherwise. Since the interaction graph satisfies Assumption 4.2,⁹ 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 $\mathcal{P}_j \in N_i(t_1) \setminus \mathcal{P}'$. Therefore

$$|B_i(a^0, \{N_i(t_1), a_i^0\})| \neq \emptyset. \quad (48)$$

This is true, because there exists at least two actions for player \mathcal{P}_i that do not increase the disagreement measure, i.e., a_i^0 , trivially, and a_j^0 , as $D(a_j^0, a_{-i}^0) \leq D(a^0)$.

Let $a'_i \in B_i(a^0, x_i(t_1))$, $a'_i \neq a_i^0$, and for notational 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

$$\begin{aligned} \phi(a^1) &= -D(a^1, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^1)}{n} \right) \\ &> -D(a^0, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^1)}{n} \right) + \delta_A \\ &> -D(a^0, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^0) + n}{n} \right) + \delta_A \\ &= \phi(a^0). \end{aligned} \quad (49)$$

⁸In the 1-D consensus problem, there are two boundaries, i.e., the maximum and minimum values. In higher dimensional spaces, one would need to be more careful with the definition of boundaries. However, the same structural form of the potential function in (43) could be used to prove that higher dimensional consensus problems with appropriately defined player objective functions constitute a sometimes weakly acyclic games.

⁹Note that Assumption 4.2 is stronger than necessary for this proof.

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

$$\begin{aligned} \phi(a^1) &= -D(a^0, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^1)}{n} \right) \\ &> -D(a^0, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^1) + 1}{n} \right) \\ &= -D(a^0, \mathcal{P}) + \delta_A \left(1 - \frac{\bar{n}(a^0)}{n} \right) \\ &= \phi(a^0) \end{aligned} \quad (50)$$

where the third equality comes from the fact that $\bar{n}(a^1) = \bar{n}(a^0) - 1$. Therefore, a^0 is not an invariant Nash equilibrium, and the game is sometimes weakly acyclic. This completes the proof. ■

Combining Proposition 4.2 and Theorem 4.1, we conclude that, if all players adhere to the better reply with inertia dynamics in a consensus problem where the interaction graph satisfies Assumption 4.2 and the players are assigned reasonable objective functions, then the joint action profile will almost surely converge to a consensus point.

This section illustrates the main advantages of designing objective functions and state dynamics within the framework of sometimes weakly acyclic games. For the consensus problem, we reduced the structural requirement on the player objective functions, thereby allowing a time-varying directed interaction graph. Furthermore, the aforementioned design ensured that all invariant Nash equilibria were desirable, i.e., consensus points. This did not hold in the potential game formulation.

V. GROUP-BASED DECISION PROCESS FOR POTENTIAL GAMES

In this section, we analyze the situation where players are allowed to collaborate with a group of other players when making a decision. In particular, we extend SAP to accommodate such a grouping structure. Our main motivation for considering group-based decision processes is the possibility of coupled constraints on the players' action sets.

A. SAP With Group-Based Decisions

We consider a variation of traditional noncooperative games to include group-based decisions. The structure of these group-based games is given as follows: There exists a finite set of players $\mathcal{P} = (\mathcal{P}_1, \dots, \mathcal{P}_n)$, each player has a finite action set \mathcal{A}_i , and each group or players $G \subseteq \mathcal{P}$ is assigned a group utility function $U_G : \mathcal{A} \rightarrow \mathbb{R}$. We will call such a game a *group-based potential game* if there exists a potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$ such that, for any group $G \subseteq \mathcal{P}$, collective group actions $a'_G, a''_G \in \mathcal{A}_G := \prod_{\mathcal{P}_i \in G} \mathcal{A}_i$, and $a_{-G} \in \prod_{\mathcal{P}_i \notin G} \mathcal{A}_i$, i.e.,

$$U_G(a''_G, a_{-G}) - U_G(a'_G, a_{-G}) = \phi(a''_G, a_{-G}) - \phi(a'_G, a_{-G}). \quad (51)$$

Notice that any group-based potential game is also a potential game.

We will now introduce a variant of SAP to accommodate group-based decisions. At each time $t > 0$, a group of players $G \subseteq \mathcal{P}$ is randomly chosen according to a fixed probability distribution $q \in \Delta(2^{\mathcal{P}})$. We will refer to q_G as the probability that group G will be chosen. We make the following assumption on the group probability distribution:

Assumption 5.1 (Completeness): For any player $\mathcal{P}_i \in \mathcal{P}$, there exists a group $G \subseteq \mathcal{P}$ such that $\mathcal{P}_i \in G$ and $q_G > 0$.

Once a group is selected, the group is unilaterally allowed to alter its collective strategy. All players not in the group must repeat their actions, i.e., $a_{-G}(t) = a_{-G}(t-1)$, where a_G is the action tuple of all players in group G , and a_{-G} is the action tuple of all players not in group G . At time t , the updating group G randomly selects a collective action from \mathcal{A}_G according to collective strategy $p_G(t) \in \Delta(\mathcal{A}_G)$, where the a_G th component $p_G^{a_G}(t)$ of the collective strategy is defined as

$$p_G^{a_G}(t) = \frac{\exp\{\beta U_G(a_G, a_{-G}(t-1))\}}{\sum_{\bar{a}_G \in \mathcal{A}_G} \exp\{\beta U_G(\bar{a}_G, a_{-G}(t-1))\}} \quad (52)$$

for some exploration parameter $\beta \geq 0$.

We will now show that the convergence properties of the learning algorithm SAP still hold with group-based decisions.

Theorem 5.1: Consider a finite n -player group-based potential game with potential function $\phi(\cdot)$ and a group probability distribution q satisfying Assumption 5.1. SAP with group-based decisions induces a Markov process over state space \mathcal{A} , where the unique stationary distribution $\mu \in \Delta(\mathcal{A})$ is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta\phi(\bar{a})\}} \quad \text{for any } a \in \mathcal{A}. \quad (53)$$

Proof: The proof follows along the lines of the proof of Theorem 6.2 in [13]. By Assumption 5.1, the Markov process induced by SAP with group-based decisions is irreducible and aperiodic; therefore, the process has a unique stationary distribution. Here, we show that this unique distribution must be (53) by verifying that the distribution (53) satisfies the detailed balanced equations

$$\mu(a)P_{ab} = \mu(b)P_{ba} \quad (54)$$

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

$$P_{ab} := \Pr[a(t) = b | a(t-1) = a]. \quad (55)$$

Note that there are now several ways to transition from a and b when incorporating group-based decisions. Let $\bar{G}(a, b)$ represent the group of players with different actions in a and b , i.e.,

$$\bar{G}(a, b) := \{\mathcal{P}_i \in \mathcal{P} : a_i \neq b_i\}. \quad (56)$$

Let $G(a, b) \subseteq 2^{\mathcal{P}}$ be the complete set of player groups for which the transition from a to b is possible, i.e.,

$$G(a, b) := \{G \in 2^{\mathcal{P}} : \bar{G}(a, b) \subseteq G\}. \quad (57)$$

Since a group $G \in G(a, b)$ has probability q_G of being chosen in any given period, it follows that

$$\begin{aligned} \mu(a)P_{ab} &= \left[\frac{\exp\{\beta\phi(a)\}}{\sum_{z \in \mathcal{A}} \exp\{\beta\phi(z)\}} \right] \\ &\times \left[\sum_{G \in G(a, b)} q_G \frac{\exp\{\beta U_G(b)\}}{\sum_{\bar{a}_G \in \mathcal{A}_G} \exp\{\beta U_G(\bar{a}_G, a_{-G})\}} \right]. \end{aligned} \quad (58)$$

Letting

$$\begin{aligned} \lambda_G &:= \left(\frac{1}{\sum_{z \in \mathcal{A}} \exp\{\beta\phi(z)\}} \right) \\ &\times \left(\frac{q_G}{\sum_{\bar{a}_G \in \mathcal{A}_G} \exp\{\beta U_G(\bar{a}_G, a_{-G})\}} \right) \end{aligned} \quad (59)$$

we obtain

$$\mu(a)P_{ab} = \sum_{G \in G(a, b)} \lambda_G \exp\{\beta\phi(a) + \beta U_G(b)\}. \quad (60)$$

Since $U_G(b) - U_G(a) = \phi(b) - \phi(a)$ and $G(a, b) = G(b, a)$, we have

$$\mu(a)P_{ab} = \sum_{G \in G(b, a)} \lambda_G \exp\{\beta\phi(b) + \beta U_G(a)\} \quad (61)$$

which leads us to

$$\mu(a)P_{ab} = \mu(b)P_{ba}. \quad (62)$$

■

B. Coupled Constraints on Group Action Sets

In the previous section, the updating group employed a strategy with a probability distribution having full support on group action set $\mathcal{A}_G = \prod_{\mathcal{P}_i \in G} \mathcal{A}_i$. In this section, we consider the situation where the actions available to a given group are constrained, i.e., $\mathcal{A}_G \subset \prod_{\mathcal{P}_i \in G} \mathcal{A}_i$.

In this setting, the updating group G randomly selects a collective action from \mathcal{A}_G according to collective strategy $p_G(t) \in \Delta(\mathcal{A}_G)$, where, for any action $a_G \in \mathcal{A}_G$

$$p_G^{a_G}(t) = \frac{\exp\{\beta U_G(a_G, a_{-G}(t-1))\}}{\sum_{\bar{a}_G \in \mathcal{A}_G} \exp\{\beta U_G(\bar{a}_G, a_{-G}(t-1))\}} \quad (63)$$

for some exploration parameter $\beta \geq 0$. Otherwise, for any action $a_G \notin \mathcal{A}_G$, $p_G^{a_G}(t) = 0$.

These dynamics define a Markov process over a *constrained* state space $\bar{\mathcal{A}} \subseteq \mathcal{A}$ that can be characterized as follows: Let $a(0)$ be the initial actions of all players. If $\bar{a} \in \bar{\mathcal{A}}$, then there exists a sequence of action profiles $a(0) = a^0, a^1, \dots, a^n = \bar{a}$, with the condition that, for all $k \in \{1, 2, \dots, n\}$, $a^k = (a_{G_k}^k, a_{-G_k}^{k-1})$ for a group $G_k \subseteq \mathcal{P}$, where $q_{G_k} > 0$ and $a_{G_k}^k \in \mathcal{A}_{G_k}$. In words, $\bar{\mathcal{A}}$ is the recurrent class of reachable states starting from $a(0)$.

Theorem 5.2: Consider a finite n -player group-based potential game with potential function $\phi(\cdot)$ and a group probability distribution q satisfying Assumption 5.1. SAP with group-based decisions and constrained group action sets $\{\mathcal{A}_G\}_{G \subseteq \mathcal{P}}$ induces a Markov process over the constrained state space $\bar{\mathcal{A}} \subseteq \mathcal{A}$. The unique stationary distribution $\mu \in \Delta(\bar{\mathcal{A}})$ is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \bar{\mathcal{A}}} \exp\{\beta\phi(\bar{a})\}} \quad \text{for any } a \in \bar{\mathcal{A}}. \quad (64)$$

The proof of Theorem 5.1 also applies to Theorem 5.2.

C. Restricted SAP With Group-Based Decisions

Extending these results to accommodate restricted action sets is straightforward. Let $a(t-1)$ be the action profile at time $t-1$. At time t , the updating group G randomly selects one trial action \hat{a}_G from the group's restricted action set $R_G(a_G(t-1)) \subseteq \mathcal{A}_G$ with the following probabilities, where z_G denotes the maximum number of actions in any restricted action set for group G , i.e., $z_G := \max_{a_G \in \mathcal{A}_G} |R_G(a_G)|$:

- $\Pr[\hat{a}_G = a_G] = (1/z_G)$ for any $a_G \in R_G(a_G(t-1)) \setminus a_G(t-1)$;
- $\Pr[\hat{a}_G = a_G(t-1)] = 1 - (|R_G(a_G(t-1))| - 1)/z_G$.

After group G selects a trial action \hat{a}_G , the updating group G selects its action $a_G(t)$ according to the collective strategy, i.e.,

$$\Pr[a_G(t) = \hat{a}_G] = \frac{\exp\{\beta U_G(\hat{a}_G, a_{-G}(t-1))\}}{D} \quad (65)$$

$$\Pr[a_G(t) = a_G(t-1)] = \frac{\exp\{\beta U_G(a(t-1))\}}{D} \quad (66)$$

where

$$D = \exp\{\beta U_G(\hat{a}_G, a_{-G}(t-1))\} + \exp\{\beta U_G(a(t-1))\} \quad (67)$$

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

As before, these dynamics define a Markov process over a constrained state space $\bar{\mathcal{A}} \subseteq \mathcal{A}$, where $\bar{\mathcal{A}}$ is the set of reachable states from $a(0)$. Following the previous discussion, if $\bar{a} \in \bar{\mathcal{A}}$, then there exists a sequence of action profiles $a(0) = a^0, a^1, \dots, a^n = \bar{a}$, with the condition that, for all $k \in \{1, 2, \dots, n\}$, $a^k = (a_{G_k}^k, a_{-G_k}^{k-1})$ for a group $G_k \subseteq \mathcal{P}$, where $q_{G_k} > 0$ and $a_{G_k}^k \in \mathcal{A}_{G_k}$. Furthermore, $a_{G_k}^k \in R_{G_k}(a_{G_k}^{k-1}) \subseteq \mathcal{A}_{G_k}$ for all $k \in \{1, \dots, n\}$.

We will state the following theorem without proof since it follows from arguments similar to the proof of Theorem 5.1.

Theorem 5.3: Consider a finite n -player group-based potential game with potential function $\phi(\cdot)$ and a group probability distribution q satisfying Assumption 5.1. If the group restricted action sets are reversible (i.e., satisfies Assumption 3.1 for all groups), then RSAP with group-based decisions and constrained group action sets $\{\mathcal{A}_G\}_{G \subseteq \mathcal{P}}$ induces a Markov process

over constrained state space $\bar{\mathcal{A}} \subseteq \mathcal{A}$. The unique stationary distribution $\mu \in \Delta(\bar{\mathcal{A}})$ is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \bar{\mathcal{A}}} \exp\{\beta\phi(\bar{a})\}} \quad \text{for any } a \in \bar{\mathcal{A}}. \quad (68)$$

VI. ILLUSTRATION

In this section, we will illustrate the broad applicability of the theoretical results presented in this paper to three separate problems: 1) power management in sensor networks; 2) dynamic sensor coverage; and 3) functional consensus.

A. Sensor Deployment Problem

In this section, we consider the sensor deployment problem described in [28] and the references therein. Consider the problem of transferring data from immobile sources to immobile destinations through the use of mobile intermediate nodes or relays. The deployment problem concerns positioning the intermediate nodes to successfully transfer the data from the sources to the destinations while optimizing some network performance metric.

We will model the nodes, both immobile and mobile, as players $\{\mathcal{P}_1, \dots, \mathcal{P}_n\}$, with finite action sets \mathcal{A}_i representing the set of physical locations that the node can reach. For example, in the case of an immobile node, the action set is a singleton consisting of only the node's fixed location. We will assume that the number of nodes and the information flow are set *a priori*. The information flow is determined by a fixed undirected graph $G(V, E)$ with the set of nodes $V = \mathcal{P}$ and the set of edges $E \subset \mathcal{P} \times \mathcal{P}$. The set of edges defines the information flow. We will adopt the notation that, if information is passed from player \mathcal{P}_i to player \mathcal{P}_j , then player \mathcal{P}_j is in the neighbor set of player \mathcal{P}_i , i.e., $\mathcal{P}_j \in N_i$.

A common metric used to assess the transmission cost between nodes is power. For a given allocation of sensors (a_1, \dots, a_n) , the power for transmitting information from sensor \mathcal{P}_i to \mathcal{P}_j typically takes on the form

$$e(a_i, a_j) = \alpha_1 + \alpha_2 \|a_i - a_j\|^2 \quad (69)$$

where α_1 and α_2 are positive constants [28], [29]. A well-studied performance objective is to find a minimum power deployment. That is, to find an allocation $(a_1^*, \dots, a_n^*) \in \mathcal{A}$ that minimizes the total transmission power used in the network

$$\sum_{\mathcal{P}_i \in \mathcal{P}} \sum_{\mathcal{P}_j \in N_i} e(a_i^*, a_j^*). \quad (70)$$

This is equivalent to maximizing performance metric $\phi: \mathcal{A} \rightarrow \mathbb{R}$, where

$$\phi(a) = - \sum_{\mathcal{P}_i \in \mathcal{P}} \sum_{\mathcal{P}_j \in N_i} \|a_i - a_j\|^2. \quad (71)$$

At this stage, it is interesting to note that the potential function used in the consensus problem (4) is equivalent (see footnote 3) to the potential function representing the minimum

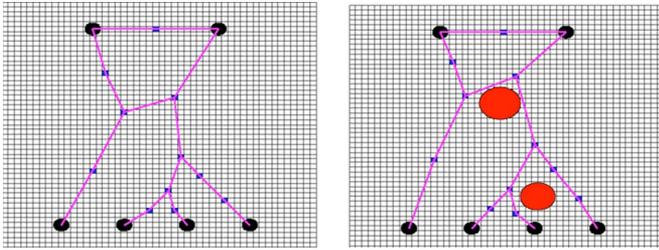


Fig. 2. Final configuration of nodes in the environment with and without obstructions.

power in the network (71). This implies that, if each node is assigned a local utility function $U_i : \mathcal{A} \rightarrow \mathbb{R}$ of the form

$$U_i(a_i, a_{-i}) = -2 \sum_{P_j \in N_i} \|a_i - a_j\|^2 \quad (72)$$

then we have an exact potential game with potential function (71). Therefore, if all agents update their locations using SAP or RSAP (assuming restricted action sets), the stationary distribution of the process is

$$\mu(a) = \frac{\exp\{\beta\phi(a)\}}{\sum_{\bar{a} \in \mathcal{A}} \exp\{\beta\phi(\bar{a})\}} \quad \text{for any } a \in \mathcal{A}. \quad (73)$$

As $\beta \uparrow \infty$, all the weight of the stationary distribution is placed on action profiles that maximize the potential function. In the problem of consensus, these action profiles represent consensus points. Alternatively, in the problem of sensor deployment, these action profiles represent minimum power allocations.

For illustration purposes, we consider a sensor deployment problem with 17 nodes (i.e., six immobile and 11 mobile nodes). We fix the location of the six immobile nodes and the interaction graph, as shown in Fig. 2, and randomly choose the starting locations of the remaining 11 mobile nodes. We consider the sensor deployment problem in two settings. In the first setting, suppose that the power of a transmission is given by

$$e(a_i, a_j) = \|a_i - a_j\|^2. \quad (74)$$

We simulated the sensor deployment problem using RSAP, with the exploration parameter chosen as $\beta(t) = 1 + t/300$. The final configuration of the sensors is given in Fig. 2(a). The evolution of the total transmission power in the network is shown in Fig. 3. One can observe that an efficient network is realized after approximately 100 iterations.

It is well known that solving for the optimal node locations in such a setting is a convex optimization problem. With that in my mind, we will make the problem more challenging (and nonconvex) by adding obstructions to the environment. Obstructions can be thought of as introducing variations in transmission costs. In this setting, the transmission power takes on the form

$$e(a_i, a_j) = \begin{cases} \|a_i - a_j\|^2, & \text{if no obstruction} \\ 1.3\|a_i - a_j\|^2, & \text{if obstruction.} \end{cases} \quad (75)$$

One could imagine that an obstruction, such as bad terrain, could require additional transmission power.

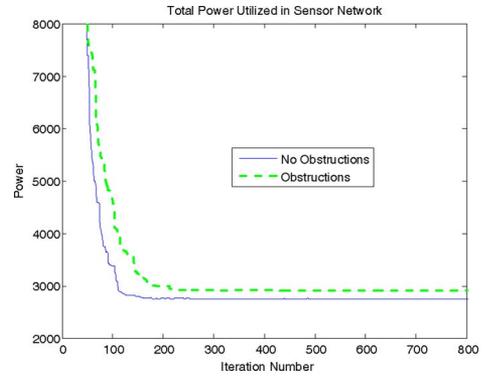


Fig. 3. Evolution of transmission power utilized in network.

We simulated this new sensor deployment problem using RSAP, with the exploration parameter chosen as $\beta(t) = 1 + t/300$. The environmental obstructions and the final configuration of the sensors are given in Fig. 2(b). The evolution of the total power utilized in the network is shown in Fig. 3. One can observe that an efficient network is realized after approximately 200 iterations. Furthermore, the environmental obstructions did not significantly impact the total transmission power as the intermediate nodes were able to adjust their positions to compensate for the obstructions.

B. Dynamic Sensor Coverage Problem

In this section, we will develop the dynamic sensor coverage problem described in [28] and the references therein to further illustrate the range applicability of the theory developed in this paper. The goal of the sensor coverage problem is to allocate a fixed number of sensors across a given “mission space” to maximize the probability of detecting a particular event.

We will divide the mission space into a finite set of sectors denoted as S . There exists an event density function or value function $V(s)$ that is defined over S . We will assume that $V(s) \geq 0 \forall s \in S$ and $\sum_{s \in S} V(s) = 1$. In the application of enemy submarine tracking, $V(s)$ could be defined as the *a priori* probability that an enemy submarine is situated in sector s .

There are a finite number of autonomous sensors denoted as $\mathcal{P} = \{\mathcal{P}_1, \dots, \mathcal{P}_n\}$ allocated to the mission space. Each sensor \mathcal{P}_i can position itself in any particular sector, i.e., the action set of sensor \mathcal{P}_i is $\mathcal{A}_i = S$. Furthermore, each sensor has limited sensing and moving capabilities. If an event occurs in sector s , the probability of sensor \mathcal{P}_i detecting the event, given its current location a_i , is denoted as $p_i(s, a_i)$. Typically, each sensor has a finite sensing radius r_i , where the probability of detection obeys the following:

$$\|s - a_i\| < r_i \Leftrightarrow p_i(s, a_i) > 0. \quad (76)$$

For a given joint action profile $a = \{a_1, \dots, a_n\}$, the joint probability of detecting an event in sector s is given by

$$P(s, a) = 1 - \prod_{\mathcal{P}_i \in \mathcal{P}} [1 - p_i(s, a_i)]. \quad (77)$$

In general, a global planner would like the sensors to allocate themselves in such a way as to maximize the following potential

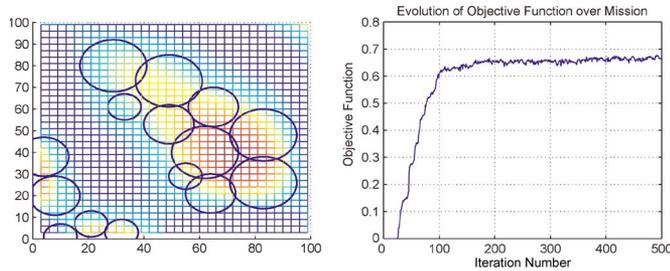


Fig. 4. Sensor coverage: final configuration and evolution of potential function over mission.

function:

$$\phi(a) = \sum_{s \in S} V(s)P(s, a). \quad (78)$$

We will assign each sensor a WLU [16], [22]. The utility of sensor \mathcal{P}_i , given any action profile $a \in \mathcal{A}$, is now

$$U_i(a) = \phi(a_i, a_{-i}) - \phi(a_i^0, a_{-i}) \quad (79)$$

where action a_i^0 is defined as the *null* action, which is equivalent to sensor \mathcal{P}_i turning off all sensing capabilities. The term $\phi(a_i^0, a_{-i})$ captures the value of the allocation of all sensors other than sensor \mathcal{P}_i . In this setting, a sensor can evaluate his utility using only local information. Furthermore, the resulting game is a potential game with potential function $\phi(\cdot)$.

In the following simulation, we have the mission space and value function as shown in Fig. 4. The mission space is $S = \{1, 2, \dots, 100\} \times \{1, 2, \dots, 100\}$, and the value function satisfies $\sum_{s \in S} V(s) = 1$. We have 18 different autonomous sensors (i.e., six sensors with a sensing radius of 6, six sensors with a sensing radius of 12, and six sensors with a sensing radius of 18). For simplicity, each sensor will have perfect sensing capabilities within its sensing radius, i.e., for any sector s satisfying $\|s - a_i\| < r_i$, then $p_i(s, a_i) = 1$. Each sensor is endowed with the WLU, as expressed in (79). All 18 sensors originally started at location (1, 1), and each sensor has range restricted action sets that are identical to those in the consensus problem shown in Fig. 1. We ran the binary RSAP with $\beta = 0.6$. Fig. 4 shows a snapshot of the sensor configuration at the final iteration, along with the evolution of the potential function over the mission. The highlighted circles indicate the sensing radii of the sensors.

C. Functional Consensus

In the consensus problem, as described in Section II-B, the global objective was for all agents to reach consensus. In this section, we will analyze the functional consensus problem where the goal is for all players to reach a specific consensus point, which is typically dependent on the initial action of all players, i.e.,

$$\lim_{t \rightarrow \infty} a_i(t) = f(a(0)) \quad \forall \mathcal{P}_i \in \mathcal{P} \quad (80)$$

where $a(0) \in \mathcal{A}$ is the initial action of all players, and $f: \mathcal{A} \rightarrow \mathbb{R}$ is the desired function. An example of such a function for an

n -player consensus problem is

$$f(a(0)) = \frac{1}{n} \sum_{\mathcal{P}_i \in \mathcal{P}} a_i(0) \quad (81)$$

for which the goal would be for all players to agree upon the average of the initial actions of all players. We will refer to this specific functional consensus problem as average consensus.

To achieve average consensus, the consensus algorithm of (1) requires that the interaction graph is connected and that the associated weighting matrix $\Omega = \{\omega_{ij}\}_{\mathcal{P}_i, \mathcal{P}_j \in \mathcal{P}}$ is doubly stochastic [6]. A doubly stochastic matrix is any matrix where all coefficients are nonnegative and all column sums and rows sums are equal to 1. The consensus algorithm takes on the following matrix form:

$$a(t+1) = \Omega a(t). \quad (82)$$

If Ω is a doubly stochastic matrix, then, for any time $t > 0$

$$\mathbf{1}^T a(t+1) = \mathbf{1}^T \Omega a(t) = \mathbf{1}^T a(t). \quad (83)$$

Therefore, the sum of the actions of all players is invariant. Hence, if the players achieve consensus, they must agree upon the average.

The consensus algorithm imposes coupled constraints on the players' action sets by requiring the sum of the actions of all players to be invariant. In this setting, if a player unilaterally acted and altered its action, the invariance of the desired function would no longer be preserved. We will seek to replicate this approach in a game-theoretic setting by modeling the functional consensus problem as a group-based potential game.

1) *Setup—Functional Consensus Problem With Group-Based Decisions:* Consider the consensus problem with a time-invariant undirected interaction graph, as described in Section II-B. To apply the learning algorithm SAP or RSAP with group-based decisions to the functional consensus problem, one needs to define both the group utility functions and the group selection process.

2) *Group Utility Function:* We will assign any group $G \subseteq \mathcal{P}$ the following local group utility function:

$$U_G(a) = -(1/2) \sum_{\mathcal{P}_i \in G} \sum_{\mathcal{P}_j \in N_i \cap G} \|a_i - a_j\| - \sum_{\mathcal{P}_i \in G} \sum_{\mathcal{P}_j \in N_i \setminus G} \|a_i - a_j\|. \quad (84)$$

It is straightforward to show that this group utility function design results in a group-based potential game with the potential function as in (4).

3) *Group Selection Process and Action Constraints:* Let $a(t-1)$ be the action profile at time $t-1$. At time t , one player \mathcal{P}_i is randomly (uniformly) chosen. Rather than unilaterally updating its action, player \mathcal{P}_i first selects a group of players $G \subseteq \mathcal{P}$, which we will assume to be the neighbors of player \mathcal{P}_i , i.e., $G = N_i$. The group is assigned a group utility function as in (84) and a constrained action set $\mathcal{A}_G \subset \prod_{\mathcal{P}_i \in G} \mathcal{A}_i$.

A central question is how one can constrain the group action set, using only location information to preserve the invariance

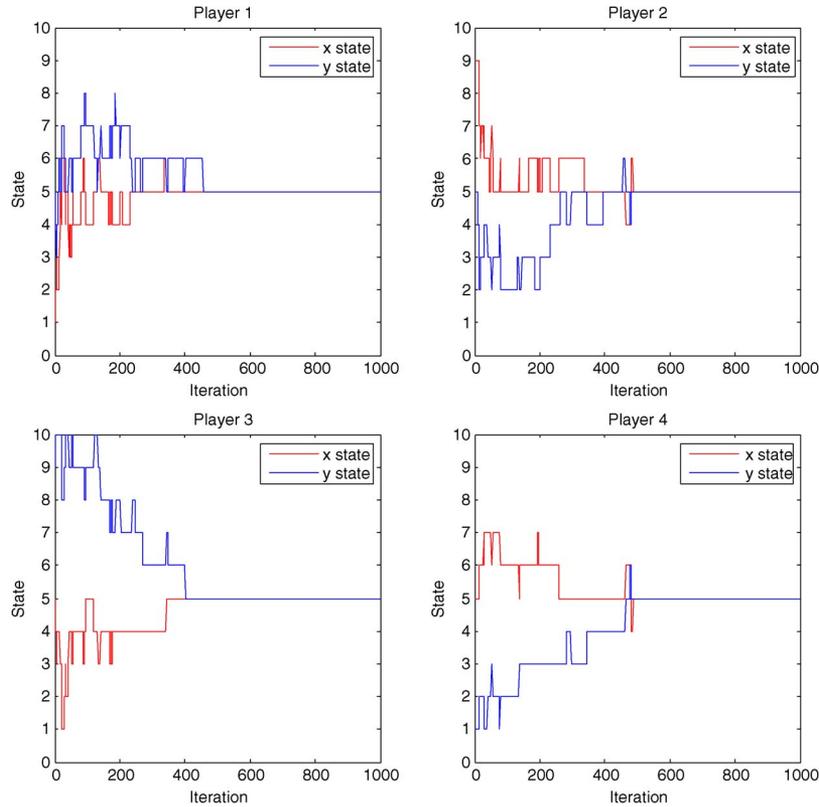


Fig. 5. Evolution of each player's action in the average consensus problem.

of the desired function f . In this case, we will restrict our attention to functions where “local” preservation is equal to “global” preservation. This means that, for each group $G \subseteq \mathcal{P}$, there exists a function $f_G : \mathcal{A}_G \rightarrow \mathbb{R}$ such that, for any group actions $a'_G, a''_G \in \mathcal{A}_G$ and $a_{-G} \in \prod_{\mathcal{P}_i \notin G} \mathcal{A}_i$

$$f_G(a'_G) = f_G(a''_G) \Rightarrow f(a'_G, a_{-G}) = f(a''_G, a_{-G}). \quad (85)$$

Examples of functions that satisfy this constraint are

$$f_G(a) = \frac{1}{|G|} \sum_{\mathcal{P}_i \in G} a_i \Rightarrow f(a) = \frac{1}{|\mathcal{P}|} \sum_{\mathcal{P}_i \in \mathcal{P}} a_i \quad (86)$$

$$f_G(a) = \max_{\mathcal{P}_i \in G} a_i \Rightarrow f(a) = \max_{\mathcal{P}_i \in \mathcal{P}} a_i \quad (87)$$

$$f_G(a) = \min_{\mathcal{P}_i \in G} a_i \Rightarrow f(a) = \min_{\mathcal{P}_i \in \mathcal{P}} a_i. \quad (88)$$

In each of these examples, the structural forms of f and f_G are equivalent. There may exist alternative functions where this is not required.

4) *Illustration*: We will illustrate this approach by solving the average consensus problem of the example developed in Section III-C. Given the initial configuration, all players should agree upon the action (5, 5). We will solve this average consensus problem using the learning algorithm binary RSAP with group-based decisions, where the group restricted action set satisfies $R_G(a_G) = \mathcal{A}_G \cap (\prod_{\mathcal{P}_i \in G} R_i(a_i))$. However, we will omit the nonconvex obstruction in this illustration. This omission is not necessary but convenient for not having to verify that consensus is possible, given the initial conditions and the constrained action sets.

Fig. 5 shows the evolution of each player's actions using the stochastic learning algorithm binary RSAP with group-based decisions and an increasing β coefficient $\beta(t) = 1.5 + t(2/1000)$.

VII. CONCLUSION

We have proposed a game-theoretic approach to cooperative control by highlighting the connection between cooperative control problems and potential games. We have introduced a new class of games and enhanced existing learning algorithms to broaden the applicability of game-theoretic methods to cooperative control setting. We have 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, 4.1, and 5.1 and the notion of a sometimes weakly acyclic game are applicable to a broader class of games and other cooperative control problems, such as the sensor deployment problem or the dynamic sensor allocation problem.

REFERENCES

- [1] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Trans. Autom. Control*, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.
- [2] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. Tsitsiklis, “Convergence in multiagent coordination, consensus, and flocking,” in *Proc. 44th IEEE Conf. Decision Control*, Dec. 2005, pp. 2996–3000.
- [3] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
- [4] L. Xiao and S. Boyd, “A scheme for robust distributed sensor fusion based on average consensus,” in *Inf. Process. Sensor Netw.*, 2005, pp. 63–70.

- [5] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
- [6] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.
- [7] L. Moreau, "Stability of continuous-time distributed consensus algorithms," in *Proc. 43rd IEEE Conf. Decision 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. Autom. Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.
- [9] A. Kashyap, T. Basar, and R. Srikant, "Consensus with quantized information updates," in *Proc. 45th IEEE Conf. Decision Control*, 2006, pp. 2728–2733.
- [10] R. Murray, "Recent research in cooperative control of multivehicle systems," *Trans. ASME, J. Dyn. Syst. Meas. Control*, vol. 129, no. 5, pp. 571–583, Sep. 2007.
- [11] J. Shamma, Ed., *Cooperative Control of Distributed Multi-Agent Systems*. Hoboken, NJ: Wiley-Interscience, 2008.
- [12] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*, ser. Applied Mathematics Series. Princeton, NJ: Princeton Univ. Press, 2008.
- [13] H. P. Young, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton Univ. Press, 1998.
- [14] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
- [15] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Auton. Agents Multi-Agent Syst.*, vol. 11, no. 3, pp. 387–434, Nov. 2005.
- [16] D. Wolpert and K. Tumor, "An overview of collective intelligence," in *Handbook of Agent Technology*, J. M. Bradshaw, Ed. Cambridge, MA: MIT Press, 1999.
- [17] D. Monderer and L. Shapley, "Potential games," *Games Econom. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.
- [18] D. Monderer and L. Shapley, "Fictitious play property for games with identical interests," *J. Econ. Theory*, vol. 68, no. 1, pp. 258–265, Jan. 1996.
- [19] D. Monderer and A. Sela, *Fictitious Play and No-Cycling Conditions*, 1997. [Online]. Available: <http://www.sfb504.uni-mannheim.de/publications/dp97-12.pdf>
- [20] J. R. Marden, G. Arslan, and J. S. Shamma, "Regret based dynamics: Convergence in weakly acyclic games," in *Proc. Int. Conf. AAMAS*, Honolulu, HI, May 2007, pp. 194–201.
- [21] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 208–220, Feb. 2009.
- [22] G. Arslan, J. R. Marden, and J. S. Shamma, "Autonomous vehicle-target assignment: A game theoretical formulation," *Trans. ASME, J. Dyn. Syst. Meas. Control*, vol. 129, no. 5, pp. 584–596, Sep. 2007.
- [23] H. P. Young, *Strategic Learning and its Limits*. London, U.K.: Oxford Univ. Press, 2005.
- [24] L. S. Shapley, "Stochastic games," *Proc. Nat. Acad. Sci. U. S. A.*, vol. 39, no. 10, pp. 1095–1100, Oct. 1953.
- [25] L. Blume, "The statistical mechanics of strategic interaction," *Games Econom. Behav.*, vol. 5, no. 3, pp. 387–424, Jul. 1993.
- [26] L. Blume, "Population games," in *The Economy as an Evolving Complex System II*, B. Arthur, S. Durlauf, and D. Lane, Eds. Reading, MA: Addison-Wesley, 1997, pp. 425–460.
- [27] J. R. Marden, H. P. Young, G. Arslan, and J. S. Shamma, "Payoff based dynamics for multi-player weakly acyclic games," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 373–396, 2009.
- [28] W. Li and C. G. Cassandras, "Sensor networks and cooperative control," *Eur. J. Control*, vol. 11, no. 4/5, pp. 436–463, 2005.
- [29] W. Heinzelman, "Application-specific protocol architectures for wireless networks," Ph.D. dissertation, MIT, Cambridge, MA, 2000.



Jason R. Marden received the B.S. and Ph.D. degrees in mechanical engineering from the University of California, Los Angeles, in 2001 and 2007, respectively.

Since 2007, he has been a Junior Fellow with the Social and Information Sciences Laboratory, California Institute of Technology, Pasadena. His research interests include game-theoretic methods for feedback control of distributed multiagent systems.



Gürdal Arslan received the Ph.D. degree in electrical engineering from the University of Illinois, Urbana, in 2001.

From 2001 to 2004, he was an Assistant Researcher with the Department of Mechanical and Aerospace Engineering, University of California, Los Angeles. In August 2004, he joined the Department of Electrical Engineering, University of Hawaii, Manoa, where he is currently an Associate Professor. His current research interests include the design of cooperative multiagent systems using

game-theoretic methods.

Dr. Arslan was the recipient of the National Science Foundation CAREER Award in May 2006.



Jeff S. Shamma received the B.S. degree from Georgia Institute of Technology (Georgia Tech), Atlanta, in 1983 and the Ph.D. degree from Massachusetts Institute of Technology, Cambridge, in 1988, both in mechanical engineering.

He has held faculty positions at the University of Minnesota, Minneapolis; the University of Texas, Austin; and the University of California, Los Angeles. In 2007, he returned to Georgia Tech, where he is currently a Professor of electrical and computer engineering and the Julian T. Hightower

Chair of Systems and Controls with the School of Electrical and Computer Engineering.