



State based potential games[☆]

Jason R. Marden¹

Department of Electrical, Computer and Energy Engineering, 425 UCB, Boulder, CO 80309, United States

ARTICLE INFO

Article history:

Received 31 May 2011

Received in revised form

27 April 2012

Accepted 13 July 2012

Available online 23 September 2012

Keywords:

Game theory

Distributed control

Resource allocation

ABSTRACT

There is a growing interest in the application of game theoretic methods to the design and control of multiagent systems. However, the existing game theoretic framework possesses inherent limitations with regards to these new prescriptive challenges. In this paper we propose a new framework, termed *state based potential games*, which introduces an underlying state space into the framework of potential games. This state space provides a system designer with an additional degree of freedom to help coordinate group behavior and overcome these limitations. Within the context of state based potential games, we characterize the limiting behavior of two learning algorithms termed *finite memory better reply processes* and *log-linear learning*. Lastly, we demonstrate the applicability of state based potential games on two cooperative control problems pertaining to distributed resource allocation and the design of local and distributed control laws.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Many engineering systems can be characterized by a collection of interacting subsystems each making local decisions in response to local information. Existing wind farms are one example of such systems where individual wind turbines make local control decisions in response to local wind conditions (Pao & Johnson, 2009). Alternative systems include autonomous vehicles for surveillance missions (Marden & Wierman, 2008; Martinez, Cortes, & Bullo, 2007; Murphey, 1999) or the routing of information through a network (Katti, Katabi, Hu, Rahul, & Médard, 2005; Roughgarden, 2005). Regardless of the specific application domain, the primary goal in such systems is to design local control policies for the individual subsystems to ensure that the emergent collective behavior is desirable with respect to the system level objective.

Developing an underlying theory for the design and control of multiagent systems remains a central goal for the field of distributed control. One of the core challenges in realizing this goal is that each application domain possesses inherent constraints that must be accounted for in the design process. For example, control strategies for individual turbines in a wind farm are constrained

by the lack of a suitable communication system and the fact that the aerodynamic interaction between the turbines is poorly understood (Pao & Johnson, 2009). Accordingly, most of the existing research in distributed control focuses on specific applications as opposed to an underlying theory for distributed control, e.g., consensus and flocking (Olfati-Saber, Fax, & Murray, 2007; Tsitsiklis, 1987), sensor coverage (Martinez et al., 2007; Murphey, 1999), and routing information over networks (Roughgarden, 2005), among many others.

Game theory is beginning to emerge as a valuable paradigm for the design and control of such multiagent systems (Arslan, Marden, & Shamma, 2007; Gopalakrishnan, Marden, & Wierman, 2011; Marden, Arslan, & Shamma, 2009a). Utilizing game theory for this purpose requires the following two step design process: (i) define the interaction framework of the agents within a game theoretic environment (game design) and (ii) define local decision making rules that specify how each agent processes available information to formulate a decision (learning design). The goal is to complete both steps (i) and (ii) to ensure that the collective behavior converges to a desirable operating point, e.g., a pure Nash equilibrium of the designed game. One of the major appeals of using game theory for multiagent systems is that game theory provides a hierarchical decomposition between the design of the interaction framework and the design of the learning rules (Gopalakrishnan et al., 2011). For example, if the interaction framework is designed as a *potential game* (Monderer & Shapley, 1996) then any learning algorithm for potential games can be utilized as a distributed control algorithm with provable guarantees on the emergent collective behavior. This decomposition could be instrumental in shifting research attention from application specific designs to an underlying theory for

[☆] This research was supported by AFOSR grants #FA9550-09-1-0538 and #FA9550-12-1-0359 and ONR grant #N00014-12-1-0643. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Ian R. Petersen.

E-mail address: jason.marden@colorado.edu.

¹ Tel.: +1 303 492 5867; fax: +1 303 492 2758.

multiagent systems. The key to this realization is the development of a broad set of tools for both game design and learning design that could be used to address these application specific challenges. Hence, a system designer could appeal to available tools in a “plug-and-play” approach to complete the design process while satisfying the inherent constraints of the application at hand.

Several recent papers focus on identifying the viability of the framework of potential games as a mediating layer for this decomposition (Li & Marden, 2010; Marden & Wierman, 2008, 2009). With a broad set of existing results for learning in potential games (Arslan & Shamma, 2004; Fudenberg & Levine, 1998; Marden, Arslan, & Shamma, 2009b, 2007; Marden, Young, Arslan, & Shamma, 2009; Shamma & Arslan, 2005; Young, 1998, 2005, 1993), the primary focus of this work is on the development of methodologies for designing the interaction framework as a potential game while meeting constraints and objectives relevant to multiagent systems, e.g., locality of agent objective functions, and efficiency guarantees for resulting equilibria, among many others. Unfortunately, the framework of potential games is not broad enough to meet this diverse set of challenges as several limitations are beginning to emerge. One such limitation involves multiagent systems where the desired collective behavior must satisfy a given coupled constraint on the agents’ actions, e.g., satisfy a desired signal to noise ratio in a wireless communication system (Li & Marden, 2010). It turns out that it is theoretically impossible to model such systems as a potential game where all resulting Nash equilibria optimize the system level objective while at the same time satisfying the given constraint (Li & Marden, 2010). Alternative limitations focus on the derivation of budget-balanced agent objective functions in networked cost sharing problems (Chen, Roughgarden, & Valiant, 2008; Marden & Wierman, 2009) and the derivation of local agent objective functions for multiagent coordination in distributed engineering systems (Li & Marden, 2011). These inherent limitations provide the analytical justification for moving beyond potential games to games of a broader structure.

The first contribution of this paper is the formalization of a new game structure, termed *state based potential games*, as a new mediating layer for this decomposition. State based potential games are an extension of potential games where there is an underlying state space introduced into the game theoretic environment. The novelty of the proposed developments is the use of this enhanced framework from a design perspective where the goal is to attain a dynamical process which yields efficient global behavior in distributed engineering systems. Here, the underlying state can take on a variety of interpretations ranging from the introduction of dummy agents to dynamics for equilibrium selection. Regardless of the interpretation, the state provides a system designer with an additional degree of freedom to help coordinate group behavior. The framework of state based potential games is rich enough to overcome the aforementioned limitations as highlighted in Section 7. Interestingly, state based potential games can be thought of in a complimentary fashion to recent work in distributed learning algorithms (Pradelski & Young, 2012; Shamma & Arslan, 2005; Young, 2009) where an underlying state space is introduced into the learning environment to help coordinate behavior. For example, in Young (2009) the authors introduce moods for each agent that impacts the agent’s behavior. The authors utilize these moods (or states) to develop a payoff based learning algorithm that converges to a pure Nash equilibrium in any game where one such equilibrium exists.

The second contribution of this paper focuses on the development of learning algorithms for state based potential games. In particular, we focus on extending two specific classes of learning algorithms for potential games to state based potential games. The first algorithm that we consider is a finite memory better reply process which converges to a (pure) Nash equilibrium in any potential

game (Young, 2005). Here, a finite memory better reply process classifies virtually any decision making rule where agents seek to improve their one-shot payoff. This result demonstrates that potential games possess an inherent robustness to decision making rules as any form of “reasonable” behavior will converge to a Nash equilibrium. This result suggests that behavior will also converge to a Nash equilibrium even when agent behavior is corrupted by any number of issues pertinent to engineering systems including delays, inconsistent clock rates, or inaccuracies in information. Our first result in Theorem 4.2 demonstrates that such robustness is also present in state based potential games. In particular, any finite memory better reply process converges to a suitably defined equilibrium of the state based potential game.

Our next result focuses on the learning algorithm log-linear learning (Alos-Ferrer & Netzer, 2010; Blume, 1993, 1997; Marden & Shamma, 2012; Shah & Shin, 2010). In potential games, log-linear learning provides guarantees on the percentage of time that the joint action profile will be at a maximizer of the potential function. One of the appeals of log-linear learning for multiagent systems is the fact that there are existing methodologies for designing agent utility functions that guarantee (i) that the resulting game is a potential game and (ii) that the action profiles that maximize the system level objective coincide with the potential function maximizers (Arslan et al., 2007; Marden & Wierman, 2008; Wolpert & Tumor, 1999). Our second result in Theorem 5.1 demonstrates that a variant of log-linear learning also provides similar analytical guarantees in state based potential games. Hence, both Theorems 4.2 and 5.1 begin the process of establishing diverse learning algorithms for this new decomposition with state based potential games as the mediating layer.

Lastly, we illustrate the developments in this paper on two classes of cooperative control problems. In Section 7.1 we focus on the problem of distributed resource allocation problems and demonstrate that the framework of state based potential games coupled with the extension of log-linear learning in Theorem 5.1 can be utilized to achieve efficient system behavior. In Section 7.2 we extensively review the results in Li and Marden (2011) which develops a systematic methodology for the design of local agent utility functions that ensures both the existence and efficiency of the resulting equilibria provided that the system level objective is concave. By local, we mean that each agent’s control policy can only depend on limited information regarding the behavior of other agents. This methodology yields a state based potential game; hence, the control design can be completed by appealing to any available learning algorithm for state based potential games that guarantees convergence to an equilibrium.

2. Preliminaries

2.1. Background: finite strategic form games

We consider finite strategic form games with n agents denoted by the set $N := \{1, \dots, n\}$. Each agent $i \in N$ has a finite action set \mathcal{A}_i and a utility function $U_i : \mathcal{A} \rightarrow \mathbb{R}$ where $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ denotes the joint action set. We refer to a finite strategic-form game as “a game”, and we sometimes use a single symbol, e.g., G , to represent the entire game, i.e., the agent set, N , action sets, \mathcal{A}_i , and cost functions U_i . We denote a game G as a tuple $G = (N, \{\mathcal{A}_i\}, \{U_i\})$ as we omit the subscript on the actions sets and utility functions for brevity, i.e., $\{\mathcal{A}_i\} = \{\mathcal{A}_i\}_{i \in N}$. For an action profile $a = (a_1, a_2, \dots, a_n) \in \mathcal{A}$, let a_{-i} denote the profile of agent actions other than agent i , i.e., $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$. With this notation, we sometimes write a profile a of actions as (a_i, a_{-i}) . Similarly, we may write $U_i(a)$ as $U_i(a_i, a_{-i})$. We also use $\mathcal{A}_{-i} = \prod_{j \neq i} \mathcal{A}_j$ to denote the set of possible collective actions of all agents other than agent i .

In this paper we focus on analyzing equilibrium behavior in distributed systems in addition to distributed learning algorithms for attaining such equilibria. The most well known form of an equilibrium is the Nash equilibrium defined as follows:

Definition 2.1 (Pure Nash Equilibrium). An action profile $a^* \in \mathcal{A}$ is called a (pure) Nash equilibrium if for each agent $i \in N$,

$$U_i(a_i^*, a_{-i}^*) = \max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}^*). \quad (1)$$

A (pure) Nash equilibrium represents a scenario for which no agent has an incentive to unilaterally deviate.

One class of games that plays a prominent role in engineering multiagent systems is that of potential games (Monderer & Shapley, 1996). In a potential game, the change in an agent's utility that results from a unilateral change in strategy equals the change in a global potential function. Definition 2.2 makes this idea precise.

Definition 2.2 (Potential games). A game $G = (N, \{\mathcal{A}_i\}, \{U_i\})$ is called an (exact) potential game if there exists a global function $\phi : \mathcal{A} \rightarrow \mathbb{R}$ such that for every agent $i \in N$, for every $a_{-i} \in \mathcal{A}_{-i}$, and for every $a_i', a_i'' \in \mathcal{A}_i$,

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

The global function ϕ is called the potential function for game G .

In potential games, any action profile maximizing the potential function is a Nash equilibrium, hence every potential game possesses at least one such equilibrium.

2.2. Learning in games

The focus of this paper is on the attainment of dynamical processes which converge to a desirable system-wide behavior in distributed systems. To that end, we consider a repeated one-shot game where at each time $t \in \{0, 1, 2, \dots\}$ each agent $i \in N$ independently chooses an action $a_i(t) \in \mathcal{A}_i$ and receives the payoff $U_i(a(t))$ where $a(t) := (a_1(t), \dots, a_n(t))$. Each agent i chooses an action $a_i(t)$ simultaneously according to the agent's strategy at time t denoted as $p_i(t) \in \Delta(\mathcal{A}_i)$, where $\Delta(\mathcal{A}_i)$ is the set of probability distributions over the finite set of actions \mathcal{A}_i . We use $p_i^{a_i}(t)$ to represent the probability that agent i selects action a_i at time t ; thus $\sum_{a_i \in \mathcal{A}_i} p_i^{a_i}(t) = 1$. An agent's strategy at time t can rely only on observations from the games played at times $\{0, 1, 2, \dots, t-1\}$. Different learning algorithms are specified by both the assumptions on available information and the mechanism by which the strategies are updated as information is gathered. In the most general form, the strategy adjustment mechanism of agent i can be written as

$$p_i(t) = F_i(a(0), \dots, a(t-1); U_i) \quad (2)$$

meaning that an agent's strategy is conditioned on all previous action profiles in addition to the structural form of the agent's utility function. In potential games, there are a wide array of distributed learning algorithms that converge to a (pure) Nash equilibrium (Arslan & Shamma, 2004; Blume, 1993, 1997; Fudenberg & Levine, 1998; Marden et al., 2009b; Marden, Younget al., 2009; Shamma & Arslan, 2005; Young, 1998, 2005, 1993).

3. State based games

In this paper we consider an extension to the framework of finite strategic form games, termed *state based games*, which introduces an underlying state space to the game theoretic

framework. State based games represent a simplification of the class of Markov games (Shapley, 1953). In the proposed state based games we focus on myopic agents and static equilibrium concepts similar to that of Nash equilibrium. The state is introduced as a coordinating entity used to improve system level behavior. To preserve the generality of the state based game architecture the following definitions impose no restrictions on the admissible state space and transition functions. Different applications will impose such restrictions as demonstrated in Section 7.

3.1. Definition of state based games

In state based games there exists a set of agents N and an underlying finite state space X . Each agent $i \in N$ has a state invariant action set \mathcal{A}_i and a state dependent utility function $U_i : \mathcal{A} \times X \rightarrow \mathbb{R}$ where \mathcal{A} is the set of joint actions.² Lastly, there is a Markovian state transition function $P : \mathcal{A} \times X \rightarrow \Delta(X)$ where $\Delta(X)$ denotes the set of probability distributions over the finite state space X . We denote a state based game G by the tuple $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, X, P\}$.

Repeated play of a state based game produces a sequence of action profiles $a(0), a(1), \dots$, and a sequence of states $x(0), x(1), \dots$, where $a(t) \in \mathcal{A}$ is referred to as the action profile at time t and $x(t) \in X$ is referred to as the state at time t . The sequence of actions and states is generated according to the following process: at any time $t \geq 0$, each agent $i \in N$ myopically selects an $a_i(t) \in \mathcal{A}_i$ according to some specified decision rule of the general form

$$a_i(t) = F_i(a(0), \dots, a(t-1), x(0), \dots, x(t); U_i), \quad (3)$$

meaning that an agent's action choice is potentially conditioned on all previous action profiles, all previous states, and the structural form of the agent's utility function. For example, if an agent used a myopic Cournot adjustment process for (3) then

$$a_i(t) \in \arg \max_{a_i \in \mathcal{A}_i} U_i(a_i, a_{-i}(t-1), x(t)). \quad (4)$$

The state $x(t)$ and the joint action profile $a(t) := (a_1(t), \dots, a_n(t))$ together determine each agent's payoff $U_i(a(t), x(t))$ at time t . After all agents select their respective action, the ensuing state $x(t+1)$ is chosen according to the state transition function $x(t+1) \sim P(a(t), x(t))$ and the process is repeated. Here we use the symbol " \sim " to signify that the state is selected randomly according to the probability distribution $P(a(t), x(t))$.

3.2. Equilibrium for state based games

We are interested in analyzing the equilibrium points of the Cournot adjustment process in (4) for state based games. Before formally defining this equilibrium concept we introduce the notion of reachable states. For an action state pair $[a^0, x^0]$, the set of reachable states by an action invariant state trajectory a^0 is defined as $X(a^0|x^0) \subseteq X$ where a state $x \in X(a^0|x^0)$ if and only if there exists a time $t > 0$ such that

$$\Pr[x(t) = x] > 0$$

conditioned on the events $x(0) = x^0$ and $x(k+1) \sim P(a^0, x(k))$ for all $k \in \{0, 1, \dots, t-1\}$. Roughly speaking, the set $X(a^0|x^0)$ represents all states that may eventually emerge along an action invariant state trajectory starting from $[a^0, x^0]$. Accordingly, we now define the notion of recurrent state equilibria as follows:

² One could also permit state dependent action sets where the set of available actions for agent i given the state x is $\mathcal{A}_i^x \subseteq \mathcal{A}_i$. However, such genericity is not needed for the results in this paper.

Definition 3.1 (Recurrent State Equilibrium). The action state pair $[a^*, x^*]$ is a recurrent state equilibrium with respect to the state transition process $P(\cdot)$ if the following two conditions are satisfied:

- (i) The state x^* satisfies $x^* \in X(a^*|x)$ for every state $x \in X(a^*|x^*)$.³
- (ii) For every agent $i \in N$ and every state $x \in X(a^*|x^*)$,

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

3.3. State based potential games

A recurrent state equilibrium may or may not exist for a given state based game. We next provide a generalization of potential games to state based games which guarantee the existence of such an equilibrium. Much like potential games, the following class of games imposes an equivalence between differences in an agent's utility and a global potential function resulting from a unilateral deviation.

Definition 3.2 (State Based Potential Games). A state based game $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, X, P\}$ is a state based potential game if there exists a potential function $\phi : \mathcal{A} \times X \rightarrow \mathbb{R}$ that satisfies the following two properties for every action state pair $[a, x] \in \mathcal{A} \times X$:

- (i) For any agent $i \in N$ and action $a'_i \in \mathcal{A}_i$

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

- (ii) For any state x' in the support of $P(a, x)$

$$\phi(a, x') \geq \phi(a, x).$$

The second condition states that the potential function is non-decreasing along any action invariant state trajectory. The first condition states that each agent's utility function is directly aligned with the potential function for all possible unilateral deviations from all action state pairs.⁴ Relaxing our requirement on equality between the change in utility and the change in the potential function in Condition (i) or Definition 3.2 gives rise to the class of *ordinal state based potential games* where Condition (i) is now of the form

$$\begin{aligned} U_i(a'_i, a_{-i}, x) - U_i(a, x) &> 0 \\ \Rightarrow \\ \phi(a'_i, a_{-i}, x) - \phi(a, x) &> 0. \end{aligned}$$

The following lemma proves the existence of a recurrent state equilibrium in any ordinal state based potential game.

Lemma 3.1. *A recurrent state equilibrium exists in any ordinal state based potential game.*⁵

³ Alternatively, consider the Markov chain induced by the transitional probabilities P with a fixed action profile a^* which we denote by $P(a = a^*, \cdot)$. This process is a Markov chain over the states X . Condition (i) implies that the state x^* is recurrent given the process $P(a = a^*, \cdot)$ and an initial state $x(0) = x^*$.

⁴ The condition that “for every action state pair $[a, x] \in \mathcal{A} \times X$ ” can be relaxed by exploiting the properties of the state transition function $P(\cdot)$. For a given state x , the state transition function $P(\cdot)$ reveals all action state pairs that could have generated x , thereby restricting the class of unilateral deviations to consider. For example, if $x(t)$ is the state at time t then the state $x(t)$ must lie within the support of $P(a(t-1), x(t-1))$. This restricts the set of possible action state pairs that could have been realized at time $t-1$ and hence restricts the set of unilateral deviations to consider which are of the form $(a_i, a_{-i}(t-1))$. To that end, define $V(a, x) \subseteq X$ as the support of the distribution $P(a, x)$ and $X(a) := \cup_{x \in X} V(a, x) \subseteq X$ as the set of states that are reachable with the action profile a . Note that $X(a|x) \subseteq X(a)$. We can now replace the statement “for every action state pair $[a, x] \in \mathcal{A} \times X$ ” to “for every action profile $a \in \mathcal{A}$ and state $x \in X(a)$ ”.

⁵ It is important to note that this lemma does not make any claims regarding the uniqueness of recurrent state equilibrium in ordinal state based potential games. Consequently, in Section 5 we focus on learning algorithms which provide equilibrium selection in ordinal state based potential games.

Proof. Let $[a^*, x^*] \in \arg \max_{[a, x] \in \mathcal{A} \times X} \phi(a, x)$ be an action state pair that maximizes the potential function. Let X^* represent the recurrent states of the process $P(a = a^*, \cdot)$ starting from the initial state $x(0) = x^*$ which is by definition nonempty. Choose any state $x^{**} \in X^*$. The action state pair $[a^*, x^{**}]$ is a recurrent state equilibrium. \square

3.4. Connection to Markov games

As noted previously, the proposed state based games represent a simplification to the framework of Markov games (Shapley, 1953). The key difference between Markov games and state based games is the evaluation of future rewards. In Markov games, each agent seeks to optimize a discounted sum of future payoffs through the use of stationary strategies of the form $\pi_i^S : X \rightarrow \Delta(\mathcal{A}_i)$. Accordingly, the utility of agent i given an initial state $x(0)$ and a collection of stationary strategies $\pi^S = (\pi_1^S, \dots, \pi_n^S)$ is denoted by $\mathcal{U}_i(\pi_i^S, \pi_{-i}^S | x(0))$ which is defined as

$$\mathbb{E} \left[\sum_{k=0}^{\infty} \delta^k \cdot U_i(\pi_i^S(x(k)), \pi_{-i}^S(x(k)), x(k)) \right]$$

where $1 > \delta > 0$ is the discount factor associated with future payoff, $x(k+1) \sim P(a(k), x(k))$, and the expectation is evaluated over the probability distributions induced by both the agents' stationary strategies π^S and the state transition function P . A collection of stationary strategies $\pi^* = (\pi_1^*, \dots, \pi_n^*)$ constitutes an equilibrium if the following condition is satisfied: for any state $x \in X$, agent $i \in N$, and stationary strategy π_i^S

$$\mathcal{U}_i(\pi_i^*, \pi_{-i}^* | x) \geq \mathcal{U}_i(\pi_i^S, \pi_{-i}^* | x).$$

This is in contrast to the proposed recurrent state equilibria which focus on myopic agents ($\delta = 0$) and characterizing action state pairs as opposed to a collection of stationary strategies. For a more complete characterization of Markov games we direct the readers to Shapley (1953).

4. Learning in state based potential games

Having established the existence of a recurrent state equilibrium in any (ordinal) state based potential game, we now seek to identify how agents can learn to play such an equilibrium in a distributed fashion. We adopt the traditional framework of learning in games with the sole exception that the underlying game is not static because of the dynamic nature of the state.

4.1. Better reply with inertia dynamics

Learning Nash equilibria in potential games is inherently robust. This theme is encapsulated in Young (2005) which proves that any “finite memory better reply process” converges to an equilibrium in potential games.⁶ Informally, a finite memory better reply process refers to any process in which agents seek to improve their one-shot payoff. This robustness property is of fundamental importance to distributed engineering systems where the agents' update rules (or control strategies) are inherently heterogeneous stemming from delays or corruption in information, varying degrees of accessible information, and variations in clock rates amongst the agents which implies that the agents do not always update their actions each iteration. The result in Young (2005)

⁶ This result actually holds for a broader class of games known as weakly acyclic games.

demonstrates that such levels of heterogeneity will not impact the asymptotic behavior of such learning algorithms.

In this section we seek to replicate these results for state based potential games thereby demonstrating that state based potential games also possess this robustness property. To that end, we consider the “better reply with inertia” dynamics for state based games. Define an agent’s *strict better reply set* for any action state pair $[a, x] \in \mathcal{A} \times X$ as

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

The better reply with inertia dynamics can be described as follows. At each time $t > 0$, each agent $i \in N$ selects an action at time t according to a strategy $p_i \in \Delta(\mathcal{A}_i)$ defined as follows: if $B_i(a(t-1); x(t)) = \emptyset$ then

$$p_i^{a_i(t-1)} = 1. \quad (5)$$

Otherwise, if $B_i(a(t-1); x(t)) \neq \emptyset$ then

$$p_i^{a_i} = \epsilon, \quad \text{for } a_i = a_i(t-1), \quad (6)$$

$$p_i^{a'_i} = \frac{(1-\epsilon)}{|B_i(a(t-1); x(t))|}, \quad \forall a'_i \in B_i(a(t-1); x(t)),$$

$$p_i^{a''_i} = 0, \quad \forall a''_i \notin B_i(a(t-1); x(t)) \cup \{a_i(t-1)\},$$

for some constant $\epsilon \in (0, 1)$ referred to as the agent’s inertia.

Theorem 4.1. *Let $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, X, P\}$ be an ordinal state based potential game with potential function $\phi : \mathcal{A} \times X \rightarrow \mathbb{R}$. If all agents adhere to the better reply with inertia dynamics then the action state pair converges almost surely to an action invariant set of recurrent state equilibria.*

We omit the proof as the better reply with inertia dynamics is a special case of the finite memory better reply with inertia dynamics proved in [Theorem 4.2](#) of the ensuing section.

4.2. Finite memory better reply with inertia dynamics

We consider a finite memory better reply process with inertia for state based games. Let $[a(t-1), x(t-1)]$ be the action state pair at time $t-1$ and let $x(t)$ be the ensuing state selected according to the distribution $P(a(t-1), x(t-1))$. In this setting, each agent selects an action at time t using information from the past m action state pairs, i.e., $[a(t-m), x(t-m)], \dots, [a(t-1), x(t-1)]$, and the current state $x(t)$, where m is referred to as the memory length.

Define the m -step action history at time t as $h_a^m(t) := \{a(t-m), \dots, a(t-1)\}$, the m -step state history $h_x^m(t) := \{x(t-m), \dots, x(t-1)\}$, and the tuple as $h^m(t) := [h_a^m(t), h_x^m(t)]$. A finite memory better response function is a set valued function of the form $B_i^m : (\mathcal{A} \times X)^m \times X \rightarrow 2^{\mathcal{A}_i}$ where $2^{\mathcal{A}_i}$ denotes the power set of \mathcal{A}_i . We refer to the range of B_i^m as the (finite memory) better response set. A function B_i^m is a finite memory better response function if it satisfies the following two conditions:

- (i) If an action is in agent i ’s finite memory better response set, then it must be a better response for some admissible action state pair in the agent’s history, i.e., if $a_i \in B_i^m(h^m(t); x(t))$ then $a_i \in B_i(\tilde{a}; \tilde{x})$ for some $\tilde{a} \in h_a^m(t)$ and $\tilde{x} \in \{h_x^m(t), x(t)\}$.
- (ii) If an action a_i is a better response for every action profile in the history h^m given the current state $x(t)$, i.e., $a_i \in B_i(\tilde{a}; x(t))$ for all $\tilde{a} \in h_a^m(t)$, then the action a_i must be in the finite memory better response set, i.e., $a_i \in B_i^m(h^m(t), x(t))$.

Notice that the better response function used in the previous section satisfies both of these conditions. The finite memory better reply with inertia dynamics is identical to the better reply with inertia dynamics described in the previous section with the sole exception that the strategy highlighted in (5) and (6) now utilizes the finite memory better response set, $B_i^m(\cdot)$, as opposed to the better response set, $B_i(\cdot)$.

Theorem 4.2. *Let $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, X, P\}$ be an ordinal state based potential game with potential function $\phi : \mathcal{A} \times X \rightarrow \mathbb{R}$. If all agents adhere to the finite memory better reply with inertia dynamics then the action state pair converges almost surely to an action invariant set of recurrent state equilibria.*

Proof. Before starting the proof, we point out an observation regarding recurrent state equilibria. For any time $t^0 > 0$, if (i) the action state pair $[a(t^0), x(t^0)]$ is a recurrent state equilibrium, (ii) the action $a(t^0)$ is repeated for m timesteps, i.e., $a(t^0) = \dots = a(t^0 + m)$, and (iii) all agents adhere to any finite memory better reply process, then for any time $t^1 \geq t^0 + m$ the action state pair $[a(t^1), x(t^1)]$ will be a recurrent state equilibrium and furthermore $a(t^1) = a(t^0)$. Therefore, in order to prove the theorem, it suffices to show that there exists a finite number of timesteps $T > 0$ and a finite constant $\gamma > 0$ such that the following is true: for any time $t^0 > 0$ and action state pair $[a(t^0), x(t^0)]$, if all agents adhere to any finite memory better reply process with inertia, then $[a(t^0 + T - m), x(t^0 + T - m)]$ will be a recurrent state equilibrium and $a(t^0 + T - m) = \dots = a(t^0 + T)$ with at least probability γ .

The structure of the proof relies on two types of behavior: (i) saturate the memory and (ii) unilateral optimization.

Step 1: saturate the memory. Let $[a^0, x^0] := [a(t^0), x(t^0)]$ be the action state pair at time t^0 . Suppose the action profile a^0 is repeated m consecutive periods, i.e., $a(t^0) = \dots = a(t^0 + m)$, and the state $x(k+1) \sim P(a^0, x(k))$ for all $k \in \{t^0, \dots, t^0 + m\}$. This event happens with at least probability ϵ^{nm} because of agents’ inertia. Let $t^1 := t^0 + m$. If $[a^0, x^0]$ is a recurrent state equilibrium then $a(t) = a^0$ for all times $t \geq t^1$ and we are done.

Step 2: unilateral optimization. Suppose $[a^0, x^0]$ is not a recurrent state equilibrium. If $[a^0, x(t^1)]$ is a recurrent state equilibrium, then we can repeat the above argument and we are done. If $[a^0, x(t^1)]$ is not a recurrent state equilibrium, then there exists an agent $i \in N$ with an action $a'_i \in \mathcal{A}_i$ for some state $x' \in X(a^0|x(t^1))$ such that $U_i(a'_i, a_{-i}^0, x') > U_i(a^0, x')$ and consequently $\phi(a'_i, a_{-i}^0, x') > \phi(a^0, x')$. Since $x' \in X(a^0|x(t^1))$, there exists a time $t^2 \in \{t^1, \dots, t^1 + |X|\}$ such that

$$\Pr[x(t^2) = x'] > \delta > 0$$

conditioned on the events $x(t^1), a(t^1) = a(t^1 + 1) = \dots = a(t^2 - 1) = a^0$.⁷ The event $x(t^2) = x'$ happens with at least probability $\delta \epsilon^{n|X|}$ because of the agents’ inertia. By Condition (ii) of the finite memory better response function, $a'_i \in B_i^m(h^m(t^2); x')$. Conditioned on the event $x(t^2) = x'$, the action profile $a^1 := (a'_i, a_{-i}^0)$ will be played at time t^2 with at least probability $\epsilon^{n-1}(1-\epsilon)/|\mathcal{A}_i|$, i.e., all agents other than agent i play their current action because of inertia and agent i selects action a'_i from his finite memory better response set. Note that $\phi(a^1, x(t^2)) > \phi(a^0, x(t^2))$ because

$$\begin{aligned} U_i(a'_i, a_{-i}^0, x(t^2)) - U_i(a^0, x(t^2)) &> 0 \\ \Rightarrow \\ \phi(a'_i, a_{-i}^0, x(t^2)) - \phi(a^0, x(t^2)) &> 0. \end{aligned}$$

Furthermore, note that $\phi(a^0, x(t^2)) \geq \phi(a^0, x^0)$ because of Condition (ii) of ordinal state based potential games, i.e., the potential function is nondecreasing along any action invariant state trajectory. Therefore, $\phi(a^1, x(t^2)) > \phi(a^0, x^0)$.

Step 3: saturate the memory. Suppose the action profile a^1 is repeated m consecutive periods, i.e., $a(t^2) = \dots = a(t^2 + m) = a^1$, and the state $x(k+1) \sim P(a^1, x(k))$ for all $k \in \{t^2, \dots, t^2 + m\}$. This

⁷ We choose δ in the above setting to satisfy the given condition for any action state pair $[a', x']$ and feasible state $x'' \in X(a'|x')$.

event happens with at least probability ϵ^{nm} because of the agents' inertia. If $[a^1, x(t^2)]$ is a recurrent state equilibrium then $a(t) = a^1$ for all times $t \geq t^2$ and we are done. Otherwise, we can repeat the same argument as before.

One can repeat this argument at most a finite number of times, because the potential function is increasing with each step and both the joint action set \mathcal{A} and state space X are finite. As a result, for any time $t_0 > 0$ and any action state pair $[a(t^0), x(t^0)] \in \mathcal{A} \times X$, if all agents adhere to any finite memory better reply process with inertia, then $[a(t^0 + T - m), x(t^0 + T - m)]$ will be a recurrent state equilibrium and $a(t^0 + T - m) = \dots = a(t^0 + T)$ with at least probability γ where

$$T := m + (|X| + m) \cdot |\mathcal{A} \times X|,$$

$$\gamma := \epsilon^{nm} (\delta \epsilon^{n(|X|+m)} \epsilon^{n-1} (1 - \epsilon) / |\mathcal{A}|)^{|\mathcal{A} \times X|}. \quad \square$$

5. Equilibrium selection in state based potential games

A second feature of potential games that is desirable in engineering systems is that of equilibrium selection. In particular, there are learning dynamics for potential games that guarantee convergence to the potential function maximizer as opposed to any equilibrium, e.g., log-linear learning (Blume, 1993, 1997; Marden & Shamma, 2012). This type of equilibrium selection is significant in many engineering systems as the interaction framework associated with a distributed multiagent system can frequently be represented as a potential game where the action profile that optimizes the system level objective is precisely the potential function maximizer (Arslan et al., 2007; Marden & Wierman, 2008). In this section we demonstrate that such algorithms also extend to the class of state based potential games under some restrictions on the state transition matrix P .

5.1. Background: binary log-linear learning for potential games

Consider an exact potential game G with potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$. The following learning algorithm for potential games is known as binary log-linear learning (Blume, 1993; Marden & Shamma, 2012). At each time $t > 0$, one agent $i \in N$ is randomly chosen and allowed to alter his current action. All other agents must repeat their current action at the ensuing time step, i.e. $a_{-i}(t) = a_{-i}(t-1)$. At time t , agent i selects one trial action a_i uniformly from the agent's action set $\mathcal{A}_i \setminus \{a_i(t-1)\}$. Then agent i employs the strategy $p_i(t) \in \Delta(\mathcal{A}_i)$ where

$$p_i^{a_i(t-1)}(t) = \frac{e^{\frac{1}{\tau} U_i(a(t-1))}}{e^{\frac{1}{\tau} U_i(a(t-1))} + e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1))}}$$

$$p_i^{a_i}(t) = \frac{e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1))}}{e^{\frac{1}{\tau} U_i(a(t-1))} + e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1))}}$$

$$p_i^{a'_i}(t) = 0, \quad \forall a'_i \in \mathcal{A}_i \setminus \{a_i, a_i(t-1)\}$$

for some temperature $\tau > 0$. The temperature τ determines how likely agent i is to select a suboptimal action. As $\tau \rightarrow \infty$, agent i will select either action a_i or $a_i(t-1)$ with equal probability. As $\tau \rightarrow 0$, agent i will select the action that yields the higher payoff, i.e., comparing $U_i(a_i, a_{-i}(t-1))$ and $U_i(a(t-1))$, with arbitrarily high probability. In the case of a non-unique comparison, i.e., $U_i(a_i, a_{-i}(t-1)) = U_i(a(t-1))$, agent i will select either a_i or $a_i(t-1)$ at random (uniformly).

In the repeated potential game in which all agents adhere to binary log-linear learning, the stationary distribution of the joint action profiles is $\mu \in \Delta(\mathcal{A})$ where (Blume, 1993)

$$\mu(a; \tau) = \frac{e^{-\frac{1}{\tau} \phi(a)}}{\sum_{\bar{a} \in \mathcal{A}} e^{-\frac{1}{\tau} \phi(\bar{a})}}. \quad (7)$$

One can interpret the stationary distribution μ as follows: for sufficiently large times $t > 0$, $\mu(a; \tau)$ equals the probability that $a(t) = a$ for a fixed temperature $\tau > 0$. As one decreases the temperature, $\tau \rightarrow 0$, all the weight of the stationary distribution μ is on the joint actions that maximize the potential function.

5.2. State based log-linear learning

Consider the following variant of log-linear learning for state based games. At each time t , at most one agent i is randomly selected according to a fixed probability distribution q where $q_i > 0$ denotes the probability that agent i is selected. Furthermore, we require that $\sum_{i \in N} q_i < 1$ meaning that there is a positive probability that no agent is selected, which we define as $q_0 = 1 - \sum_{i \in N} q_i > 0$. For simplicity, we assume that $q_i = 1/(n+1)$ for all $i \in N$. All non-selected agents must repeat their current action at the ensuing time step, i.e. $a_{-i}(t) = a_{-i}(t-1)$. At time t , agent i selects one trial action a_i uniformly from the agent's action set $\mathcal{A}_i \setminus \{a_i(t-1)\}$. Then agent i employs the strategy $p_i(t) \in \Delta(\mathcal{A}_i)$ where

$$p_i^{a_i(t-1)}(t) = \frac{e^{\frac{1}{\tau} U_i(a(t-1), x(t))}}{e^{\frac{1}{\tau} U_i(a(t-1), x(t))} + e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1), x(t))}}$$

$$p_i^{a_i}(t) = \frac{e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1), x(t))}}{e^{\frac{1}{\tau} U_i(a(t-1), x(t))} + e^{\frac{1}{\tau} U_i(a_i, a_{-i}(t-1), x(t))}}$$

$$p_i^{a'_i}(t) = 0, \quad \forall a'_i \in \mathcal{A}_i \setminus \{a_i, a_i(t-1)\}$$

for some temperature $\tau > 0$. Let $a_i \sim p_i(t)$ be the action selected at time t . The action profile at time t is $a(t) = (a_i, a_{-i}(t))$. The ensuing state $x(t+1)$ is then chosen randomly according to the state transitional probabilities $P(a(t), x(t))$. Here, it is important to note that the state transitional probabilities $P(\cdot)$ and agent selection probabilities q are fixed for all τ .

The analysis of log-linear learning for potential games characterizes the precise stationary distribution as a function of the temperature τ . This analysis relies on the fact that the underlying process is reversible which in general is not satisfied in state based games. However, the importance of the result is not solely the explicit form of the stationary distribution, but rather the fact that as $\tau \rightarrow 0$ the support of the limiting stationary distribution is precisely the set of action profiles that maximize the potential function. The support of the limiting distribution is referred to as the *stochastically stable states*. More precisely, an action profile $a \in \mathcal{A}$ is stochastically stable if and only if $\lim_{\tau \rightarrow 0^+} \mu(a; \tau) > 0$. We now demonstrate that the process described above guarantees that the stochastically stable states, which are now action state pairs $[a, x]$, are contained in the set of state based potential function maximizers.

Theorem 5.1. *Let $G = \{N, \{\mathcal{A}_i\}, \{U_i\}, X, P\}$ be an ordinal state based potential game with a state invariant potential function $\phi : \mathcal{A} \rightarrow \mathbb{R}$ that satisfies the following three conditions for any action profile $a \in \mathcal{A}$ and state $x \in X(a)$.⁸*

- (i) *The action invariant state transition process $P(a, \cdot)$ is aperiodic and irreducible over the states $X(a)$.*
- (ii) *For any agent $i \in N$ and action $a'_i \in \mathcal{A}_i$*

$$U_i(a'_i, a_{-i}, x) - U_i(a, x) \leq \phi(a'_i, a_{-i}) - \phi(a).$$

⁸ Recall that $V(a, x) \subseteq X$ is the support of the distribution $P(a, x)$ and $X(a) = \cup_{x \in X} V(a, x) \subseteq X$ represents the set of states that are reachable with the action profile a .

(iii) For any agent $i \in N$ and action $a'_i \in \mathcal{A}_i$ there exists a state $x' \in X(a)$ such that

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

For such state based potential games, the process log-linear learning guarantees that an action state pair $[a^*, x^*]$ is stochastically stable if and only if the action profile $a^* \in \arg \max_{a \in \mathcal{A}} \phi(a)$.

Note that Conditions (ii) and (iii) of [Theorem 5.1](#) provide a relaxation to the potential game structure by relaxing the equality constraint. We present the proof of [Theorem 5.1](#) in the [Appendix](#). See [Section 7.1](#) for an illustration of this algorithm.

6. An alternative formulation of state based games

A common assumption in game theory is that an agent can select any action in the agent's action set at any instance in time. In multiagent systems this assumption is not necessarily true. Rather, each agent has the ability to influence his action through different control strategies. Accordingly, we focus on the situation where each agent i has a set of control strategies Π_i that the agent can use to influence the agent's action choice. Let $\Pi := \prod_i \Pi_i$ be the set of joint control strategies. We represent the action transition function of agent i by a deterministic (or stochastic) transition function $g_i : \mathcal{A}_i \times \Pi \rightarrow \mathcal{A}_i$. In a repeated state based game, we adopt the convention that $a_i(t+1) = g_i(a_i(t), \pi(t))$ for any agent $i \in N$ and time $t \geq 1$ where $\pi(t) = (\pi_1(t), \dots, \pi_n(t))$ is the joint control decision at time t . This implies that an agent's ensuing action is potentially influenced by the control strategies of all agents. We assume throughout that each agent has a null control strategy $\pi_i^0 \in \Pi_i$ such that for any agent $i \in N$ and action $a_i \in \mathcal{A}_i$ we have $a_i = g_i(a_i, \pi^0)$ where $\pi^0 = (\pi_1^0, \dots, \pi_n^0)$.

We re-formulate the state based game described in [Section 3](#) for control based decisions as follows. First, we embed the action profiles \mathcal{A} and the original state space X into a new state space $Y = \mathcal{A} \times X$. Each agent $i \in N$ now has a state invariant control strategy set Π_i as described above and a new state dependent utility function of the form $U_i : \Pi \times Y \rightarrow \mathbb{R}$. Lastly, there is a Markovian state transition function $Q : \Pi \times Y \rightarrow \Delta(Y)$ which encompasses both the previous state transition function $P(\cdot)$ and the new action transition functions $\{g_i(\cdot)\}$. Repeated play of a state based game proceeds in the same fashion as before. All previous definitions and results directly carry over to this control based formulation by integrating the null action appropriately into the definitions and learning algorithms as action invariant trajectories are now encoded by the null control strategy π^0 . For example, [Condition \(ii\) of Definition 3.2](#) now takes the following form: for any policy state pair $[\pi, y]$ and any state y' in the support of $Q(\pi, y)$ the potential function satisfies $\phi(\pi^0, y') \geq \phi(\pi, y)$. We omit the remaining details for brevity.

7. Illustrations

In this section we illustrate the viability of the framework of state based potential games on two independent problems related to multiagent systems. In both settings we demonstrate that the framework of state based games provides the design freedom necessary to meet the given performance objectives.

7.1. Distributed resource allocation

To motivate the challenges inherent to distributed resource allocation we focus on the classical vehicle target assignment problem where there exists a finite set of targets (or resources),

denoted by \mathcal{R} , and each target $r \in \mathcal{R}$ has a relative worth $w_r \geq 0$ ([Murphey, 1999](#)). There are a finite number of vehicles denoted by N and the set of possible assignments for each vehicle $i \in N$ is $\mathcal{A}_i \subseteq 2^{\mathcal{R}}$. Lastly, each vehicle $i \in N$ is parameterized with an invariant success probability $p_i \in [0, 1]$. The goal of the vehicle target assignment problem is to establish a distributed control algorithm that converges to a joint assignment that maximizes the global objective

$$W(a) = \sum_{r \in \mathcal{R}: \{a\}_r \neq \emptyset} w_r \left(1 - \prod_{i \in \{a\}_r} [1 - p_i] \right) \quad (8)$$

where $\{a\}_r := \{i \in N : r \in a_i\}$ is the set of agents that selected target r and $\left(1 - \prod_{i \in \{a\}_r} [1 - p_i] \right)$ represents the joint probability of successfully eliminating target r .

Dealing with the informational limitations to the system designer represents one of the fundamental challenges associated with completing the above goal. That is, in general the system designer does not have information *a priori* pertaining to the number of targets \mathcal{R} , their relative worth $\{w_r\}_{r \in \mathcal{R}}$, or the layout of the mission space which gives rise to the possible assignments for each vehicle $\{\mathcal{A}_i\}_{i \in N}$. Our goal is to develop a general approach for the design of the vehicles' utility functions to ensure both the existence and efficiency of pure Nash equilibria irrespective of the mission space. To that end, we focus on the design of local utility functions of the form

$$U_i(a_i, a_{-i}) = \sum_{r \in a_i} f_r(i, \{a\}_r) \quad (9)$$

where $f_r(\cdot)$, which we refer to as the protocol at target r , designates how each vehicle evaluates the benefit of selecting a particular target given the assignment of the other vehicles. Note that defining a protocol $f_r(\cdot)$ for each target $r \in \mathcal{R}$ results in a well-defined game for any collection of action sets $\{\mathcal{A}_i\}_{i \in N}$.

Incremental marginal costs represent a promising approach to utility design, or protocol design, for distributed engineering systems ([Marden & Wierman, 2009](#)). A protocol defined using incremental marginal costs assigns payoff shares to the agents in accordance with their marginal contribution relative to a given ordering $x = \{x^i\}_{i \in N}$ where each agent $i \in N$ is assigned a unique order $x^i \in \{1, \dots, n\}$. For any ordering x , target $r \in \mathcal{R}$, and subset of agents $S \subseteq N$, the incremental marginal cost for agent $i \in S$ is $f_r(i, S; x)$ which is defined as

$$W_r(\{j \in S : x^j \leq x^i\}) - W_r(\{j \in S : x^j < x^i\})$$

where $W_r(S) = w_r \left(1 - \prod_{i \in S} [1 - p_i] \right)$. Irrespective of the state of the mission space, incremental marginal costs ensure (i) the interaction framework is a potential game and (ii) all pure Nash equilibria are guaranteed to be at least 50% efficient provided that the objective functions in [\(8\)](#) are submodular ([Marden & Wierman, 2009](#)).⁹ However, the optimal allocation is not necessarily a Nash equilibrium as the following example highlights.

Example 7.1. Consider a vehicle target assignment problem with the following specifications: (i) vehicles $N = \{v_1, v_2\}$, (ii) invariant detection probabilities $p_1 = 1$ and $p_2 = 0.5$, (iii) targets $\mathcal{R} = \{r_1, r_2, r_3\}$, (iv) values $w_{r_1} = 1$, $w_{r_2} = 2$, and $w_{r_3} = 2 - \epsilon$ where $\epsilon > 0$, and (v) action sets $\mathcal{A}_1 = \{r_1, r_2\}$ and $\mathcal{A}_2 = \{r_2, r_3\}$. Suppose utility functions for the vehicles were derived using incremental marginal costs. The payoff matrix associated with both orderings ($x^1 = 1$ and $x^2 = 2$) and ($x^1 = 2$ and $x^2 = 1$) are

⁹ A function $W_r : 2^N \rightarrow \mathbb{R}$ is submodular if for any agent sets $S \subseteq T \subseteq N$ and any agent $i \in N$ we have $W_r(S \cup \{i\}) - W_r(S) \geq W_r(T \cup \{i\}) - W_r(T)$.

		v_2	
		r_2	r_3
v_1	r_1	1, 1	1, $1 - \epsilon/2$
	r_2	2, 0	2, $1 - \epsilon/2$

Payoffs ($x^1 = 1, x^2 = 2$)

		v_2	
		r_2	r_3
v_1	r_1	1, 1	1, $1 - \epsilon/2$
	r_2	1, 1	2, $1 - \epsilon/2$

Payoffs ($x^1 = 2, x^2 = 1$)

		v_2	
		r_2	r_3
v_1	r_1	2	2 - $\epsilon/2$
	r_2	2	3 - $\epsilon/2$

Global Objective W

Note that the ordering only impacts the payoffs for the allocation (r_2, r_2) . For the ordering $x^1 = 1$ and $x^2 = 2$, which we will express as $1 \leftarrow 2$, the optimal allocation (r_2, r_3) is a pure Nash equilibrium. For the ordering $2 \leftarrow 1$, the optimal allocation is not a pure Nash equilibrium.¹⁰

We will now present an approach for dynamically adjusting the order of the vehicles to provide stronger efficiency guarantees. This approach is adapted from the approach presented in Marden and Wierman (2009) and we will explicitly highlight the differences below. We focus predominantly on the vehicle target assignment problem for explanation; however, the approach and forthcoming guarantees immediately extend to any resource allocation problems where the objective functions $W_r(\cdot)$ are submodular.

Agents. The vehicles $N = \{1, 2, \dots, n\}$.

Actions. The action set for each vehicles $i \in N$ is \mathcal{A}_i .

States. Define $\mathcal{X}(S)$ as the set of possible orderings for the agent set $S \subseteq N$. An ordering $x = \{x^i\}_{i \in S} \in \mathcal{X}(S)$ assigns a unique entry $x^i \in \{1, 2, \dots, |S|\}$ to each agent $i \in S$. The states of the system are $X = \prod_{r \in \mathcal{R}} X_r$ where $X_r = \cup_{S \subseteq N} \mathcal{X}(S)$ represent the set of states for each target r .

State dynamics. Let $[a(t), x(t)] \in \mathcal{A} \times X$ represent the allocation and state of the system at time t . The order for target r , denoted by $x_r(t)$, evolves according to a local state transition function of the form $x_r(t+1) \sim P_r(x_r(t), \{a(t)\}_r)$ that satisfies the following three rules:

– **Rule #1.** The order $x_r(t+1) \in \mathcal{X}(\{a(t)\}_r)$. We adopt the convention that $x_r^j(t+1) = \infty$ for each vehicle $j \notin \{a(t)\}_r$.

– **Rule #2.** The order $x_r(t+1)$ satisfies the following for all vehicles $i, j \in \{a(t)\}_r \cap \{i \in N : x_r^i(t) < \infty\}$

$$x_r^i(t) < x_r^j(t) \Rightarrow x_r^i(t+1) < x_r^j(t+1).$$

Note that this rule ensures a unique ordering over the vehicle set $\{a(t)\}_r \cap \{i \in N : x_r^i(t) < \infty\}$. The remaining components of this order are chosen according to any deterministic or stochastic rule.

¹⁰ Incremental marginal costs, also referred to as ordered protocols, ensures that the payoff shares are budget-balanced, i.e., for any target $r \in \mathcal{R}$, agent set $S \subseteq N$, and order x , we have $\sum_{i \in S} f_r(i, S; x) = W_r(S)$. In general, it is impossible to design budget-balanced utility functions which ensures that the optimal allocation is a pure Nash equilibrium irrespective of the mission space (Marden & Wierman, 2009).

– **Rule #3.** If $\{a(t)\}_r = \{i \in N : x_r^i(t) < \infty\}$ then $x_r(t+1)$ is chosen randomly according to a probability distribution with full support on the order set $\mathcal{X}(\{a(t)\}_r)$, i.e., orders are periodically reshuffled.¹¹

Agent utility functions. Recall that $x(t)$ is a valid ordering for the action profile $a(t-1)$, i.e., $\{a(t-1)\}_r = \{i \in N : x_r^i(t) \neq \infty\}$. First, consider any action state pair $[a(t), x(t)]$ where $x(t)$ is a valid ordering for the action profile $a(t)$. For this situation, the payoff for any agent $i \in N$ is

$$U_i(a(t), x(t)) = \sum_{r \in a_i(t)} f_r(i, \{a(t)\}_r; x_r(t)).$$

Now consider the situation where $x(t)$ is not a valid ordering for the action profile $a(t)$, i.e., $a(t) \neq a(t-1)$. For this situation, we begin by recovering the previous action profile $a(t-1)$ from the state $x(t)$ by evaluating $\{a(t-1)\}_r = \{i \in N : x_r^i(t) \neq \infty\}$. For this situation, the payoff for any agent $i \in N$ is defined as

$$\begin{aligned} U_i(a(t), x(t)) &= U_i(a_i(t), a_{-i}(t-1), x(t)) \\ &= \sum_{r \in a_i(t) \setminus a_i(t-1)} f_r(i, \{i\} \cup \{a(t-1)\}_r; \tilde{x}_r(t)) \\ &\quad + \sum_{r \in a_i(t) \cap a_i(t-1)} f_r(i, \{a(t-1)\}_r; x_r(t)) \end{aligned}$$

where $\tilde{x}_r(t) = P_r(x_r, \{i\} \cup \{a(t-1)\}_r)$ for any resource $r \in a_i(t) \setminus a_i(t-1)$, i.e., $\tilde{x}_r^j(t) = x_r^j(t-1)$ for any agent $j \in \{a(t-1)\}_r$ and $\tilde{x}_r^i(t) = |\{a(t-1)\}_r| + 1$. Hence, a vehicle evaluates his utility for alternative actions by considering the case for which he is the only deviating player.

Claim 1. The above game is an ordinal state based potential game with a state invariant potential function $\phi(a) = W(a)$.

Proof. Let $[a, x]$ represent any action state pair where $x = \{x_r\}_{r \in \mathcal{R}}$ represents a valid ordering for the allocation a , i.e., $x \in X(a)$. Because the welfare functions are submodular, we know that for any vehicle $i \in N$ and target $r \in a_i$

$$f_r(i, \{a\}_r; x_r) \geq W_r(\{a\}_r) - W_r(\{a\}_r \setminus \{i\}).$$

The utility of vehicle i for any allocation of the form $a' = (a'_i, a_{-i})$ is

$$\begin{aligned} U_i(a', x) &= \sum_{r \in a'_i \cap a_i} f_r(i, \{a\}_r; x_r) \\ &\quad + \sum_{r \in a'_i \setminus a_i} (W_r(\{a_{-i}\}_r \cup \{i\}) - W_r(\{a_{-i}\}_r)). \end{aligned}$$

We can express the utility difference $U_i(a', x) - U_i(a, x)$ as

$$\sum_{r \in a'_i \setminus a_i} (W_r(\{a'\}_r) - W_r(\{a\}_r)) - \sum_{r \in a_i \setminus a'_i} f_r(i, \{a\}_r; x_r).$$

By submodularity, we know that for any resource $r \in a_i \setminus a'_i$

$$f_r(i, \{a\}_r; x_r) \geq W_r(\{a\}_r) - W_r(\{a\}_r \setminus \{i\}).$$

Combining the above equations and simplifying gives us

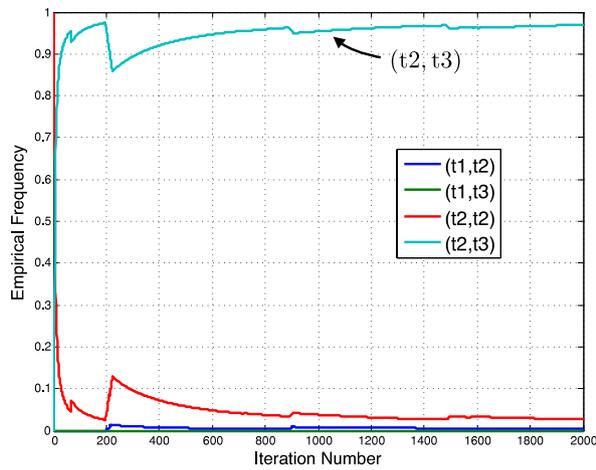
$$U_i(a', x) - U_i(a, x) \geq W(a') - W(a).$$

Hence, we verified Condition (i) of ordinal state based potential games which shows that

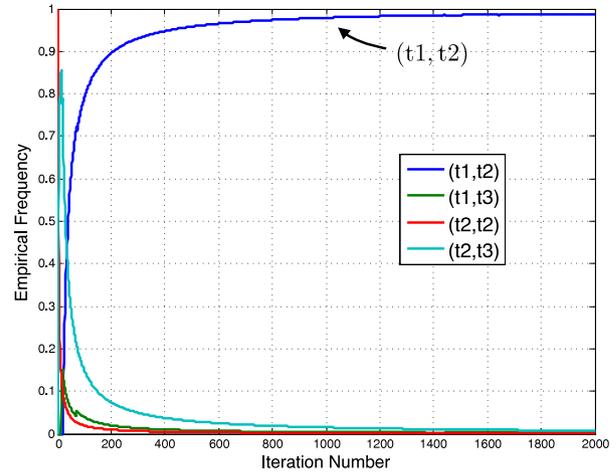
$$U_i(a'_i, a_{-i}, x) - U_i(a, x) > 0 \Rightarrow W(a'_i, a_{-i}) - W(a) > 0.$$

Condition (ii) is verified trivially because W is state invariant. \square

¹¹ The proposed state based game is similar to the approach in Marden and Wierman (2009). The difference stems from the addition of Rule #3 which is necessary for invoking log-linear learning to get the desired equilibrium selection.



(a) Case #1.



(b) Case #2.

Fig. 1. Case #1 focuses on a vehicle target assignment problem with the following specifications: (i) vehicles $N = \{1, 2\}$, (ii) detection probabilities $p_1 = 1$ and $p_2 = 0.5$, (iii) targets $\mathcal{R} = \{r_1, r_2, r_3\}$, (iv) values $w_{r_1} = 1$, $w_{r_2} = 2$, and $w_{r_3} = 0.9$, and (v) action sets $\mathcal{A}_1 = \{r_1, r_2\}$ and $\mathcal{A}_2 = \{r_2, r_3\}$. Case #2 focuses on a vehicle target assignment problem with the same specifications as Case #1 except for the values which are now $w_{r_1} = 1.8$, $w_{r_2} = 2$, and $w_{r_3} = 0.9$. Note that the distributed control algorithm resulting from the proposed state based game design and the learning algorithm log-linear learning guarantees convergence to the optimal allocation in both settings. In Case #1 the optimal allocation in (r_2, r_3) which yields a total welfare of $2 + 0.45 = 2.45$ while in Case #2 the optimal allocation in (r_1, r_2) which yields a total welfare of $1.8 + 1 = 2.8$.

This implies that a recurrent state equilibrium exists, any finite memory better response process converges almost surely to such an equilibrium, and that the optimal allocation is a recurrent state equilibrium. The following claim demonstrates that state based log-linear learning can be utilized to provide convergence to the potential function maximizer, i.e., the maximizer of W .

Claim 2. *The process log-linear learning guarantees that an action state pair $[a^*, x^*]$ is stochastically stable if and only if the action profile $a^* \in \arg \max_{a \in \mathcal{A}} W(a)$.*

Proof. To prove this we verify the three conditions of Theorem 5.1 for any action state pair $[a, x]$ where $x \in X(a)$. Condition (i) follows immediately from Rule #3 of the state dynamics. Condition (ii) follows from the proof of Claim 1. As for Condition (iii), consider any allocation $a \in \mathcal{A}$ and an order $\tilde{x} = \{\tilde{x}_r\}_{r \in \mathcal{R}} \in X(a)$ where for some agent $i \in N$, $\tilde{x}_r^i = |\{a\}_r|$ for all $r \in a_i$. Following the same arguments as in Claim 1, it is straightforward to show that for any action $a'_i \in \mathcal{A}_i$ we have

$$U_i(a'_i, a_{-i}, \tilde{x}) - U_i(a_i, a_{-i}, \tilde{x}) = W(a'_i, a_{-i}) - W(a). \quad \square$$

Fig. 1 presents two simulations of the learning algorithm log-linear learning on the vehicle target assignment problem highlighted in Example 7.1. Both simulations utilize the following parameters: $\tau = 0.1$, $q_0 = q_1 = q_2 = 1/3$, and Rule #3 selects an order according to a uniform distribution over the orders $1 \leftarrow 2$ and $2 \leftarrow 1$. In Fig. 1(a), we simulated the setting in Example 7.1 with the following specifications: (i) vehicles $N = \{1, 2\}$, (ii) detection probabilities $p_1 = 1$ and $p_2 = 0.5$, (iii) targets $\mathcal{R} = \{r_1, r_2, r_3\}$, (iv) values $w_{r_1} = 1$, $w_{r_2} = 2$, and $w_{r_3} = 0.9$, and (v) action sets $\mathcal{A}_1 = \{r_1, r_2\}$ and $\mathcal{A}_2 = \{r_2, r_3\}$. Here, the optimal allocation is (r_2, r_3) and the order which ensures that the optimal allocation is a Nash equilibrium is $1 \leftarrow 2$. This simulation shows the empirical frequency of the allocation (r_2, r_3) converges to 1 relatively quickly. In Fig. 1(b), we simulated a variation of the setting in Example 7.1 with the following specifications: (i) vehicles $N = \{1, 2\}$, (ii) detection probabilities $p_1 = 1$ and $p_2 = 0.5$, (iii) targets $\mathcal{R} = \{r_1, r_2, r_3\}$, (iv) values $w_{r_1} = 1.8$, $w_{r_2} = 2$, and $w_{r_3} = 0.9$, and (v) action sets $\mathcal{A}_1 = \{r_1, r_2\}$ and $\mathcal{A}_2 = \{r_2, r_3\}$. Here, the optimal allocation is (r_1, r_2) and the order which ensures that the optimal allocation is a Nash equilibrium is $2 \leftarrow 1$. This simulation

shows the empirical frequency of the allocation (r_1, r_2) converges to 1 relatively quickly. Note that both settings utilized the same distributed control algorithm resulting from the proposed state based game design and the learning algorithm log-linear learning. Furthermore, note that vehicle 1 never had access to w_{r_3} while vehicle 2 never had access to w_{r_1} . Nonetheless, this approach yields the optimal behavior irrespective of the problem setting.

7.2. Local control design

Consider a multiagent system consisting of a set of agents N , an action set $\mathcal{A}_i = \mathbb{R}$ for each agent $i \in N$, and a system level objective of the form $W : \mathcal{A} \rightarrow \mathbb{R}$ that a system designer seeks to maximize. We assume throughout that W is concave, continuously differentiable, and that a solution is guaranteed to exist. Here, we focus on the design of local control laws where the information available to each agent is represented by an undirected and connected graph $\mathcal{G} = \{N, \mathcal{E}\}$ with nodes N and edges \mathcal{E} . These local control policies produce a sequence of action profiles $a(0), a(1), a(2), \dots$ where at each iteration $t \in \{0, 1, \dots\}$ each agent i makes a decision independently according to a local control law of the form:

$$a_i(t) = F_i(\{\text{Information about agent } j \text{ at time } t\}_{j \in N_i}) \quad (10)$$

where $N_i := \{j \in N : (i, j) \in \mathcal{E}\}$ denotes the neighbors of agent i . The goal in this setting is to design the local controllers $\{F_i(\cdot)\}_{i \in N}$ within the given informational constraints such that the collective behavior converges to a joint decision $a^* \in \arg \max_{a \in \mathcal{A}} W(a)$.

Consider a distributed gradient ascent algorithm where each agent's control policy represents a gradient ascent process with respect to the system level objective W . This approach ensures that the action profile will converge to the optimizer of the system level objective W . However, it is important to highlight that the structure of the agents' control policies may not satisfy our desired local constraints in (10). An alternative approach is to have each agent's control policy represent a gradient ascent process with respect to a given local utility function U_i as opposed to the system level objective W . Is it possible to design local utility functions for the individual agents such that (i) all resulting equilibria optimize the system level objective and (ii) the game possesses a structure that can be exploited in distributed learning?

We now review the methodology developed in Li and Marden (2011) which accomplishes this task. The underlying design fits into the formulation of state based games depicted in Section 6. Accordingly, the structure of the agents' local utility functions is of the form

$$U_i : \prod_{j \in N_i} (\Pi_j \times Y_j) \rightarrow R \quad (11)$$

where Π_j is the set of control policies for agent j and Y_j is the set of local state variables for agent j . The forthcoming design embodies continuous action sets (or control policies) as opposed to finite action sets as considered in the rest of this paper. While the learning results contained in this paper do not readily apply to this setting, the designed game is a state based potential game. Accordingly, results in Li and Marden (2011) prove that gradient ascent dynamics guarantee convergence to recurrent state equilibria for such continuous state based potential games. The details of the design are as follows:

Agents. The agent set is $N = \{1, 2, \dots, n\}$.

States. The starting point of the design is an underlying state space Y where each state $y \in Y$ is defined as a tuple $y = (a, e)$, where $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ is the action profile and $e = (e_1, \dots, e_n)$ is a profile of agent based estimation terms where $e_i = (e_i^1, \dots, e_i^n) \in \mathbb{R}^n$ is agent i 's estimation for the joint action profile a . The term e_i^k captures agent i 's estimate of agent k 's action a_k .

Action sets. Each agent i is assigned a set of control policies Π_i that permits the agents to change their value and change their estimation through communication with neighboring agents. Specifically, a control for agent i is defined as a tuple $\pi_i = (\hat{a}_i, \hat{e}_i)$ where $\hat{a}_i \in \mathbb{R}$ indicates a change in the agent's action and $\hat{e}_i := (\hat{e}_i^1, \dots, \hat{e}_i^n)$ indicates a change in the agent's estimation terms e_i . We represent the changes in estimation terms for an agent k by the tuple $\hat{e}_i^k := \{\hat{e}_{i \rightarrow j}^k\}_{j \in N_i}$ where $\hat{e}_{i \rightarrow j}^k \in \mathbb{R}$ represents the estimation value that agent i "passes" to agent j regarding the action of agent k .

State dynamics. We now describe how the state evolves as a function of the control choices $\pi(0), \pi(1), \dots$, where $\pi(k)$ is the control profile at stage k . Define the initial state as $y(0) = [a(0), e(0)]$ where $a(0) = (a_1(0), \dots, a_n(0))$ is the initial action profile and $e(0)$ is an initial estimation profile that satisfies

$$\sum_{i \in N} e_i^k(0) = n \cdot a_k(0) \quad (12)$$

for each agent $k \in N$; hence, the initial estimation values are contingent on the initial action profile.¹² We represent the state transition function $Q(\pi, y)$ by a set of local state transition functions $\{Q_i^a(\pi, y)\}_{i \in N}$ and $\{Q_{i,k}^e(\pi, y)\}_{i,k \in N}$. For any distinct agents $i, k \in N$, state $y = (a, e)$, and control choice $\pi = (\hat{a}, \hat{e})$, the local state transition functions take on the form

$$\begin{aligned} Q_i^a(\pi, y) &= a_i + \hat{a}_i \\ Q_{i,i}^e(\pi, y) &= e_i^i + n \cdot \hat{a}_i + \sum_{j \in N: i \in N_j} \hat{e}_{j \rightarrow i}^i - \sum_{j \in N_i} \hat{e}_{i \rightarrow j}^i \\ Q_{i,k \neq i}^e(\pi, y) &= e_i^k + \sum_{j \in N: i \in N_j} \hat{e}_{j \rightarrow i}^k - \sum_{j \in N_i} \hat{e}_{i \rightarrow j}^k. \end{aligned} \quad (13)$$

It is straightforward to show that for any sequence of control choices $\pi(0), \pi(1), \dots$, the resulting state trajectory $y(t) = (a(t), e(t)) = Q(\pi(t-1), y(t-1))$ satisfies for all times $t \geq 1$ and

agents $k \in N$

$$\sum_{i=1}^n e_i^k(t) = n \cdot a_k(t). \quad (14)$$

Agent utility functions. The last part of our design is the agents' utility functions. For any state $y \in Y$ and admissible control $\pi \in \Pi$ the utility function of agent i is defined as

$$U_i(\pi, y) = \sum_{j \in N_i} W(\tilde{e}_j^1, \tilde{e}_j^2, \dots, \tilde{e}_j^n) + \sum_{j \in N_i} \sum_{k \in N} [\tilde{e}_i^k - \tilde{e}_j^k]^2$$

where $(\tilde{v}, \tilde{e}) = Q(\pi, y)$ represents the ensuing state. Note that the agents' utility functions are local and of the form (11).

In Li and Marden (2011) the authors prove two properties regarding the designed state based game. First, the designed game is a state based potential game with potential function $\phi(\pi, y)$ of the form

$$\sum_{i \in N} W(\tilde{e}_i^1, \tilde{e}_i^2, \dots, \tilde{e}_i^n) + \frac{1}{2} \sum_{i \in N} \sum_{j \in N_i} \sum_{k \in N} [\tilde{e}_i^k - \tilde{e}_j^k]^2$$

where $(\tilde{a}, \tilde{e}) = Q(\pi, y)$ represents the ensuing state. This ensures (i) the existence of a recurrent state equilibrium and (ii) that the game possesses an underlying structure that can be exploited by distributed learning algorithms, e.g., gradient play (Li & Marden, 2011). The second property is that all equilibria of the designed game are optimal with respect to W provided that the designed interaction graph satisfies some relatively weak connectivity conditions.¹³ More specifically, a control state pair $[\pi, y] = [(\hat{a}, \hat{e}), (a, e)]$ is a recurrent state equilibrium if and only if the following conditions are satisfied:

- (i) The action profile a is optimal, i.e., $a \in \arg \max_{a' \in \mathcal{A}} W(a')$.
- (ii) The estimation profile e is consistent with a , i.e., for any $\forall i, k \in N$ we have $e_i^k = a_k$.
- (iii) The change in action profile satisfies $\hat{a} = \mathbf{0}$.
- (vi) The change in estimation profile satisfies the following for all agents $i, k \in N$,

$$\sum_{j \in N_i} \hat{e}_{i \rightarrow j}^k = \sum_{j \in N: i \in N_j} \hat{e}_{j \rightarrow i}^k.$$

Hence, this design provides a systematic methodology for distributing an optimization problem under virtually any desired degree of locality in agent objective functions.

8. Conclusions

The beauty of game theory centers on its generality which makes its tools and structure accessible to a broad class of distributed systems encompassing both the social and engineering sciences. Unfortunately, as game theory evolves from a descriptive tool for social systems to a prescriptive tool for engineering systems, the existing framework is not suitable for meeting this new set of design challenges. The primary goal of this paper is to provide a new game theoretic framework that is better suited for handling this prescriptive agenda—state based potential games. Within the context of state based potential games, we introduced new notions of equilibria in addition to distributed learning algorithms for finding such equilibria. Lastly, we demonstrated the applicability of state based potential games on two cooperative control problems pertaining to distributed resource allocation and the design of local and distributed control laws.

¹² Note that satisfying condition (12) is trivial as we can set $e_i^j(0) = n \cdot a_j(0)$ and $e_i^i(0) = 0$ for all agents $i, j \in N$ where $i \neq j$.

¹³ One such condition is that the interaction graph is connected, undirected, and contains at least one odd-length cycle (non-bipartite). We direct the reader to (Li & Marden, 2011) for alternative sufficient conditions on the interaction graph.

Appendix

We start with an outline of the proof of [Theorem 5.1](#). The learning algorithm state based log-linear learning induces a finite Markov process over the state space $\mathcal{A} \times X$. The analysis of this Markov process is challenging stemming from the fact that the process is not reversible. However, in [Lemma A.1](#) we prove that this Markov process fits into the class of *regular perturbed processes* as introduced in [Young \(1993\)](#). This realization allows for this process to be analyzed in a tractable fashion using graph theoretic techniques termed resistance trees ([Young, 1993](#)). [Lemmas A.2](#) and [A.3](#) establish two key analytical properties of resistance trees for state based log-linear learning which simplify the analysis of the stationary distribution for the given process. Lastly, [Lemma A.4](#) complete the proof by exploiting these properties to provide the desired characterization.

Before proceeding with the proof, we first introduce a brief background on the analysis of regular perturbed processes using the theory of resistance trees ([Young, 1993](#)).

A.1. Background on regular perturbed processes

For a detailed review of the theory of resistance trees, please see [Young \(1993\)](#). Let P^0 denote the probability transition matrix for a finite state Markov chain over the state space Z . Consider a “perturbed” process such that the size of the perturbation can be parameterized by a scalar $\epsilon > 0$, and let P^ϵ be the associated transition probability matrix. The process P^ϵ is called a *regular perturbed Markov process* if P^ϵ is ergodic for all sufficiently small $\epsilon > 0$ and P^ϵ approaches P^0 at an exponentially smooth rate ([Young, 1993](#)). Specifically, the latter condition means that $\forall z, z' \in Z$,

$$\lim_{\epsilon \rightarrow 0^+} P_{zz'}^\epsilon = P_{zz'}^0,$$

and

$$P_{zz'}^\epsilon > 0 \text{ for some } \epsilon > 0 \Rightarrow 0 < \lim_{\epsilon \rightarrow 0^+} \frac{P_{zz'}^\epsilon}{\epsilon^{R(z \rightarrow z')}} < \infty,$$

for some nonnegative real number $R(z \rightarrow z')$, which is called the *resistance* of the transition $z \rightarrow z'$.¹⁴(Note in particular that if $P_{zz'}^0 > 0$ then $R(z \rightarrow z') = 0$.) We will adopt the convention that if $P_{zz'}^\epsilon = 0$ for all sufficiently small ϵ , then the resistance $R(z \rightarrow z') = \infty$.

Now construct a complete directed graph with $|Z|$ vertices, one for each state. The weight on the directed edge $z_i \rightarrow z_j$ is $R(z_i \rightarrow z_j)$. A tree, T , rooted at vertex z_j , or z_j -tree, is a set of $|Z| - 1$ directed edges such that, from every vertex different from z_j , there is a unique directed path in the tree to z_j . The resistance of a rooted tree, T , is the sum of the resistances $R(z_i \rightarrow z_k)$ on the $|Z| - 1$ edges that compose it. The *stochastic potential*, $\gamma(z_j)$, of state z_j is defined to be the minimum resistance over all trees rooted at z_j . The following theorem gives a simple criterion for determining the stochastically stable states ([Lemma 1, \(Young, 1993\)](#)).

Theorem A.1. *Let P^ϵ be a regular perturbed Markov process, and for each $\epsilon > 0$ let μ^ϵ be the unique stationary distribution of P^ϵ . Then $\lim_{\epsilon \rightarrow 0} \mu^\epsilon$ exists and the limiting distribution μ^0 is a stationary distribution of P^0 . The stochastically stable states (i.e., the support of μ^0) are precisely those states with minimum stochastic potential.*

¹⁴ We adopt the convention that $P_{zz'}$ denotes the probability of transitioning from state z to state z' .

A.2. Proof of [Theorem 5.1](#)

Log-linear learning induces a finite and aperiodic process over the state space $\mathcal{A} \times X$. We will now show that the trees rooted at recurrent state equilibria that optimize the potential $\phi(\cdot)$ will have minimum stochastic potential over all possible states. Therefore, according to [Theorem A.1](#), the probability that the state will be at a recurrent state equilibrium that optimizes the potential $\phi(\cdot)$ can be made arbitrarily close to 1.

For convention, we define a state of our Markov process by the tuple $[a(t-1), x(t)] \in \mathcal{A} \times X$. We denote the probability that the state x is selected according to the transition probabilities $P(a(t), x(t))$ as $P(x|a(t), x(t))$. There are two types of feasible transitions for this process. The first type of transition is when the action profile changes by a unilateral deviation, i.e., $a \rightarrow a' = (a'_i, a_{-i})$ where $a'_i \neq a_i$ and is of the form

$$[a, x] \rightarrow [a', x']$$

where $P(x'|a', x) > 0$. The second type of transition is when the action profile stays constant and is of the form

$$[a, x] \rightarrow [a, x']$$

where $P(x'|a, x) > 0$. We will now prove [Theorem 5.1](#) using the following sequence of lemmas. The first lemma states that the considered process is in fact a regular perturbed process.

Lemma A.1. *Log-linear learning for state based potential games induces a regular perturbed Markov process. Let $z = [a, x]$ be any action state pair. The resistance of a feasible transition of the form $z \rightarrow z' = [(a'_i, a_{-i}), x']$ is*

$$R(z \rightarrow z') = \max_{a_i^* \in \{a_i, a'_i\}} U_i(a_i^*, a_{-i}, x) - U_i(a'_i, a_{-i}, x). \quad (15)$$

The resistance of a feasible transition of the form $z \rightarrow z'' = [a, x']$ is of the form

$$R(z \rightarrow z'') = 0. \quad (16)$$

Proof. We analyze this process with respect to $\epsilon := e^{-\frac{1}{\tau}}$ rather than the temperature τ . Let P^ϵ denote the associated transition probability matrix. The probability of transitioning from $[a^0, x^0]$ to $[a^1, x^1]$ where $a^1 := (a_i^1, a_{-i}^0)$ for some agent i and action $a_i^1 \neq a_i^0$ is

$$P_{[a^0, x^0] \rightarrow [a^1, x^1]}^\epsilon = c_0^1 \frac{\epsilon^{-U_i(a_i^1, a_{-i}^0, x^0)}}{\sum_{a_i \in \{a_i^0, a_i^1\}} \epsilon^{-U_i(a_i, a_{-i}^0, x^0)}} \quad (17)$$

where

$$c_0^1 = \frac{q_i}{|\mathcal{A}_i| - 1} \cdot P(x^1|x^0, a^1).$$

The probability is divided into three components: (i) the probability of selecting agent i and trial action a_i^1 , (ii) the probability of agent i selecting action a_i^1 , and (iii) the probability of the state traversing to x^1 . Let $a_i^* \in \arg \max_{a_i \in \{a_i^0, a_i^1\}} U_i(a_i, a_{-i}^0, x^0)$. Multiplying the numerator and denominator of (17) by $\epsilon^{U_i(a_i^*, a_{-i}^0, x^0)}$, we obtain

$$c_0^1 \frac{\epsilon^{U_i(a_i^*, a_{-i}^0, x^0) - U_i(a_i^1, a_{-i}^0, x^0)}}{\sum_{a_i \in \{a_i^0, a_i^1\}} \epsilon^{U_i(a_i^*, a_{-i}^0, x^0) - U_i(a_i, a_{-i}^0, x^0)}}.$$

Accordingly,

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \frac{P_{[a^0, x^0] \rightarrow [a^1, x^1]}^\epsilon}{\epsilon^{U_i(a_i^*, a_{-i}^0, x^0) - U_i(a_i^1, a_{-i}^0, x^0)}} \\ = \frac{q_i}{|A_i| - 1} \cdot \frac{1}{|Z_i|} \cdot P(x^1 | x^0, a^1) > 0 \end{aligned}$$

where $z_i = \{a_i \in \{a_i^0, a_i^1\} : U_i(a_i, a_{-i}^0) = \max_{a_i \in \{a_i^0, a_i^1\}} U_i(a_i, a_{-i}^0, x^0)\}$ denotes agent i 's best response set. This implies that the transitional probability decays at an exponentially smooth rate and the resistance of the transition is

$$\begin{aligned} R([a^0, x^0] \rightarrow [a^1, x^1]) &= U_i(a_i^*, a_{-i}^0, x^0) - U_i(a_i^1, a_{-i}^0, x^0) \\ &\geq 0. \end{aligned}$$

Similarly, the probability of transitioning from $[a^0, x^0]$ to $[a^0, x^1]$ is

$$P_{[a^0, x^0] \rightarrow [a^0, x^1]}^\epsilon \geq q_0 \cdot P(x^1 | x^0, a^0) > 0.$$

The resistance of this transition is

$$R([a^0, x^0] \rightarrow [a^0, x^1]) = 0. \quad \square$$

The second lemma proves a property about resistance trees with minimum stochastic potential. Let $[a^0, x^0] \rightarrow [a^1, x^1]$ represent a feasible transition where $a^1 := (a_i^1, a_{-i}^0)$ for some agent i and action $a_i^1 \neq a_i^0$. Suppose $U_i(a_i^0, a_{-i}^0, x^0) > U_i(a_i^1, a_{-i}^0, x^0)$. By Lemma A.1 and Condition (ii) of Theorem 5.1, we know that

$$\begin{aligned} R([a^0, x^0] \rightarrow [a^1, x^1]) &= U_i(a_i^0, a_{-i}^0, x^0) - U_i(a_i^1, a_{-i}^0, x^0) \\ &\geq \phi(a^0) - \phi(a_i^1, a_{-i}^0). \end{aligned}$$

Note that if $U_i(a_i^0, a_{-i}^0, x^0) \leq U_i(a_i^1, a_{-i}^0, x^0)$ the associated resistance would be 0. We call such an edge “easy” if

$$R([a^0, x^0] \rightarrow [a^1, x^1]) = \max\{\phi(a^0) - \phi(a^1), 0\}.$$

Furthermore, we also say that all edges of the form $[a^0, x^0] \rightarrow [a^0, x^1]$ are easy.

Lemma A.2. *All edges of a tree with minimum stochastic potential must be easy.*

Proof. We prove this lemma by contradiction. Suppose there exists a tree T rooted at a state $[a, x]$ of minimum stochastic potential which does not consist of only easy edges. Then there exist an edge in the tree T of the form $[a^0, x^0] \rightarrow [a^1, x^1]$, where $a^1 = (a_i^1, a_{-i}^0)$ for some agent $i \in N$ and action $a_i^1 \neq a_i^0$, such that

$$R([a^0, x^0] \rightarrow [a^1, x^1]) > \max\{\phi(a^0) - \phi(a^1), 0\}.$$

By Condition (iii) we know that there exists a state $x' \in X(a^0)$ such that

$$U_i(a_i^1, a_{-i}^0, x') - U_i(a_i^0, a_{-i}^0, x') = \max\{\phi(a^0) - \phi(a^1), 0\}.$$

Consider a sequence of feasible transitions, or path which we denote by \mathcal{P} , of the form

$$\left\{ [a^0, x^0] \rightarrow \dots \rightarrow [a^0, x'] \rightarrow [a^1, x'] \rightarrow \dots \rightarrow [a^1, x^1] \right\}.$$

Note that both sequences of action invariant transitions, represented by “ $\rightarrow \cdot^{a^0} \rightarrow$ ” and “ $\rightarrow \cdot^{a^1} \rightarrow$ ”, are feasible because of Condition (i) and have a total resistance of 0. The intermediate transition $[a^0, x'] \rightarrow [a^1, x']$ has a resistance = $\max\{\phi(a^0) - \phi(a^1), 0\}$.

Let $\mathcal{P}' = \{[a^1, x^1] \rightarrow \dots \rightarrow [a, x]\}$, be the unique sequence of transitions from $[a^1, x^1]$ to $[a, x]$ in the original tree T . We now demonstrate two different approaches for augmenting the graph structure depending on whether there are any common edges between \mathcal{P} and \mathcal{P}' .

Case #1. We first start with the easy case where there are no commonalities between the intermediate states in \mathcal{P} and \mathcal{P}' ; this is, if an edge $[\tilde{a}, \tilde{x}] \rightarrow [\cdot \cdot \cdot]$ is in the path \mathcal{P} then there is no edge leaving the state $[\tilde{a}, \tilde{x}]$, i.e., an edge of the form $[\tilde{a}, \tilde{x}] \rightarrow [\cdot \cdot \cdot]$, in the path \mathcal{P}' . Consider a new tree T' rooted still at $[a, x]$ constructed from the original tree T as follows:

- (i) Add the edges \mathcal{P} to the tree T . The new edges have a combined resistance = $\max\{\phi(a^0) - \phi(a^1), 0\}$.
- (ii) Remove the edge $[a^0, x^0] \rightarrow [a^1, x^1]$ in the original tree T . This edge has a resistance $> \max\{\phi(a^0) - \phi(a^1), 0\}$.
- (iii) Remove the redundant edges from tree T . That is, if we added the edge $[\tilde{a}, \tilde{x}] \rightarrow [\cdot \cdot \cdot]$ in step (i) then remove the edge exiting the state $[\tilde{a}, \tilde{x}]$ in the original tree T . The total resistance of these edges has a resistance ≥ 0 .

Note that T' is also a valid tree rooted at $[a, x]$. For example, any state $[\tilde{a}, \tilde{x}] \in \mathcal{P}$ has a unique path to $[a^1, x^1]$ along \mathcal{P} and then a unique path from $[a^1, x^1]$ to $[a, x]$ along \mathcal{P}' . This is true because there are no commonalities between \mathcal{P} and \mathcal{P}' ; hence, \mathcal{P}' still exists in the new tree T' . Since all edges leaving states not in \mathcal{P} are unchanged given the graph augmentation above, each state still possesses a unique path either to \mathcal{P} or directly to $[a, x]$. Hence, T' is a valid tree rooted at $[a, x]$. The resistance of the tree T' , denoted by $R(T')$, is

$$\begin{aligned} R(T') &\leq R(T) + R([a^0, x^0] \rightarrow [a^1, x^1]) - R([a^0, x^0] \rightarrow [a^1, x^1]) \\ &< R(T). \end{aligned}$$

Hence, the original tree T does not have a minimum stochastic potential.

Case #2. Now, consider the case where there are commonalities between the intermediate states in \mathcal{P} and \mathcal{P}' , i.e., there exists a state $[\tilde{a}, \tilde{x}]$ such that $[\tilde{a}, \tilde{x}] \rightarrow [\tilde{a}', \tilde{x}']$ is an edge in \mathcal{P} and $[\tilde{a}, \tilde{x}] \rightarrow [\tilde{a}'', \tilde{x}'']$ is an edge in \mathcal{P}' . In this case, the tree construction depicted above need not lead to a valid tree rooted at $[a, x]$ because of the removal of these “common” edges. To rectify this situation, let $\tilde{\mathcal{P}}$ be the complete set of transitions in \mathcal{P} such that there are no conflicts of this sort. Construct the new tree T'' in the same fashion depicted above with the sole difference being in step (i) where we now add the edges $\tilde{\mathcal{P}}$, as opposed to \mathcal{P} , to the tree T . Note that T'' is also a valid tree rooted at $[a, x]$. For example, any state $[\tilde{a}, \tilde{x}] \in \mathcal{P}$ has a unique path to the root $[a, x]$ either through $[a^1, x^1]$ and the path \mathcal{P}' or a truncated version of the path \mathcal{P} and the original tree T . Since all edges leaving states not in \mathcal{P} are unchanged given the graph augmentation above, each state still possesses a unique path either to \mathcal{P} or directly to $[a, x]$. Hence, T'' is a valid tree rooted at $[a, x]$. The resistance of the tree T'' also satisfies

$$R(T'') < R(T).$$

Hence, the original tree T does not have minimum stochastic potential proving the contradiction. It is straightforward to repeat this argument to show that the edges of the form $[a^0, x^0] \rightarrow [a^0, x^1]$ must also be easy, i.e., have resistance 0. \square

We next utilize the fact that all edges in a tree with minimum stochastic potential must be easy to derive a relationship between the stochastic potential of different states.

Lemma A.3. *Let T_0 be a tree of minimum stochastic potential rooted at the state $[a^0, x^0]$ and T_1 be a tree of minimum stochastic potential rooted at the state $[a^1, x^1]$ where $a^1 = (a_i^1, a_{-i}^0)$ for some agent $i \in N$ and action $a_i^1 \neq a_i^0$. Suppose there exists a set of edges in tree T_0 of the form*

$$[a^1, x^1] \rightarrow \dots \rightarrow [a^1, x'] \rightarrow [a^0, x'] \rightarrow \dots \rightarrow [a^0, x^0] \quad (18)$$

and a set of edges in tree T_1 of the form

$$[a^0, x^0] \rightarrow \overset{a^0}{\dots} \rightarrow [a^0, x'''] \rightarrow [a^1, x'''''] \rightarrow \overset{a^1}{\dots} \rightarrow [a^1, x^1] \quad (19)$$

for some states $x', x'', x''', x'''' \in X$. Then, the stochastic potential of the two states is related as follows

$$\gamma([a^0, x^0]) - \gamma([a^1, x^1]) = \phi(a^1) - \phi(a^0).$$

Proof. Consider a new tree $T_{0 \rightarrow 1}$ rooted at $[a^1, x^1]$ constructed from the original tree T_0 as follows:

(i) Consider an easy path \mathcal{P} of the form

$$[a^0, x^0] \rightarrow \overset{a^0}{\dots} \rightarrow [a^0, \tilde{x}] \rightarrow [a^1, \tilde{x}'] \rightarrow \overset{a^1}{\dots} \rightarrow [a^1, x^1].$$

Add the edges of \mathcal{P} to the tree T_0 . These new edges have a combined resistance equal to $\max\{\phi(a^0) - \phi(a^1), 0\}$.

(ii) Remove the edge $[a^1, x'] \rightarrow [a^0, x'']$ in the original tree T_0 . This edge has a resistance $\max\{\phi(a^1) - \phi(a^0), 0\}$.

(iii) Consider an easy path \mathcal{P}' of the form

$$[a^1, x'] \rightarrow \overset{a^1}{\dots} \rightarrow [a^1, x^1].$$

Add the edges of \mathcal{P}' to the tree T_0 . These new edges have a combined resistance equal to 0.

(iv) Remove the edge leaving $[a^1, x^1]$ in the original tree T_0 . This edge has a resistance ≥ 0 .

(v) Remove the other redundant edges in the original tree T_0 . These edges have a total resistance ≥ 0 .

Note that $T_{0 \rightarrow 1}$ is a tree rooted at $[a^1, x^1]$ since there is a unique path from any state to this new root. To see this consider any state $[\tilde{a}, \tilde{x}]$ in the first set of action invariant transitions “ $\rightarrow \overset{a^1}{\dots} \rightarrow$ ” in the path (18). Then, there is a unique path to the root $[a^1, x^1]$ through the action state pair $[a^1, x']$ and the path \mathcal{P}' . Consider any state $[\tilde{a}, \tilde{x}]$ in the second set of action invariant transitions “ $\rightarrow \overset{a^0}{\dots} \rightarrow$ ” in the path (18). Then, there is a unique path to the root $[a^1, x^1]$ through the action state pair $[a^0, x^0]$ and the path \mathcal{P} . This property trivially holds for all alternative states since all edges leaving states not in \mathcal{P} or \mathcal{P}' are unchanged given the graph augmentation above.

The stochastic potential of $[a^1, x^1]$ is bounded above by

$$\begin{aligned} \gamma([a^1, x^1]) &\leq R(T_{0 \rightarrow 1}) \\ &= \gamma([a^0, x^0]) + \phi(a^0) - \phi(a^1). \end{aligned}$$

Repeating the above analysis starting from the tree T_1 yields

$$\gamma([a^0, x^0]) \leq \gamma([a^1, x^1]) + \phi(a^1) - \phi(a^0)$$

which gives us

$$\gamma([a^0, x^0]) - \gamma([a^1, x^1]) = \phi(a^1) - \phi(a^0). \quad \square$$

Before completing the proof of [Theorem 5.1](#), we introduce the following notation. We denote a path of easy transition as highlighted in [Lemma A.3](#) compactly as

$$[a^0, x^0] \xrightarrow{\text{easy}} [a^1, x^1].$$

We will now use the previous lemma to finish the proof.

Lemma A.4. *An action state pair $[a^*, x^*]$ has minimum stochastic potential over all states if and only if the action profile a^* maximizes the potential function ϕ .*

Proof. Suppose that an action state pair $[a^0, x^0]$ has minimum stochastic potential over all states and $\phi(a^0) < \phi(a^*)$ for some

action profile $a^* \in \mathcal{A}$. Let T_0 be a minimum resistance tree rooted at $[a^0, x^0]$. In the tree T_0 , there exists a series of easy transition paths of the form

$$[a^*, x^*] = [a^m, x^m] \xrightarrow{\text{easy}} [a^{m-1}, x^{m-1}] \xrightarrow{\text{easy}} \dots \xrightarrow{\text{easy}} [a^0, x^0]$$

for some states $x^1, x^2, \dots, x^m \in X$ where each transition is highlighted by a unilateral deviation, i.e., for each $k \in \{1, 2, \dots, m\}$ the action profile $a^k = (a_i^k, a_{-i}^{k-1})$ for some agent $i \in N$ with action $a_i^k \neq a_i^{k-1}$. Construct a new tree T_1 , from the original tree T_0 , rooted at $[a^1, x^1]$ according to the procedure highlighted in [Lemma A.3](#). The new tree has resistance

$$R(T_1) = R(T_0) + \phi(a^0) - \phi(a^1).$$

Construct a new tree T_2 , from the tree T_1 , rooted at $[a^2, x^2]$ according to the procedure highlighted in [Lemma A.3](#). The new tree has resistance

$$\begin{aligned} R(T_2) &= R(T_1) + \phi(a^1) - \phi(a^2) \\ &= R(T_0) + \phi(a^0) - \phi(a^2). \end{aligned}$$

Repeat this process m times to construct a tree T_m rooted at $[a^m, x^m]$. The new tree has resistance

$$\begin{aligned} R(T_m) &= R(T_0) + \phi(a^0) - \phi(a^*) \\ &< R(T_0). \end{aligned}$$

Therefore, the action state pair $[a^0, x^0]$ does not have minimum stochastic potential over all states and hence the contradiction. The same analysis can be repeated to show that all action state pairs $[a^*, x]$ where $a^* \in \arg \max_{a \in \mathcal{A}} \phi(a)$ and $x \in X(a^*)$ are stochastically stable. \square

References

- Alos-Ferrer, C., & Netzer, N. (2010). The logit-response dynamics. *Games and Economic Behavior*, 68, 413–427.
- Arslan, G., Marden, J. R., & Shamma, J. S. (2007). Autonomous vehicle-target assignment: a game theoretical formulation. *ASME Journal of Dynamic Systems, Measurement and Control*, 129, 584–596.
- Arslan, G., & Shamma, J.S. (2004). Distributed convergence to Nash equilibria with local utility measurements. In *43rd IEEE conference on decision and control* (pp. 1538–1543).
- Blume, L. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5, 387–424.
- Blume, L. (1997). Population games. In B. Arthur, S. Durlauf, & D. Lane (Eds.), *The economy as an evolving complex system II* (pp. 425–460). Reading, MA: Addison-Wesley.
- Chen, H-L, Roughgarden, T., & Valiant, G. (2008). Designing networks with good equilibria. In *Proceedings of the nineteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 854–863).
- Fudenberg, D., & Levine, D. (1998). *The theory of learning in games*. Cambridge, MA: MIT Press.
- Gopalakrishnan, R., Marden, J. R., & Wierman, A. (2011). An architectural view of game theoretic control. *ACM SIGMETRICS Performance Evaluation Review*, 38(3), 31–36.
- Katti, S., Katabi, D., Hu, W., Rahul, H., & Médard, M. (2005). The importance of being opportunistic: practical network coding for wireless environments. In *33rd allerton annual conference on communications, control, and computing*. Monticello, IL: IEEE, invited paper.
- Li, N., & Marden, J.R. (2010). Designing games to handle coupled constraints. In *Proceedings of the 48th IEEE conference on decision and control*. December.
- Li, N., & Marden, J.R. (2011). Designing games for distributed optimization. In *Proceedings of the 49th IEEE conference on decision and control*. December.
- Marden, J. R., Arslan, G., & Shamma, J. S. (2009a). Connections between cooperative control and potential games. *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, 39, 1393–1407.
- Marden, J. R., Arslan, G., & Shamma, J. S. (2009b). Joint strategy fictitious play with inertia for potential games. *IEEE Transactions on Automatic Control*, 54, 208–220.
- Marden, J.R., Arslan, G., & Shamma, J.S. (2007). 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.
- Marden, J. R., & Shamma, J. S. (2012). Revisiting log-linear learning: asynchrony, completeness and a payoff-based implementation. *Games and Economic Behavior*, 75, 788–808.
- Marden, J.R., & Wierman, A. (2008). Distributed welfare games with applications to sensor coverage. In *Proceedings of the 46th IEEE conference on decision and control*. December.

- Marden, J.R., & Wierman, A. (2009). Overcoming limitations of game-theoretic distributed control. In *Proceedings of the 47th IEEE conference on decision and control*. December.
- Marden, J. R., Young, H. P., Arslan, G., & Shamma, J. S. (2009). Payoff based dynamics for multi-player weakly acyclic games. *SIAM Journal on Control and Optimization*, 48, 373–396.
- Martinez, S., Cortes, J., & Bullo, F. (2007). Motion coordination with distributed information. *Control Systems Magazine*, 27(4), 75–88.
- Monderer, D., & Shapley, L. (1996). Potential games. *Games and Economic Behavior*, 14, 124–143.
- Murphey, R. A. (1999). Target-based weapon target assignment problems. In P. M. Pardalos, & L. S. Pitsoulis (Eds.), *Nonlinear assignment problems: algorithms and applications* (pp. 39–53). Alexandria, Virginia: Kluwer Academic.
- Olfati-Saber, R., Fax, J. A., & Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*.
- Pao, L., & Johnson, K.E. (2009). A tutorial on the dynamics and control of wind turbines and wind farms. In *Proceedings of the 2009 american control conference*.
- Pradelski, B. R., & Young, H. P. (2012). "Learning Efficient Nash Equilibria in Distributed Systems" (with Bary S.R. Pradelski). *Games and Economic Behavior*, 75, 882–897.
- Roughgarden, T. (2005). *Selfish routing and the price of anarchy*. Cambridge, MA, USA: MIT Press.
- Shah, D., & Shin, J. (2010). Dynamics in congestion games. *ACM SIGMETRICS*.
- Shamma, J. S., & Arslan, G. (2005). Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3), 312–327.
- Shapley, L. S. (1953). Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America*, 39(10), 1095–1100.
- Tsitsiklis, J.N. (1987). Decentralized detection by a large number of sensors. Technical report. MIT, LIDS.
- Wolpert, D., & Tumor, K. (1999). An overview of collective intelligence. In J. M. Bradshaw (Ed.), *Handbook of agent technology*. AAAI Press/MIT Press.
- Young, H. P. (1998). *Individual strategy and social structure*. Princeton, NJ: Princeton University Press.
- Young, H. P. (2005). *Strategic learning and its limits*. Oxford University Press.
- Young, H. P. (1993). The evolution of conventions. *Econometrica*, 61(1), 57–84.
- Young, H. P. (2009). Learning by trial and error. *Games and Economic Behavior*, 65, 626–643.



Jason R. Marden is an Assistant Professor in the Department of Electrical, Computer, and Energy Engineering at the University of Colorado. He received a B.S. degree in mechanical engineering in 2001 from UCLA, and a Ph.D. in mechanical engineering in 2007, also from UCLA, where he was awarded the Outstanding Graduating Ph.D. Student in Mechanical Engineering. After graduating from UCLA, he served as a junior fellow in the Social and Information Sciences Laboratory at the California Institute of Technology until 2010 after which he joined the University of Colorado. He received the American Automatic Control Council Donald P. Eckman Award in 2012, an AFOSR Young Investigator Award in 2012, and his student's paper was a finalist for the Best Student Paper Award at the IEEE Conference on Decision and Control in 2011. His research interests focus on game theoretic methods for distributed and networked control systems.