

Annual Review of Control, Robotics, and Autonomous Systems

Game Theory and Control

Jason R. Marden¹ and Jeff S. Shamma²

¹Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA; email: jrmarden@ece.ucsb.edu

²Computer, Electrical, and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; email: jeff.shamma@kaust.edu.sa

Annu. Rev. Control Robot. Auton. Syst. 2018. 1:105–34

First published as a Review in Advance on November 9, 2017

The Annual Review of Control, Robotics, and Autonomous Systems is online at control.annualreviews.org

https://doi.org/10.1146/annurev-control-060117-105102

Copyright © 2018 by Annual Reviews. All rights reserved



Explore related articlesSearch keywords

Keywords

game theory, zero-sum games, team games, distributed control, learning in games, mechanism design

Abstract

Game theory is the study of decision problems in which there are multiple decision makers and the quality of a decision maker's choice depends on both that choice and the choices of others. While game theory has been studied predominantly as a modeling paradigm in the mathematical social sciences, there is a strong connection to control systems in that a controller can be viewed as a decision-making entity. Accordingly, game theory is relevant in settings with multiple interacting controllers. This article presents an introduction to game theory, followed by a sampling of results in three specific control theory topics where game theory has played a significant role: (a) zero-sum games, in which the two competing players are a controller and an adversarial environment; (b) team games, in which several controllers pursue a common goal but have access to different information; and (c) distributed control, in which both a game and online adaptive rules are designed to enable distributed interacting subsystems to achieve a collective objective.

1. INTRODUCTION AND OVERVIEW

In his 1997 textbook, Myerson (1, p. 1) defines game theory as "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers." From a control theory perspective, one can interpret a control law as a sort of intelligent rational decision maker that was designed to produce a desired effect. Accordingly, again from a control theory perspective, game theory can be viewed as the study of conflict and cooperation between interacting controllers, where the identities (who are the controllers?) and preferences (what are they trying to achieve?) depend on the specific setting.

One setting, which is one of conflict, where game theory and control theory have had a long-standing connection is strictly competitive games, more commonly known as zero-sum games. There are two players, and what is to the benefit of one player is to the detriment of the other. The connection to control theory is as follows. A driving motivation for feedback is to assure satisfactory performance in the face of an uncertain environment (see the robust control literature on this topic, e.g., as described in Reference 2). From a conflict perspective, the controller is one player and environmental uncertainty is the other player.

Another setting, which is one of cooperation, that is connected to control theory is team games. Team games can include many players, and their defining feature is that all players have the same preferences, typically expressed as a utility function. The complication is that different players have access to different information, and this distribution of information prohibits a conventional centralized controller implementation.

A third setting, which is one of neither strict conflict nor cooperation, is where a controller can be viewed as a collection of interacting subsystems that are distributed over an operating domain with limited communication capabilities (mobile sensor networks, robotic swarms, etc.). Each subsystem can be viewed as a player in a game, but now the challenge is to design the game itself (i.e., both preferences and online adaptation rules) to evoke a desirable emergent behavior of the collective subsystems (see 3, 4).

This review begins with a basic introduction to the elements of game theory and proceeds with a limited sampling of selected results in each of these three settings. The analysis and theorems stated herein are presented informally; technical details can be found in the associated references.

2. BASIC CONCEPTS AND NOTATION

We begin with a review of some basic material. The discussion is deliberately brief, given the many texts on the topic of game theory directed at different audiences, including economists (1, 5, 6), computer scientists (7–9), and engineers (10–12).

Standard notation and basic concepts from control theory are used throughout, with clarifications provided as needed. Some specific notation is as follows. The symbol \mathbb{Z}_+ denotes the set of nonnegative integers, $\{0,1,2,\ldots\}$. For a set \mathcal{X} , $2^{\mathcal{X}}$ denotes the (power) set of subsets. For a function $f: \mathbb{R}_+ \to \mathbb{R}^n$, define $\|f\|_2 = \left(\int_0^\infty f^{\mathrm{T}}(t)f(t)\,\mathrm{d}t\right)^{1/2}$. A function f is in \mathcal{L}_2 if $\|f\|_2 < \infty$. For a vector $x \in \mathbb{R}^n$, $|x|_\infty = \max_{i=1}^n |x_i|$. The notation $\xi \sim \mathcal{N}(x,X)$ means that ξ is a Gaussian random variable with mean x and covariance X.

2.1. Game Elements

A game is described by three elements. First, there is the set of players, \mathcal{P} . We limit the discussion to a finite set of players, i.e.,

$$\mathcal{P} = \{1, 2, \dots, P\}.$$

For each player, $p \in \mathcal{P}$, there is a set of strategies, S_p . The joint strategy set is

$$S = S_1 \times \cdots \times S_p$$
.

A joint strategy $s \in S$ is represented as

$$s = (s_1, s_2, \ldots, s_p).$$

We sometimes represent a joint strategy from the perspective of a specific player, $p \in \mathcal{P}$, as

$$s = (s_1, \ldots, s_p, \ldots, s_p) = (s_p, s_{-p}).$$

Here, s_{-p} denotes the set of strategies of players in \mathcal{P} or p, i.e., players other than player p. Finally, for each player, $p \in \mathcal{P}$, there is a utility function

$$u_{b}: \mathcal{S} \to \mathbb{R}$$

that captures the player's preferences over joint strategies. That is, for any two joint strategies, $s, s' \in S$, player p strictly prefers s to s' if and only if

$$u_p(s) > u_p(s');$$

i.e., larger is better. If $u_p(s) = u_p(s')$, then player p is indifferent between joint strategies s and s'. The vector of utility functions is denoted by u, i.e.,

$$u = (u_1, u_2, \ldots, u_P) : \mathcal{S} \to \mathbb{R}^P$$

It is sometimes more convenient to express a game in terms of cost functions rather than utility functions. In this case, for each player, $p \in \mathcal{P}$, there is a cost function

$$c_n: \mathcal{S} \to \mathbb{R}$$
,

and player p prefers the joint strategy s to s' if and only if

$$c_p(s) < c_p(s');$$

i.e., smaller is better.

2.2. Examples

A game is fully described by the triplet of (a) the player set, \mathcal{P} ; (b) the joint strategy set, \mathcal{S} ; and (c) the vector of utility functions, u (or cost functions, c). The following examples illustrate the versatility of this framework.

2.2.1. Two-player matrix games. In a two-player matrix game, there are two players: a row player and a column player. Each player has a finite set of actions. Accordingly, the utility functions can be represented by two matrices, M_{row} and M_{col} , for the row and column players, respectively. The elements $M_{\text{row}}(i,j)$ and $M_{\text{col}}(i,j)$ indicate the utility to each player when the row player selects its ith strategy and the column player selects its jth strategy.

Such games admit the convenient visualization shown in **Figure 1***a*. In this illustration, the row player's strategy set is labeled {T, B}, and the column player's strategy set is {L, R}. In this

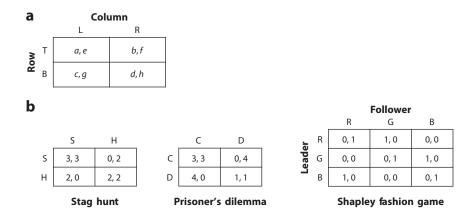


Figure 1

Two-player matrix games. (a) A basic visualization. The players here are labeled "row" and "column"; the row player's strategies are denoted as top (T) and bottom (B), and the column player's strategies are denoted as left (L) and right (R). In each cell, the first variable is the payoff for the row player, and the second is the payoff for the column player. (b) Three examples of two-player matrix games. In a stag hunt, the players can choose to hunt either a stag (S) or a hare (H); a stag is worth more than a hare, but successfully hunting a stag requires the cooperation of the other player. In the prisoner's dilemma, the players can either cooperate (C) or defect (D); choosing to defect when the other player chooses to cooperate provides the largest payoff for the defecting player, but a defection by both players results in a smaller payoff than mutual cooperation. In the Shapley fashion game, the players can choose to wear red (R), green (G), or blue (B); the fashion leader (row player) wants to wear a color that contrasts with what the follower (column player) is wearing, whereas the follower wants to wear the same color as the leader.

case,

$$M_{\text{row}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
 and $M_{\text{col}} = \begin{pmatrix} e & f \\ g & b \end{pmatrix}$.

If the players have more than two strategies, then the matrix representations simply have more rows and columns. **Figure 1***b* shows representative two-player matrix games that have received considerable attention for their illustration of various phenomena.

2.2.2. Randomized (mixed) strategies. Continuing with the discussion of two-player matrix games, let us redefine the strategy sets as follows. For any positive integer n, define $\Delta[n]$ to be the n-dimensional probability simplex, i.e.,

$$\Delta[n] = \left\{ \pi \in \mathbb{R}^n : \pi_i \ge 0 \text{ and } \sum_{i=1}^n \pi_i = 1 \right\}.$$

Now redefine the strategy sets as follows. If the matrices M_{row} and M_{col} are of dimension $n \times m$, define

$$S_{\text{row}} = \Delta[n]$$
 and $S_{\text{col}} = \Delta[m]$.

The strategies of each player can be interpreted as the probabilities of selecting a row or column. The utility functions over these randomized strategies are now defined as

$$u_{\text{row}}(s_{\text{row}}, s_{\text{col}}) = s_{\text{row}}^{\text{T}} M_{\text{row}} s_{\text{col}},$$

$$u_{\text{col}}(s_{\text{row}}, s_{\text{col}}) = s_{\text{row}}^{\text{T}} M_{\text{col}} s_{\text{col}}.$$

These redefined utility functions can be identified with the expected outcome of the previous deterministic utility functions over independent randomized strategies.

2.2.3. Parallel congestion games. In a parallel congestion game, there is a collection of parallel roads from A to B,

$$\mathcal{R} = \{1, 2, \dots, R\}$$
.

Although there is a finite set of players, the number of players is typically much larger than the number of roads, i.e., $P \gg R$. Each player must select a single road, and so the strategy set of each player is \mathcal{R} . For each road, there is a function that expresses the congestion on that road as a function of the number of players using the road. That is, for each $r \in \mathcal{R}$, there is a road-specific congestion function

$$\kappa_r: \mathbb{Z}_+ \to \mathbb{R}.$$

A natural assumption is that the congestion functions are increasing. For the joint strategy s, let $\nu_r(s)$ denote the number of players such that $a_p = r$, i.e., the number of users of road r.

Players seek to avoid congestion. Accordingly, the cost function of player p depends on the selected road, r, and the road selection of other players, s_{-p} . In terms of the congestion functions, the cost function of player p evaluated at the joint strategy $s = (r, s_{-p})$ is

$$c_p(r, s_{-p}) = \kappa_r(\nu_r(r, s_{-p})),$$

i.e., the congestion experienced on the selected road, r.

Although this example is for parallel congestion games, one can define a similar setup for a network of roads and player-specific starting and destination points. For an extensive discussion of this topic, see Reference 13.

2.2.4. Repeated prisoner's dilemma. In the presentation of the prisoner's dilemma shown in **Figure 1***b*, the strategy set of each player is {C, D}, whose elements stand for cooperate and defect, respectively. We now consider a scenario in which the matrix game is played repeatedly over an infinite series of stages, t = 0, 1, 2, ... At stage t, both players simultaneously select an action in {C, D}. The actions are observed by both players, and the process repeats at stage t + 1 and so on. Let a(t) denote the joint actions of both players at stage t, so that a(t) can take one of four values: (C, C), (C, D), (D, C), or (D, D). At stage t, the observed history of play is

$$H(t) = (a(0), a(1), \dots, a(t-1)).$$

Let \mathcal{H}_* denote the set of such finite histories. In the repeated prisoner's dilemma, a strategy is a mapping

$$s:\mathcal{H}_*\to \{C,D\}.$$

In words, a strategy is a reaction rule that dictates the action a player will take as a function of the observed history of play. Two example strategies that have played an important role in the analysis of the prisoner's dilemma are grim trigger and tit-for-tat. In grim trigger, the player cooperates (i.e., selects C) as long as the other player has always cooperated:

$$a_p(t) = \begin{cases} C, & a_{-p}(\tau) \neq D, \forall \tau < t; \\ D, & \text{otherwise.} \end{cases}$$

In tit-for-tat, the player selects the action that the other player selected in the previous round:

$$a_p(t) = \begin{cases} C, & t = 0; \\ a_{-p}(t-1), & t \ge 1. \end{cases}$$

Here, p stands for either the row or column player, and as usual, -p stands for the other player.

To complete the description of the game, we still need to define the utility functions. Let $s = (s_{row}, s_{col})$ denote a pair of strategies for the row and column players. Any pair of strategies, s, induces an action stream, $a(0; s), a(1; s), \ldots$ Let $U_p(a(t; s))$ denote the single-stage utility to player p at stage t under the (strategy-dependent) joint action a(t; s). Now define the repeated game utility function:

$$u_p(s) = (1 - \delta) \sum_{t=0}^{\infty} \delta^t U_p(a(t; s)),$$

where the discount factor, δ , satisfies $0 < \delta < 1$. In words, the utility function is the future discounted sum of the single-stage payoffs.

The repeated prisoner's dilemma is an example of a dynamic game, which is a model of an evolving scenario. This example illustrates that the basic framework of Section 2.1 is versatile enough to accommodate (a) infinite strategy sets, since there are an infinite number of reaction rules, and (b) utility functions that reflect payoff streams that are realized over an infinite time.

2.3. Nash Equilibrium

Game theory has its origins in the mathematical social sciences as a modeling framework in settings that involve multiple decision-making actors. As a modeler, one would like to propose an outcome, or solution concept, once the sets of players, strategies, and utility functions have been specified. Nash equilibrium is one such proposal. Note that there is an extensive literature—which is not discussed here—that examines and critiques Nash equilibrium in the face of experimental evidence and presents a variety of alternative proposals (see 14, 15).

At a Nash equilibrium, each player's strategy is optimal with respect to the strategies of other players.

Definition 1. The set of strategies $(s_1^*, s_2^*, \dots, s_p^*) \in \mathcal{S}$ is a Nash equilibrium if for each $p \in \mathcal{P}$,

$$u_p(s_p^*, s_{-p}^*) \ge u_p(s_p, s_{-p}^*), \quad \forall s_p \in \mathcal{S}_p.$$

An alternative representation is in terms of best-response functions.

Definition 2. The best-response function $BR_p: \mathcal{S}_{-p} \to 2^{\mathcal{S}_p}$ is defined by

$$\mathrm{BR}_p(s_{-p}) = \{s_p \in \mathcal{S}_p : u_p(s_p, s_{-p}) \geq u_p(s_p', s_{-p}) \text{ for all } s_p' \in \mathcal{S}_p\}.$$

In words, $BR_p(s_{-p})$ is the set of strategies that maximize the utility of player p in response to the strategies s_{-p} of other players. Note that there need not be a unique maximizer, and so $BR_p(s_{-p})$ is, in general, set valued.

In terms of the best-response functions, a Nash equilibrium is a set of strategies, $(s_1^*, s_2^*, \dots, s_p^*) \in \mathcal{S}$, such that for all $p \in \mathcal{P}$,

$$s_p^* \in \mathrm{BR}_p(s_{-p}^*).$$

Let us now revisit the examples in Section 2.2.

2.3.1. Nash equilibria in two-player matrix games. For the stag hunt, there are two Nash equilibria, (S, S) and (H, H); for the prisoner's dilemma, the sole Nash equilibrium is (D, D); and for the Shapley fashion game, there is no Nash equilibrium. These examples illustrate that games can exhibit multiple, unique, or no Nash equilibria.

In the case of randomized strategies, the stag hunt inherits the same two Nash equilibria, now expressed as probability vectors:

$$s_{\text{row}} = s_{\text{col}} = \begin{pmatrix} \mathbf{Pr}[S] \\ \mathbf{Pr}[H] \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } s_{\text{row}} = s_{\text{col}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Furthermore, there is a new Nash equilibrium:

$$s_{\text{row}} = s_{\text{col}} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix}.$$

The prisoner's dilemma continues to have a unique Nash equilibrium. The Shapley fashion game now has a single Nash equilibrium:

$$s_{\text{row}} = s_{\text{col}} = \begin{pmatrix} \mathbf{Pr} [\mathbf{R}] \\ \mathbf{Pr} [\mathbf{G}] \\ \mathbf{Pr} [\mathbf{B}] \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}.$$

- **2.3.2.** Nash equilibria in parallel congestion games. At a Nash equilibrium, no player has an incentive to use a different road. Accordingly, all of the roads that are being used have (approximately) the same level of congestion.
- **2.3.3.** Nash equilibria in repeated prisoner's dilemma games. One can show that grim trigger versus grim trigger constitutes a Nash equilibrium in the case that the discount factor, δ , is sufficiently close to 1. The standard interpretation is that cooperation can be incentivized provided that the players have a long-term outlook. By contrast, the single-stage (or even finitely repeated) prisoner's dilemma has a unique Nash equilibrium of always defect versus always defect.

3. ZERO-SUM AND MINIMAX DYNAMIC GAMES

In zero-sum games, there are two players with strategy sets S_1 and S_2 , and the utility functions satisfy the zero-sum property:

$$u_1(s_1, s_2) + u_2(s_1, s_2) = 0, \quad \forall s_1 \in S_1, s_2 \in S_2.$$

Accordingly, an increase in utility for one player results in a decrease in utility for the other player (by the same amount). Given this special structure, zero-sum games are usually expressed in terms of a single objective function, $\phi(s_1, s_2)$, that is the cost function of player 1 (the minimizer) and

the utility function of player 2 (the maximizer). In terms of the original formulation,

$$\phi(s_1, s_2) = -u_1(s_1, s_2) = u_2(s_1, s_2).$$

Define

$$\frac{\underline{\mathrm{val}}[\phi] = \sup_{s_2 \in \mathcal{S}_2} \inf_{s_1 \in \mathcal{S}_1} \phi(s_1, s_2),}{\overline{\mathrm{val}}[\phi] = \inf_{s_1 \in \mathcal{S}_1} \sup_{s_2 \in \mathcal{S}_2} \phi(s_1, s_2).}$$

The quantity $\overline{\text{val}}[\phi]$ represents the best guaranteed cost for the minimizer in the worst-case scenario of its strategy, s_1 , being known to the maximizer. The quantity $\underline{\text{val}}[\phi]$ has an analogous interpretation for the maximizing player. In general,

$$\underline{\operatorname{val}}[\phi] \leq \overline{\operatorname{val}}[\phi].$$

In the case of equality, the zero-sum game is said to have a value:

$$\operatorname{val}[\phi] = \operatorname{val}[\phi] = \overline{\operatorname{val}}[\phi].$$

For zero-sum games, a Nash equilibrium is characterized by the following saddle-point condition. The pair of strategies s_1^* and s_2^* is a Nash equilibrium if

$$\phi(s_1^*, s_2') \le \phi(s_1^*, s_2^*) \le \phi(s_1', s_2^*), \quad \forall s_1' \in S_1, s_2' \in S_2.$$

This inequality indicates that s_1^* is the best response (for the minimizer) to s_2^* and vice versa. If a zero-sum game has a Nash equilibrium, then it has a value, with

$$val[\phi] = \phi(s_1^*, s_2^*).$$

The converse need not be true, depending on whether the associate infimum or supremum operations are achieved.

In control problems, one is often interested in a worst-case analysis against an environmental opponent. Taking the viewpoint that the control design is the minimizer, define

$$\phi_{\text{wc}}(s_1) = \sup_{s_2 \in S_2} \phi(s_1, s_2),$$

i.e., the worst-case outcome as a function of the strategy s_1 . Let us call a strategy, s_1 , that achieves

$$\phi_{wc}(s_1) < \ell$$

a security strategy that guarantees ℓ . It may be the case that one is interested solely in constructing a security strategy (for the sake of establishing a guaranteed level of performance) without determining the optimal ℓ (which is $\overline{\text{val}}[\phi]$) or even whether a value or Nash equilibrium exists.

3.1. Pursuit-Evasion Games

Pursuit–evasion games have long received attention in the controls literature. A very early summary article was written by Ho et al. (16), who also directed readers to the seminal work of Isaacs (17). The following discussion follows the work of Başar & Olsder (10, their section 8.2].

As the terminology implies, in this game, a mobile pursuer is trying to approach a mobile evader. A typical modeling setup is that the pursuer (labeled "p") and evader (labeled "e") are

modeled by controlled differential equations:

$$\dot{x}^{p} = f^{p}(x^{p}, u), \quad x^{p}(0) = x_{0}^{p},$$

 $\dot{x}^{e} = f^{e}(x^{e}, v), \quad x^{e}(0) = x_{0}^{e}.$

The pursuer and evader control inputs $u(t) \in \mathcal{U}$ and $v(t) \in \mathcal{V}$, respectively, which may be subject to various constraints that model effects such as bounded velocity, acceleration, or turning radius, as captured by the sets \mathcal{U} and \mathcal{V} . Let $x = (x^p \quad x^e)$ denote the concatenated state vector. The evader is considered to be captured by the pursuer at time t if

$$\ell(x(t), t) = 0,$$

where the function $\ell(\cdot, \cdot)$ can represent events such as the pursuer being sufficiently close to the evader.

An example is the so-called homicidal chauffeur problem (17), where the pursuer is faster than the evader but the evader is more agile. These constraints are modeled in two dimensions as follows. The pursuer is subject to Dubins (or unicycle) dynamics:

$$\dot{x}_1^{\mathrm{p}} = V^{\mathrm{p}} \cos(\theta),$$

$$\dot{x}_2^{\mathrm{p}} = V^{\mathrm{p}} \sin(\theta),$$

$$\dot{\theta} = u,$$

with |u(t)| < 1. The evader is a simple integrator:

$$\dot{x}_1^e = v_1,$$

 $\dot{x}_2^e = v_2.$

The constraints are $\sqrt{v_1^2(t) + v_2^2(t)} \le V^e$, where V^e is the maximum evader velocity and $V^e < V^p$. We are interested in finding closed-loop strategies for the pursuer and evader of the form

$$u(t) = s^{p}(t, x^{p}(t), x^{e}(t)),$$

 $v(t) = s^{e}(t, x^{p}(t), x^{e}(t)),$

which are time- and state-dependent feedback laws. One can also formulate alternative variations, such as partial observations of the opponent's state, in which case a closed-loop strategy may be a function of the history of observations.

Finally, we define an objective function of the form

$$\phi(s^{p}, s^{e}) = \int_{0}^{T} g(t, x(t), u(t), v(t)) dt + q(T, x(T)),$$
 2.

where the final time, T, is defined as the shortest time that satisfies the termination condition of Equation 1. An example setup from Ho et al. (16) is

$$\phi(s^{p}, s^{e}) = x^{T}(T)Qx(T) + \int_{0}^{T} u(t)R^{p}u(t) - v^{T}(t)R^{e}v(t) dt,$$

$$\ell(x,t) = T_{\text{max}} - t$$
.

In this formulation, the game ends in a specified terminal time, T_{max} . The terminal objective is some quadratic function of the pursuer and evader states, such as the norm of the distance between

the pursuer and evader locations. The integrated objective penalizes the weighted control energy of both the (minimizing) pursuer and (maximizing) evader.

Of course, the remaining challenge is to characterize strategies that constitute a Nash equilibrium. Note that for a fixed strategy of the opponent, the remaining player is left with a one-sided optimal control problem (see 18). This insight leads to the Isaacs equation that defines a partial differential equation for an unknown value function, J(t, x). Let

$$f(x, u, v) = \begin{pmatrix} f^{p}(x^{p}, u) \\ f^{e}(x^{e}, v) \end{pmatrix}$$

denote the concatenated dynamics vector. The Isaacs equation is

$$-\frac{\partial J}{\partial t} = \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \left[\frac{\partial J}{\partial x} f(x, u, v) + g(t, x, u, v) \right]$$

$$= \max_{v \in \mathcal{V}} \min_{u \in \mathcal{U}} \left[\frac{\partial J}{\partial x} f(x, u, v) + g(t, x, u, v) \right].$$
3.

If a solution to the Isaacs equation exists, then one can define strategies $s^p(t, x)$ and $s^e(t, x)$ as the associated minimizer and maximizer, respectively.

A sufficient condition for the existence of a Nash equilibrium is the following theorem.

Theorem 1 (from Reference 10, theorem 8.1). Assume that there exists a continuously differentiable function, J(t,x), that satisfies (a) the Isaacs equation shown as Equation 3 and (b) the boundary condition J(T,x) = q(T,x) on the set $\ell(T,x) = 0$. Assume further that the associated trajectories induced by $u(t) = s^p(t,x)$ and $v(t) = s^e(t,x)$ generate trajectories that terminate in finite time. Then the pair (s^p, s^e) constitutes a Nash equilibrium.

Theorem 1 leaves open the issue of constructing a solution to the Isaacs equation, which can be computationally challenging even for low-dimensional problems. This issue has motivated a variety of constructive approaches.

3.1.1. Backwards reachability. Suppose that the capture condition of Equation 1 does not depend on time but is only a function of the pursuer and evader states. Now define the goal set

$$\mathcal{G}_0 = \left\{ x : \ell(x) = 0 \right\}.$$

The backwards reachability problem (for the pursuer) is to find all states that can be driven to \mathcal{G}_0 regardless of the control inputs of the evader, i.e., a forced capture. Backwards reachability approaches are closely related to methods of set invariance (e.g., 19) and viability theory (e.g., 20) and are connected to the discussion of optimal disturbance rejection in Section 3.3.

It is convenient to describe the main idea using the discrete-time pursuit-evasion dynamics

$$x(t+1) = f(x(t), u(t), v(t)), \quad x(0) = x_0,$$

subject to constraints $u(t) \in \mathcal{U}$ and $v(t) \in \mathcal{V}$. First, define the following set:

$$G_1 = \{x : \exists u \in \mathcal{U} \text{ s.t. } f(x, u, v) \in G_0, \forall v \in \mathcal{V} \}.$$

In words, G_1 is the set of all states that can be forced (i.e., regardless of the evader's action, v) to the target set in one time step. Proceeding in a similar manner, recursively define

$$\mathcal{G}_{k+1} = \{x : \exists u \in \mathcal{U} \text{ s.t. } f(x, u, v) \in \mathcal{G}_k, \forall v \in \mathcal{V}\}.$$

By construction, if ever $x_0 \in \mathcal{G}_k$ for some k, then the state trajectory starting from an initial condition of x_0 can be forced to the target set in k time steps, again regardless of the evader's future actions. Alternatively, if an initial condition x_0 does not lie in $\bigcup_{k=1}^{\infty} \mathcal{G}_k$, then the evader can perpetually avoid the target set.

Mitchell et al. (21) developed these ideas for continuous-time dynamics. In particular, they associated backwards reachability sets with a terminal-penalty pursuit–evasion problem [the integral cost $g(\cdot) = 0$ in Equation 2] and represented the set \mathcal{G}_{τ} (for nondiscrete values of τ) as the level set of a numerically constructed function.

3.1.2. Approximate methods. The backwards reachability approach seeks to exactly characterize all states that lead to capture. An alternative is to exploit special simplified settings to derive a specific strategy that guarantees capture (not necessarily optimizing any specific criterion). One approach, taken by Zhou et al. (22), is to compute the Voronoi cell of the evader, i.e., the set of points that the evader can reach before the pursuer. The control law derived by Zhou et al. (22) minimizes the instantaneous time rate of change of the area of the evader's Voronoi cell. The authors went on to derive the minimizing control law and establish that this control law guarantees eventual capture, where capture is defined as the proximity condition $||x^p - x^e|| \le \delta$ for some capture radius δ .

Karaman & Frazzoli (23) considered a probabilistic approach to compute an escape path for an evader, if one exists. The algorithm is based on random trajectory generation for both the evader and pursuer. The outcome is an open-loop trajectory (i.e., not a feedback law) that, with high probability, guarantees escape for all trajectories of the pursuer. Here, the term high probability refers to a high probability of the algorithm producing a guaranteed escape trajectory—if one exists—but the movements of the evader and pursuer are not randomized. Randomized pursuit trajectories were considered by Isler et al. (24). The game is played in a nonconvex polygonal region, where the pursuer's objective is to have the evader in its line of sight. The authors showed that randomized open-loop pursuit trajectories can lead to capture in settings where deterministic trajectories can be perpetually evaded.

Note that backwards reachability and approximate methods are more aligned with the abovementioned worst-case or minimax formulation in that they do not attempt to address the construction of a pair of strategies that constitutes a Nash equilibrium. Rather, the objective is to derive a strategy for one of the players that guarantees a desired outcome regardless of the behavior of the other player.

3.2. Zero-Sum Linear-Quadratic Games and Robust Control

One of the foundational motivations for feedback control is to mitigate the effects of environmental uncertainty, which can come in the form of exogenous disturbances, model misspecification, and parameter variations, among others. The robust control program (e.g., 2) sought to address such issues as optimization problems. Quoting the seminal work of Zames (25, p. 301), "can the classical 'lead-lag' controllers be derived from an optimization problem?"

There is an immediate connection to zero-sum or minimax games, where the controller to be designed is in direct competition with the uncertain environment. In this section, we review a problem of induced-norm minimization that makes this connection explicit. For an extensive technical presentation and a historical overview, see Reference 26.

The induced-norm minimization problem is to design a controller that minimizes the effect of exogenous disturbances. This objective can be stated as follows. Let T[K] denote an operator that represents a closed-loop system as a function of the controller, K. For example (see 2),

$$T[K] = \begin{pmatrix} W_1(I + PK)^{-1} \\ W_2K(I + PK)^{-1} \end{pmatrix},$$

where P is the plant to be controlled and W_1 and W_2 are dynamic weighting operators (e.g., frequency-shaping filters).

For any stabilizing controller, define the induced norm

$$||T[K]||_{\text{ind},2} = \sup_{v \in \mathcal{L}_2} \frac{||T(K)v||_2}{||v||_2}.$$

We are interested in the optimization

$$\min_{K \text{ stabilizing}} ||T[K]||_{\text{ind,2}}$$

In addition to the face-value norm minimization, this objective function is also relevant to the problem of robust stabilization, i.e., stabilization in the presence of dynamic modeling uncertainty (see 27). One can also formulate optimal estimation problems in a similar manner (e.g., 28).

An example setup is the case of state feedback controllers. For the linear system,

$$\dot{x} = Ax + Bu + Lv, \quad x(0) = x_0,$$

$$z = \begin{pmatrix} Cx \\ u \end{pmatrix},$$

u is the control signal, v is an exogenous disturbance, and z gathers signals to be made small in the presence of disturbances (tracking error, control authority, etc.). A state feedback controller, $K \in \mathbb{R}^{m \times n}$, is stabilizing if A - BK is a stability matrix. (The eigenvalues have strictly negative real parts.) Given a stabilizing state feedback, the closed-loop system is

$$\dot{x} = (A - BK)x + Lv, \quad x(0) = x_0,
z = \begin{pmatrix} Cx \\ -Kx \end{pmatrix},$$
4.

and the resulting induced norm is

$$\sup_{v \in \mathcal{L}_2} \frac{\|z\|_2}{\|v\|_2}$$

along solutions of Equation 4 with zero initial conditions ($x_0 = 0$). Clearly, this objective depends on the stabilizing K.

To make the connection to zero-sum games, we modify the problem slightly. Rather than attempting to minimize the norm, we ask the question of whether it is possible to make the induced norm less than some specified level, $\gamma > 0$, as in

$$\sup_{v \in \mathcal{L}_2} \frac{\|z\|_2}{\|v\|_2} \le \gamma$$

or, equivalently,

$$\sup_{v \in \mathcal{L}_2} \|z\|^2 - \gamma^2 \|v\|^2 \le 0.$$

By iterating on γ , one can then seek to derive controllers that make γ as small as possible.

The associated objective function for the zero-sum game formulation, with u as the minimizer and v the maximizer, is

$$\phi(u,v) = \int_0^\infty x^{\mathrm{T}}(t)C^{\mathrm{T}}Cx(t) + u^{\mathrm{T}}(t)u(t)\,\mathrm{d}t - \gamma^2 \int_0^\infty v^{\mathrm{T}}(t)v(t)\,\mathrm{d}t$$
 5.

along solutions of Equation 4, where it is not assumed that $x_0 = 0$.

The following theorem combines the presentations of Başar & Bernhard (26, their theorem 4.11) and Limebeer et al. (29).

Theorem 2. Assume that the pair (A, B) is stabilizable and (A, C) is observable. The following statements are equivalent:

- 1. The zero-sum game Equation 5 has a value for any x_0 .
- 2. There exists a positive definite solution to the algebraic Riccati equation

$$XA + A^{\mathrm{T}}X + X\left(BB^{\mathrm{T}} - \frac{1}{\gamma^2}LL^{\mathrm{T}}\right) + C^{\mathrm{T}}C = 0$$

with $(A - (BB^{T} - \frac{1}{v^{2}}LL^{T})X)$ a stability matrix.

3. There exists a stabilizing state feedback K such that $||T[K]||_{ind,2} < \gamma$.

If any of these statements is satisfied, the state feedback $K = -BB^{T}X$ achieves $||T[K]||_{\text{ind},2} < \gamma$.

The above is just a sample of the extensive results concerning linear-quadratic games; for a dedicated discussion, see Reference 30.

3.3. Minimax Dynamic Programming

In the previous section, a large adversarial disturbance, v, was penalized in Equation 5 by the energy term

$$-\gamma^2 \int_0^\infty v^{\mathrm{T}}(t)v(t)\,\mathrm{d}t,$$

but otherwise the disturbance could have an arbitrarily large magnitude. Another model of an adversarial disturbance is that it takes values in a bounded domain. This unknown-but-bounded approach has its origins in References 31 and 32 and is related to set-invariance methods (19), viability theory (20), and the backwards reachability method discussed in Section 3.1.

The discussion here follows Bertsekas (33, his sections 1.6 and 4.6.2). We consider discrete-time dynamics of the form

$$x(t+1) = f(x(t), u(t), v(t)), \quad x(0) = x_0,$$

subject to constraints $u(t) \in \mathcal{U}$ and $v(t) \in \mathcal{V}$. (For added generality, one can also have the constraint sets be state dependent.) We are interested in finding the (possibly time-dependent) state feedback

$$u(t) = \mu_t(x(t)) \in \mathcal{U}$$

to minimize

$$\max_{v(0),v(1),\dots,v(T-1)\in\mathcal{V}} \left[q(x(T)) + \sum_{t=0}^{T-1} g(x(t),\mu_t(x(t)),v(t)) \right]$$

over some time horizon, [0, 1, ..., T]. Let $\vec{\mu}$ denote a collection of state feedback laws $(\mu_0(\cdot), \mu_1(\cdot), ..., \mu_{T-1}(\cdot))$. The order of operations here is that the controller (or minimizer) commits to $\vec{\mu}$ and then the constrained disturbance reacts to maximize the cost.

The dynamic programming solution follows the recursive procedure of value iteration. Define

$$J_0(x) = q(x)$$

and recursively define the cost-to-go functions

$$J_k(x) = \min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \left(g(x, u, v) + J_{k-1}(f(x, u, v)) \right).$$

Then, by standard dynamic programming arguments, one can show that

$$J_T(x_0) = \min_{\vec{\mu} = (\mu_0, \dots, \mu_{T-1})} \max_{v(0), v(1), \dots, v(T-1) \in \mathcal{V}} \left[q(x(T)) + \sum_{t=0}^{T-1} g(x(t), \mu_t(x(t)), v(t)) \right].$$

Furthermore, the optimal policy at stage t is the minimizing

$$\mu_t(x(t)) = \arg\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \left(g(x(t), u, v) + J_{T - (t+1)}(f(x(t), u, v)) \right).$$

The connection to zero-sum games is that minimax dynamic programming provides a procedure to construct an optimal security strategy from the perspective of the minimizer.

3.3.1. Controlled invariance. The problem of controlled invariance is to maintain the state in a specified region in the presence of disturbances. In the case of linear systems, consider

$$x(t+1) = Ax(t) + Bu(t) + Lv(t),$$

subject to $|v|_{\infty} \le 1$ and $|u|_{\infty} \le u_{\max}$. Our objective is to maintain the state in the bounded region $|x|_{\infty} \le 1$. One can map this problem to the minimax dynamic programming approach by defining the stage and terminal costs as

$$q(x) = g(x, u, v) = \begin{cases} 0, & |x|_{\infty} \le 1; \\ 1, & \text{otherwise.} \end{cases}$$

Then an initial condition, x_0 , satisfies $J_T(x_0) = 0$ if and only if there exists a policy, $\vec{\mu}$, that assures that the state will satisfy $|x(t)|_{\infty} \le 1$ for all t = 0, 1, ..., T.

One can mimic the backwards reachability algorithm to construct a representation of the costto-go functions, $J_k(\cdot)$, as follows. Define

$$G_0 = \{x : |x|_{\infty} < 1\}$$

and then define

$$\mathcal{G}_1 = \left\{ x : \exists u \text{ with } |u|_{\infty} \le u_{\text{max}} \text{ s.t. } Ax + Bu + Lv \in \mathcal{G}_0, \forall |v| \le 1 \right\} \cap \mathcal{G}_0.$$

Then $x_0 \in \mathcal{G}_1$ if and only if $J_1(x_0) = 0$, or, equivalently, there exists a control action that keeps $|x(t)|_{\infty} \le 1$ for t = 0 and t = 1. Proceeding recursively, define

$$\mathcal{G}_k = \{x : \exists u \text{ with } |u|_{\infty} \le u_{\text{max}} \text{ s.t. } Ax + Bu + Lv \in \mathcal{G}_{k-1}, \forall |v| \le 1\} \cap \mathcal{G}_{k-1}.$$

Note that the only difference between this algorithm and that of backwards reachability is the recursive intersection. Although computationally costly, one can use polyhedral representations of these sets (e.g., 19). That is, one can recursively construct matrices, M_k , such that

$$\mathcal{G}_k = \{x : |M_k x|_{\infty} \le 1\}.$$

By construction, $J_T(x_0) = 0$ if and only if $x_0 \in \mathcal{G}_T$. Furthermore, subject to certain technical conditions (32), if $x_0 \in \bigcap_{k=0}^{\infty} \mathcal{G}_k$, then there exists a policy that keeps $|x(t)|_{\infty} \le 1$ indefinitely.

It turns out that the controlled invariance problem is also related to minimizing an induced norm of the form

$$\left\|T\left[K\right]\right\|_{\mathrm{ind},\infty}=\sup_{v\in\mathcal{L}_{\infty}}\frac{\left\|T\left(K\right)v\right\|_{\infty}}{\left\|v\right\|_{\infty}},$$

where, for a function $f: \mathbb{Z}_+ \to \mathbb{R}^n$,

$$||f||_{\infty} = \sup_{t=0,1} |f(t)|_{\infty}.$$

This optimization is known as ℓ_1 optimal control (see 34 and references therein). The connection between ℓ_1 optimal control and controlled invariance was investigated in References 35 and 36.

3.3.2. Linear-parameter-varying optimal control. Linear-parameter-varying systems are linear systems that depend on an exogenous time-varying parameter whose values are unknown a priori. Stemming from the connection to the nonlinear design methodology of gain scheduling, the linear-parameter-varying paradigm can serve as a middle-ground model between linear time-invariant and nonlinear models (for an overview and further discussion, see 37).

Consider the controlled discrete-time linear-parameter-varying dynamics

$$x(t+1) = A(\theta(t))x(t) + Bu(t), \quad x(0) = x_0.$$
 6.

For this discussion, we assume that the parameter takes on finite values.

$$\theta(t) \in \mathcal{Q}, \quad \forall t = 0, 1, 2, \dots,$$

for some finite set, Q. One can also add constraints like bounds on the rate of change, such as $|\theta(t+1) - \theta(t)| \leq \delta_{\text{max}}$.

We are interested in deriving a collection of state feedback laws $(\mu_0(\cdot), \mu_1(\cdot), \dots, \mu_{T-1}(\cdot))$ to optimize

$$\sup_{\theta(0),\theta(1),\dots,\theta(T-1)\in\mathcal{Q}}\sum_{t=0}^{T-1}g(x(t),\mu_t(x(t)))$$

along solutions of Equation 6. This structure falls within the above-mentioned minimax dynamic programming, where the game is between the control, u, and parameter, θ .

The following results, along with explicit computational methods, were presented in Reference 38.

Theorem 3. Assume that the stage cost has the polyhedral representation

$$g(x,u) = \left| M_g x + N_g u \right|_{\infty}.$$

Then there exists a sequence of matrices, M_k , so that the cost-to-go functions of minimax dynamic programming satisfy

$$J_k(x) = |M_k x|_{\infty}$$
.

Furthermore, the stationary (receding horizon) feedback policy

$$\mu_T(x(t)) = \arg\min_{u} \max_{\theta} \left(g(x(t), u) + J_T \left(A(\theta)x(t) + Bu \right) \right)$$

asymptotically approximates the optimal infinite horizon policy for sufficiently large horizon lengths, ${\cal T}$.

3.4. Asymmetric Information and Randomized Deception

Thus far, we have assumed an environmental opponent that is all knowing in that we guard against all possible opponent actions after committing to a control policy. Another formulation of interest is where the opposing environment has limited information.

A specific mathematical model is the setting of repeated zero-sum matrix games. In a one-shot zero-sum game, there is a matrix, $M \in \mathbb{R}^{I \times J}$, whose ij th element is the penalty to the (minimizing) row player and payoff to the (maximizing) column player when the row player uses its ith (out of I) action and the column player uses its jth (out of J) action.

In a repeated zero-sum game, this game is repeated over a sequence of stages, $t=0,1,2,\ldots$ (see Section 2.2.4), and a strategy is a mapping from past actions over stages $0,1,\ldots,t-1$ to the selected action in stage t. In the case of asymmetric information, the characteristic matrix, M, is selected by nature from a collection of possible matrices, \mathcal{M} , as in

$$M \in \mathcal{M} = \{M^1, M^2, \dots, M^K\},$$

according to some known probability distribution, $\pi \in \Delta[K]$, where π_k is the probability that the matrix M_k is selected.

The informed player knows which matrix was selected, whereas the uninformed player does not. Furthermore, a common assumption is that the uninformed player cannot measure the realized payoff at each stage. Accordingly, the strategy of the informed player can depend on the selected characteristic matrix, whereas the strategy of the uninformed player can depend only on past actions. Following the notation of Section 2.2.4, let \mathcal{H}_* denote the set of finite histories of play. Then a (behavioral) strategy for the informed row player is a mapping

$$s_{\text{row}}: \mathcal{H}_* \times \mathcal{M} \to \Delta[P],$$

whereas a strategy for the uninformed column player depends only on the history of play, as in

$$s_{\rm col}:\mathcal{H}_*\to\Delta[Q].$$

This setup, as with more complicated variations, has received considerable attention, from the seminal work of Aumann et al. (39) through the more recent monograph by Mertens et al. (40).

A motivating setting is network interdiction (e.g., 41–43). In network interdiction problems, the activities of a network owner (defender) are being observed by an attacker. The

network owner is the informed player in that the owner knows the details of the network, e.g., which parts of the network are protected or susceptible. Based on these observations, the attacker launches a limited attack to disable a portion of the network. The network owner faces a trade-off of exploitation versus revelation. If the network owner exploits its superior information, it may reveal sensitive information to the attacker and thereby increase the likelihood of a critical attack. (For further discussion of game theory for network security, see 44.)

Returning to the repeated zero-sum game problem, it turns out that the optimal strategy of the informed player is to randomize its actions based on the following model of the uninformed player. First, assume that the uninformed player is executing Bayesian updates to compute the posterior probability

$$\pi_k^{\text{post}}[H(t)] = \mathbf{Pr}[M_k \mid H(t)],$$

which is the posterior probability that the matrix M_k was selected based on the observed sequence of past actions, H(t). The action of the uninformed player is then modeled as a myopically optimal randomized strategy with respect to these posterior probabilities:

$$\sigma_{\text{col}}(H(t)) = \arg\max_{\pi_{\text{col}} \in \Delta[Q]} \left(\min_{\pi_{\text{row}} \in \Delta[P]} \pi_{\text{row}}^{\text{T}} \left(\sum_{k=1}^{K} \pi_k^{\text{post}}[H(t)] M_k \right) \pi_{\text{col}} \right).$$

A classic result (see 39) is that the informed row player's optimal security strategy is a best response to the above column player strategy. The resulting strategy will randomize according to probabilities that depend on both the stage of play and the online outcome of prior stages. One can think of this strategy as a deception attempt to influence the posterior beliefs of the uninformed attacker. Recent work has presented efficient algorithms for the computation of these probabilities based on recursive linear programs (42, 43, 45, 46).

4. TEAM GAMES

We now discuss another class of games that has had a long-standing relationship with control theory: team games. In team games, there can be multiple players, 1, 2, ..., P. The defining characteristic is that all players have the same preferences. Recall that $S = S_1 \times \cdots \times S_P$ is the joint strategy set for players 1, 2, ..., P. In team games, 1 for some cost function, $C : S \to \mathbb{R}$,

$$c_1(s) = c_2(s) = \cdots = c_P(s) = \mathcal{C}(s).$$

While team games have been of interest for several decades (e.g., 47), there has been a recent surge of interest related to distributed control applications, where there is no single centralized controller that has authority over all control inputs or access to all measurements (e.g., 48–51) (for an alternative approach to distributed control, see Section 5).

The main issue in team games is that each player has access to different information. If all players had the same information, then one could approach the game as a conventional centralized control design problem. Players with different information can significantly complicate the characterization and, more importantly, the explicit and efficient computational construction of optimal strategies.

¹In accordance with much of the literature, we assume that players are minimizers.

Another issue in team games is the distinction between a team-optimal strategy and a Nash equilibrium. The joint strategy, s^* , is team optimal if

$$C(s^*) \leq C(s'), \quad \forall s' \in \mathcal{S},$$

but a Nash equilibrium if

$$\mathcal{C}(s_1^*,\ldots,s_p^*,\ldots,s_p^*) \leq \mathcal{C}(s_1^*,\ldots,s_p',\ldots,s_p^*), \quad \forall s_p' \in \mathcal{S}_p, \forall p \in \mathcal{P}.$$

In much of the team game literature, a Nash equilibrium is referred to as person-by-person optimal. Clearly, a team-optimal strategy is a Nash equilibrium, but the converse need not hold.

4.1. Static Linear-Quadratic Team Games

A classic result that dates back to work by Radner (52) is as follows [the discussion here follows Ho (47)]. There are P players, and player p can measure the scalar (for simplicity) random variable

$$z_p = H_p \xi,$$

where $\xi \in \mathbb{R}^n$ is Gaussian; $\xi \sim \mathcal{N}(0, X)$, with X > 0; and H_p is a matrix of appropriate dimension. The strategy of player p is a mapping

$$s_n:\mathbb{R}\to\mathbb{R}$$

from a player's measurement, z_p , to its actions, u_p , i.e., $u_p = s_p(z_p)$. The common cost function is

$$C(s_1,\ldots,s_p) = \mathbf{E}\left[\frac{1}{2}u^{\mathrm{T}}Qu + u^{\mathrm{T}}S\xi\right],$$

where matrices $Q = Q^{T} > 0$ and S are cost function parameters and

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_P \end{pmatrix} = \begin{pmatrix} s_1(z_1) \\ \vdots \\ s_P(z_P) \end{pmatrix}.$$

Theorem 4 (from Reference 52). The unique Nash equilibrium strategies are team optimal and linear, i.e.,

$$s_p^*(z_p) = F_p z_p$$

for matrices F_p .

For this specific case, person-by-person optimality also implies team optimality.

4.2. Information Structures

In more general models, the measurements of each player can depend on the actions of other players, as in

$$z_p = H_p \xi + \sum_{q=1}^{p} D_{pq} u_q$$
 7.

for matrices D_{pq} . A specific case that has received considerable attention is the Witsenhausen counterexample (53; see also 54). Here, there are two players, and the random variable $\xi = (\xi_1, \xi_2)$ is in \mathbb{R}^2 . The measurements are

$$z_1 = \xi_1,$$

 $z_2 = \xi_2 + u_1,$

and the joint cost function is

$$C(s_1, s_2) = \mathbf{E} \left[\alpha \cdot (u_1 - \xi_1)^2 + \beta \cdot (u_2 - u_1)^2 \right]$$

for positive weights α , $\beta > 0$. The standard interpretation of this example is that player 1, through its action u_1 , faces the dual objective to either cancel the disturbance effect of ξ_1 or signal its own value to u_2 to overcome the noise effect of ξ_2 .

Despite the linear–quadratic look and feel of this example, Witsenhausen (53) showed that there exist nonlinear team-optimal strategies that strictly outperform the best linear strategies. The team optimality of nonlinear strategies leads to the question of when linear strategies are team optimal, which ultimately relates to the desire to explicitly construct team-optimal strategies. For example, one well-known case is that of partially nested information. Continuing with the general formulation of Equation 7, suppose that the measurements can be decomposed as follows:

$$z_p = egin{pmatrix} z_q, & \forall q \in \mathcal{O}_p \ H_p \xi \end{pmatrix},$$

where the sets \mathcal{O}_p are defined as

$$\mathcal{O}_p = \{ q \mid D_{pq} \neq 0 \}.$$

In words, the measurement of player p includes the information of player q whenever the action of player q affects the measurement of player p (as indicated by $D_{pq} \neq 0$).

Theorem 5 (from Reference 55). Under partially nested information, team-optimal strategies are unique and linear.

Again, person-by-person optimality implies team optimality.

4.3. Common Information

A more recent development for team games exploits when players have some measurement information in common (56, 57). To establish the approach, consider a variation of the setup in Section 4.1. Let ξ be a random variable, and let z_0, z_1, z_2 be three random variables that are functions of ξ .

For this discussion, there are two players. Player 1 measures (z_0, z_1) , and player 2 measures (z_0, z_2) . As before, each player's strategy is a mapping from its measurements to an action, and so

$$u_1 = s_1(z_0, z_1),$$

 $u_2 = s_2(z_0, z_2).$

The common cost function is

$$\mathcal{C}(s_1, s_2) = \mathbf{E} \left[\ell(\xi, u_1, u_2) \right].$$

We see that both players have access to the common information, z_0 .

The approach of Nayyar et al. (56, 57) is to transform the team problem into a centralized optimization as follows. First, let z_0 , z_1 , z_2 , u_1 , and u_2 take values in the sets \mathcal{Z}_0 , \mathcal{Z}_1 , \mathcal{Z}_2 , \mathcal{U}_1 , and \mathcal{U}_2 , respectively. Define \mathcal{S}_1 to be the set of mappings

$$\mathcal{S}_1 = \{ \phi_1 \mid \phi_1 : \mathcal{Z}_1 \to \mathcal{U}_1 \} .$$

Likewise, define

$$S_2 = \{ \phi_1 \mid \phi_2 : \mathcal{Z}_2 \to \mathcal{U}_2 \}.$$

Finally, let \mathcal{R} denote the set of mappings

$$\mathcal{R} = \{ \rho \mid \rho : \mathcal{Z}_0 \to \mathcal{S}_1 \times \mathcal{S}_2 \}.$$

Now imagine a hypothetical coordinator that has access to the common information, z_0 . Upon obtaining a measurement, the coordinator recommends a pair of mappings $r_1(\cdot; z_0)$ and $r_2(\cdot; z_0)$ to the two players to apply to their private information, z_1 and z_2 , respectively. The optimization problem for the coordinator is then

$$\min_{(r_1,r_2)\in\mathcal{R}} \mathbf{E}\left[\ell(\xi,r_1(z_1;z_0),r_2(z_2;z_0))\right],$$

i.e., to derive the optimal recommended mappings. The main results of Nayyar et al. (56, 57) establish conditions under which the optimal recommendations, $(r_1^*(\cdot; z_0), r_2^*(\cdot; z_2))$, constitute teamoptimal strategies, with the associations

$$s_1^*(z_0, z_1) = r_1^*(z_1; z_0),$$

 $s_2^*(z_0, z_2) = r_2^*(z_2; z_0).$

4.4. Dynamic Settings

The discussion thus far has been restricted to static games, where there is no explicit underlying dynamical system. (Indeed, dynamic team games can often be mapped to a static formulation by associating the control actions over different times as different players.) The question of the structure of team-optimal controllers in dynamic games also has received significant interest. A prototypical setup (here with two players) is stochastic linear dynamics:

$$x(t+1) = Ax(t) + B_1u_1(t) + B_2u_2(t) + Lw(t), x(0) \sim \mathcal{N}(x_0, \Sigma),$$

$$y_1(t) = C_1x(t) + D_1w(t),$$

$$y_2(t) = C_2x(t) + D_2w(t),$$

where a strategy for player 1, s_1 , is a dynamical system mapping measurements, $y_1(\cdot)$, to control actions, $u_1(\cdot)$. Likewise, a strategy for player 2, s_2 , is a dynamical system mapping measurements, $y_2(\cdot)$, to control actions, $u_2(\cdot)$. The common cost function to be minimized is

$$C(s_1, s_2) = \mathbf{E} \left[\sum_{t=0}^{T-1} x(t)^{\mathrm{T}} Q x(t) + u_1(t)^{\mathrm{T}} u_1(t) + u_2(t)^{\mathrm{T}} u_2(t) + x(T)^{\mathrm{T}} S x(T) \right]$$

for some positive definite weighting matrices Q, S, and Σ .

Of particular interest is when the structure of the dynamics and measurements admit linear team-optimal strategies whose computation is tractable. Examples include one-step delay information sharing (e.g., 58–60), funnel causality (61), and quadratic invariance (62).

5. GAME-THEORETIC DISTRIBUTED CONTROL

In this section, we begin to view game theory from a design perspective as opposed to the primarily analytical focus discussed in the preceding sections. Our motivation for this new direction is the problem of distributed control, which focuses on systems where decision making is distributed throughout the system. More formally, each system comprises several individual subsystems, each constrained to making independent decisions in response to locally available information. Some examples are the following:

- Networking routing: In network routing, the goal of a system operator is to allocate demand over a network to optimize a given measure of performance, e.g., throughput (63, 64). In most situations, centralized routing policies are infeasible because routing decisions are made at a far more local level. Lasaulce et al. (65) and Ozdaglar & Menache (66) provide a broader discussion of network applications.
- Wind farms: Wind farm control seeks to optimize the power production. Employing control
 strategies where the individual turbines independently optimize their own power production
 is not optimal with regard to optimizing the power production in the wind farm (67–70).
- Area coverage: In coverage problems, a collection of mobile sensors in an unknown environment seeks to maximize the area under surveillance. Furthermore, some portions of the environment may carry more weight than others. Applications range from following a drifting spill to intruder detection (71–73).

The overarching goal of such distributed control problems is to characterize admissible decisionmaking policies for the individual subsystems that ensure the emergent collective behavior is desirable.

The design of distributed control policies can be derived from a game-theoretic perspective, where the subsystems are modeled as players in a game with designed utility functions. We no longer view equilibrium concepts such as Nash equilibrium as plausible outcomes of a game. Rather, we view these equilibrium concepts as stable outcomes associated with distributed learning where the individual subsystems adjust their behavior over time in response to information regarding their designed utility function and the behavior of the other subsystems.

The following sections present highlights of a game-theoretic approach. Two recent overview articles are References 3 and 4.

5.1. Learning in Games

We begin by focusing on the question of how players can reach an equilibrium. To that end, we consider the framework of repeated one-shot games in which a game is repeated over time and players are allowed to revise their strategies in response to available information regarding previous plays of the game (74). More formally, consider any finite strategic form game with player set \mathcal{P} , finite strategy sets \mathcal{S} , and utility functions u. A repeated one-shot game yields a sequence of strategy profiles s(0), s(1), s(2), ..., where at any time $t \ge 1$, the strategy profile s(t) is derived from a set of decision-making rules $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_n)$ of the form

$$s_{h}(t) = \mathbf{d}_{h}(s(0), s(1), \dots, s(t-1); u_{h}),$$
 8.

meaning that each player adjusts its strategy at time *t* using knowledge of the previous decisions of the other players as well as information regarding the structural form of the player's utility function. Any learning algorithm that can be expressed in this form is termed uncoupled, as players are not allowed to condition their choice on information regarding the utility functions of other players. An example of a decision-making rule of this form is the well-studied best-response

dynamics, in which

$$s_{b}(t) \in \mathrm{BR}_{b}(s_{-b}(t-1)),$$
 9.

where $BR_p(\cdot)$ is the best-response set defined in Definition 2. In the best-response dynamics, each player selects a best response to the behavior of the other players at the previous play of the game.

An alternative class of learning algorithms to Equation 8 that imposes less informational demands on the players is termed completely uncoupled dynamics or payoff-based dynamics (see 75–77) and is of the form

$$s_b(t) = \mathbf{d}_b(u_b(s(0)), u_b(s(1)), \dots, u_b(s(t-1))).$$
 10.

In completely uncoupled dynamics, each agent is now given only the payoff that the agent received at each play of the game. A player is no longer able to observe the behavior of the other agent or access the utility that the agent would have received for any alternative choices at any stage of the game. In problems such as network routing, learning algorithms of the form shown in Equation 10 may be far more reasonable than those in Equation 8.

Ignoring the informational demands for the moment, the theory of learning in games has sought to establish whether there exist distributed learning algorithms of the form shown in Equation 8 that will always ensure that the resulting collective behavior reaches a Nash equilibrium (or its generalizations). Representative papers include References 74, 78, and 79 in the economic game theory literature and References 80–82 in the control literature.

The following theorem highlights an inherent challenge associated with distributed learning for Nash equilibrium in general.

Theorem 6 (from Reference 75). There are no natural dynamics of the form shown in Equation 8 that converge to a pure Nash equilibrium for all games.

The term natural used in Theorem 6, which we do not explicitly define here, seeks to eliminate learning dynamics that exhibit phenomena such as exhaustive search or centralized coordination. A key point regarding this negative result is the phrase "for all games," which effectively means that there are complex games for which no natural dynamics exist that converge to a Nash equilibrium. A central question that we seek to address here is whether a system designer can exploit the freedom to design the players' utility functions to avoid this negative result.

5.2. Game Structures

In this section, we turn our attention to games that possess an underlying structure that negates the negative result given in Theorem 6. One such class of games, termed potential games (83), is defined as follows.

Definition 3. A game $(\mathcal{P}, \mathcal{S}, u)$ is a potential game if there exists a potential function $\phi : \mathcal{S} \to \mathbb{R}$ such that for any strategy profile $s \in \mathcal{S}$, player $p \in \mathcal{P}$, and strategy $s_p' \in \mathcal{S}_p$,

$$u_p(s_p', s_{-p}) - u_p(s_p, s_{-p}) = \phi(s_p', s_{-p}) - \phi(s_p, s_{-p}).$$
 11.

In a potential game, each player's utility is directly aligned with a common potential function in the sense that the change in the utility that a player would receive by unilaterally deviating from a strategy s_p to a new strategy s_p' when all other players are selecting s_{-p} is equal to the difference of the potential function over those two strategy profiles. Note that there is no mention of what happens to the utility of the other players $j \neq i$ across these two strategy profiles.

There are several desirable properties regarding potential games that are of interest to distributed control. First, a pure Nash equilibrium is guaranteed to exist in any potential game because any strategy $s \in \arg\max_{s \in S} \phi(s)$ is a pure Nash equilibrium. Second, there are natural dynamics that do in fact lead to a pure Nash equilibrium for any potential game. In fact, the following mild modification of the best-response dynamics given in Equation 12, termed best-response dynamics with inertia, accomplishes this goal:

$$s_p(t) \in \begin{cases} BR_p(s_{-p}(t)) & \text{with probability } (1 - \epsilon), \\ s_p(t - 1) & \text{with probability } \epsilon, \end{cases}$$
12.

where $\epsilon > 0$ is referred to as the player's inertia (79). More formally, the learning algorithm in Equation 12 converges almost surely to a pure Nash equilibrium in any potential game. Similar positive results also hold for relaxed versions of potential games, e.g., generalized ordinal potential games or weakly acyclic games (79), which seek to relax Equation 11 through either the equality or the condition for all players.

One of the other interesting facets of potential games is the availability of distributed learning algorithms that exhibit equilibrium selection properties, i.e., learning algorithms that favor one Nash equilibrium over other Nash equilibria. For example, consider the algorithm log-linear learning (see 84–87), where at each time $t \ge 1$ a single agent i is given the opportunity to revise its strategy, i.e., $s_{-p}(t) = s_{-p}(t-1)$, and this updating player adjusts its choice according to the following mixed strategy:

$$s_p(t) = s_p \text{ with probability } \frac{e^{\beta \cdot u_p(s_p, s_{-p}(t-1))}}{\sum_{s_p \in S_p e^{\beta \cdot u_p(s_p', s_{-p}(t-1))}}},$$
13.

where $\beta > 0$. For any $\beta \geq 0$, this process induces an aperiodic and irreducible Markov process over the finite strategy set S with a unique stationary distribution $q = \{q^s\}_{s \in S} \in \Delta(S)$. When the game is a potential game, the stationary distribution satisfies

$$q^{s} = \frac{e^{\beta \cdot \phi(s)}}{\sum_{s' \in \mathcal{S}e^{\beta \cdot \phi(s')}}},$$
14.

which means that when $\beta \to \infty$, the support of the stationary distribution is contained in the strategy profiles with highest potential value.

5.3. Efficiency Guarantees

A key facet of any distributed control problem is a measure of the performance of the collective behavior, which we define by a nonnegative welfare function $W: \mathcal{S} \to \mathbb{R}$. Furthermore, we now define a game by the tuple $(\mathcal{P}, \mathcal{S}, u, W)$. Consider a game G where a pure Nash equilibrium is known to exist. We measure the efficiency of this Nash equilibrium according to two worst-case measures, termed the price of anarchy and the price of stability, defined as follows (88):

$$\operatorname{PoA}(G) = \min_{s^{\text{ne}} \in G} \left\{ \frac{W(s^{\text{ne}})}{\max_{s \in S} W(s)} \right\},$$

$$\operatorname{PoS}(G) = \max_{s^{\text{ne}} \in G} \left\{ \frac{W(s^{\text{ne}})}{\max_{s \in S} W(s)} \right\},$$

$$15.$$

$$PoS(G) = \max_{s \text{ ne} \in G} \left\{ \frac{W(s^{\text{ne}})}{\max_{s \in S} W(s)} \right\},$$
16.

where we use the notation $s^{ne} \in G$ to mean a pure Nash equilibrium in the game G. The price of anarchy seeks to bound the performance of any pure Nash equilibrium relative to the optimal strategy profile, while the price of stability focuses purely on the best Nash equilibrium. Note that $1 \ge PoS(G) \ge PoA(G) \ge 0$.

Now consider a family of games \mathcal{G} where a pure Nash equilibrium is known to exist for each game $G \in \mathcal{G}$. The price of anarchy and price of stability extend to this family of games \mathcal{G} in the following manner:

$$PoA(\mathcal{G}) = \min_{G \in \mathcal{G}} PoA(G),$$
17.

$$PoS(\mathcal{G}) = \min_{G \in \mathcal{G}} PoS(G).$$
 18.

In essence, the price of anarchy and price of stability provide worst-case performance guarantees when restricting attention to a specific type of equilibrium behavior.

5.4. Utility Design

We now turn to the question of how to design agent utility functions. To that end, suppose that a system operator has knowledge of the player set \mathcal{P} , an overestimate of the strategy sets $\bar{\mathcal{S}} = \prod_j \bar{\mathcal{S}}_j$, and a welfare function $W: \bar{\mathcal{S}} \to \mathbb{R}$. By overestimate, we mean that the true strategy sets satisfy $S_p \subseteq \bar{\mathcal{S}}_p$ for all i, but the system designer does not know this information a priori. Accordingly, the question that we seek to address is whether a system design can commit to a specific design of agent utility functions that ensures desirable properties irrespective of the chosen strategy sets \mathcal{S} . The following theorem from References 89–92 provides one such mechanism.

Theorem 7. Consider any player set \mathcal{P} , an overestimate of the strategy set $\bar{\mathcal{S}}$, and a welfare function $W: \bar{\mathcal{S}} \to \mathbb{R}$. Define an overestimate of the utility function where for any $s \in \bar{\mathcal{S}}$ and player $p \in \mathcal{P}$

$$u_p(s_p, s_{-p}) = W(s_p, s_{-p}) - W(s_p^{\text{BL}}, s_{-p}),$$
 19.

where $s_p^{\mathrm{BL}} \in \bar{\mathcal{S}}_p$ is a fixed baseline strategy. Then for any game $G = (\mathcal{P}, \mathcal{S}, u, W)$ where $\mathcal{S}_p \subseteq \bar{\mathcal{S}}_p$ for all p, the game G is a potential game with potential function W restricted to the domain \mathcal{S} .

Consider the family of games G induced by the utility design mechanism given in Equation 19 for any baseline strategy s^{BL} . Theorem 7 implies that the price of stability satisfies

$$PoS(\mathcal{G}) = 1$$

or, alternatively, that the optimal strategy profile is a Nash equilibrium for any game $G \in \mathcal{G}$, i.e., any realization with strategy sets $\mathcal{S}_p \subseteq \bar{\mathcal{S}}_p$. Not only is the optimal strategy profile a pure Nash equilibrium, but this optimal strategy profile is also the potential function maximizer. Consequently, by appealing to the algorithm log-linear learning defined above, one can ensure that the resulting collective behavior is nearly optimal when β is sufficiently high.

5.4.1. Optimizing the price of anarchy. Note that Theorem 7 makes no claim regarding the price of anarchy. In fact, what utility design mechanism optimizes the price of anarchy over the set of induced games \mathcal{G} remains an open question. Recent work has sought to address this question in several different domains, including concave cost-sharing problems (93, 94), submodular resource allocation problems (95, 96), network coding problems (97), set-covering problems (98), routing problems (64, 99), and more general cost-sharing problems with a restriction to utility design mechanisms that enforce $\sum_{b} u_{p}(s) = W(s)$ for all $s \in \mathcal{S}$ (100–102).

5.4.2. Local utility design. The utility design given in Theorem 7 also makes no reference to the structure or the informational dependence or locality of the derived utility functions. This question has been looked at extensively in the cost-sharing literature, where agents' utility functions are required to depend only on information regarding the selected resources; for example, in network routing problems, the utility of a player for selecting a route can depend only on the edges within that route and the other players that selected these edges. Confined to such local agent utility functions, Gopalakrishnan et al. (103) derived the complete set of local agent utility functions that ensure the existence of a pure Nash equilibrium; however, optimizing over such utility functions to optimize the price of anarchy is very much an open question, as highlighted above. Other recent work (104–106) has introduced a state in the game environment as a design parameter to design player objective functions of a desired locality.

6. CONCLUDING REMARKS

This review has presented an overview of selected topics in game theory of particular relevance to control systems, namely zero-sum games, team games, and game-theoretic distributed control. Of course, with limited space, there are bound to be some major omissions. Topics not discussed here include the following:

- General sum dynamic games: These are dynamic games that need not be zero-sum or team
 games. Of particular interest is the existence of Nash equilibria (for a summary, see 107) as
 well as reduced-complexity solution concepts (e.g., 108, 109) and learning (e.g., 110, 111).
- Large-population games: Here, a very large number of players can be approximated as a continuum. Sandholm (112) provides a general discussion, and Quijano et al. (113) provide a recent overview tailored to a control audience. Also related to large populations is the topic of mean-field games (114).
- Cooperative game theory: All of the models discussed above fall under the framework of noncooperative game theory. Cooperative game theory is a complementary formalism that relates to problems such as bargaining, matching, and coalition formation (e.g., for various engineering applications, see 115–117). Fele et al. (118) provide a recent overview tailored to a control audience.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by Office of Naval Research grant N00014-15-1-2762, National Science Foundation grant ECCS-1351866, and funding from King Abdullah University of Science and Technology (KAUST).

LITERATURE CITED

- 1. Myerson R. 1997. Game Theory. Cambridge, MA: Harvard Univ. Press
- 2. Zhou K, Doyle J. 1998. Essentials of Robust Control. Upper Saddle River, NJ: Prentice Hall
- 3. Marden JR, Shamma JS. 2015. Game theory and distributed control. In *Handbook of Game Theory with Economic Applications*, Vol. 4, ed. H Young, S Zamir, pp. 861–99. Amsterdam, Neth.: Elsevier

- Marden JR, Shamma JS. 2018. Game-theoretic learning in distributed control. In *Handbook of Dynamic Game Theory*, ed. T Başar, G Zaccour. Cham, Switz.: Springer. In press
- 5. Fudenberg D, Tirole J. 1991. Game Theory. Cambridge, MA: MIT Press
- 6. Osborne M, Rubinstein A. 1994. A Course in Game Theory. Cambridge, MA: MIT Press
- Cesa-Bianchi N, Lugosi G. 2006. Prediction, Learning, and Games. Cambridge, UK: Cambridge Univ. Press
- Leyton-Brown K, Shoham Y. 2008. Essentials of Game Theory: A Concise, Multidisciplinary Introduction. San Rafael, CA: Morgan & Claypool
- Nisan N, Roughgarden T, Tardos E, Vazirani VV. 2007. Algorithmic Game Theory. Cambridge, UK: Cambridge Univ. Press
- 10. Başar T, Olsder G. 1999. Dynamic Noncooperative Game Theory. Philadelphia, PA: Soc. Ind. Appl. Math.
- 11. Bauso D. 2016. Game Theory with Engineering Applications. Philadelphia, PA: Soc. Ind. Appl. Math.
- 12. Hespanha J. 2017. Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists. Princeton, NJ: Princeton Univ. Press
- 13. Roughgarden T. 2005. Selfish Routing and the Price of Anarchy. Cambridge, MA: MIT Press
- Camerer C. 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton, NJ: Princeton Univ. Press
- 15. Rubinstein A. 1998. Modeling Bounded Rationality. Cambridge, MA: MIT Press
- Ho Y, Bryson A, Baron S. 1965. Differential games and optimal pursuit-evasion strategies. IEEE Trans. Automat. Control 10:385–89
- 17. Isaacs R. 1965. Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization. New York: Wiley
- Liberzon D. 2012. Calculus of Variations and Optimal Control Theory: A Concise Introduction. Princeton, NJ: Princeton Univ. Press
- 19. Blanchini F, Miani S. 2007. Set-Theoretic Methods in Control. Basel, Switz.: Birkhäuser
- 20. Aubin J. 1991. Viability Theory. Basel, Switz.: Birkhäuser
- Mitchell IM, Bayen AM, Tomlin CJ. 2005. A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. IEEE Trans. Automat. Control 50:947–57
- Zhou Z, Zhang W, Ding J, Huang H, Stipanović DM, Tomlin CJ. 2016. Cooperative pursuit with Voronoi partitions. *Automatica* 72:64–72
- 23. Karaman S, Frazzoli E. 2011. Incremental sampling-based algorithms for a class of pursuit-evasion games. In Algorithmic Foundations of Robotics IX: Selected Contributions of the Ninth International Workshop on the Algorithmic Foundations of Robotics, ed. D Hsu, V Isler, JC Latombe, MC Lin, pp. 71–87. Berlin: Springer
- Isler V, Kannan S, Khanna S. 2005. Randomized pursuit-evasion in a polygonal environment. IEEE Trans. Robot. 21:875–84
- Zames G. 1981. Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. IEEE Trans. Automat. Control 26:301–20
- Başar T, Bernhard P. 1991. H-Infinity-Optimal Control and Related Minimax Design Problems. Basel, Switz.: Birkhäuser
- 27. Doyle J, Francis B, Tannenbaum A. 2009. Feedback Control Theory. Mineola, NY: Dover
- Banavar RN, Speyer JL. 1991. A linear-quadratic game approach to estimation and smoothing. In 1991 American Control Conference (ACC), pp. 2818–22. New York: IEEE
- Limebeer DJN, Anderson BDO, Khargonekar PP, Green M. 1992. A game theoretic approach to H[∞] control for time-varying systems. SIAM J. Control Optim. 30:262–83
- 30. Engwerda J. 2005. LQ Dynamic Optimization and Differential Games. Hoboken, NJ: Wiley
- Bertsekas D, Rhodes I. 1971. On the minimax reachability of target sets and target tubes. Automatica 7:233–47
- Bertsekas D. 1972. Infinite time reachability of state-space regions by using feedback control. IEEE Trans. Automat. Control 17:604–13
- 33. Bertsekas D. 2005. Dynamic Programming and Optimal Control. Belmont, MA: Athena Sci.

- Dahleh M, Diaz-Bobillo I. 1995. Control of Uncertain Systems: A Linear Programming Approach. Upper Saddle River, NJ: Prentice Hall
- 35. Shamma JS. 1993. Nonlinear state feedback for ℓ₁ optimal control. Syst. Control Lett. 21:265–70
- 36. Shamma JS. 1996. Optimization of the ℓ_{∞} -induced norm under full state feedback. *IEEE Trans. Automat. Control* 41:533–44
- 37. Shamma JS. 2012. An overview of LPV systems. In *Control of Linear Parameter Varying Systems with Applications*, ed. J Mohammadpour, CW Scherer, pp. 3–26. Boston: Springer
- Shamma JS, Xiong D. 1999. Set-valued methods for linear parameter varying systems. Automatica 35:1081–89
- Aumann RJ, Maschler M, Stearns RE. 1995. Repeated Games with Incomplete Information. Cambridge, MA: MIT Press
- 40. Mertens J, Sorin S, Zamir S. 2015. Repeated Games. Cambridge, UK: Cambridge Univ. Press
- 41. Smith JC, Prince M, Geunes J. 2013. Modern network interdiction problems and algorithms. In *Handbook of Combinatorial Optimization*, ed. PM Pardalos, DZ Du, RL Graham, pp. 1949–87. New York: Springer
- Zheng J, Castañón DA. 2012. Stochastic dynamic network interdiction games. In 2012 American Control Conference (ACC), pp. 1838–44. New York: IEEE
- Zheng J, Castañón DA. 2012. Dynamic network interdiction games with imperfect information and deception. In 2012 51st Annual IEEE Conference on Decision and Control (CDC), pp. 7758–63. New York: IEEE
- 44. Alpcan T, Başar T. 2010. Network Security: A Decision and Game-Theoretic Approach. Cambridge, UK: Cambridge Univ. Press
- Li L, Shamma JS. 2015. Efficient computation of discounted asymmetric information zero-sum stochastic games. In 2015 54th Annual IEEE Conference on Decision and Control (CDC), pp. 4531–36. New York: IEEE
- Li L, Langbort C, Shamma J. 2017. Computing security strategies in finite horizon repeated bayesian games. In 2017 American Control Conference (ACC), pp. 3664

 –69. New York: IEEE
- 47. Ho YC. 1980. Team decision theory and information structures. Proc. IEEE 68:644–54
- 48. Lewis FL, Zhang H, Hengster-Movric K, Das A. 2013. Cooperative Control of Multi-Agent Systems: Optimal and Adaptive Design Approaches. New York: Springer
- 49. Shamma JS. 2008. Cooperative Control of Distributed Multi-Agent Systems. Hoboken, NJ: Wiley
- 50. van Schuppen J, Villa T. 2014. Coordination Control of Distributed Systems. Cham, Switz.: Springer
- Wang Y, Zhang F. 2017. Cooperative Control of Multi-Agent Systems: Theory and Applications. Hoboken, NJ: Wiley
- 52. Radner R. 1962. Team decision problems. Ann. Math. Stat. 33:857–81
- 53. Witsenhausen HS. 1968. A counterexample in stochastic optimum control. SIAM J. Control 6:131-47
- 54. Başar T. 2008. Variations on the theme of the Witsenhausen counterexample. In 2008 47th IEEE Conference on Decision and Control, pp. 1614–19. New York: IEEE
- Ho YC, Chu K. 1974. Information structure in dynamic multi-person control problems. Automatica 10:341–51
- Nayyar A, Mahajan A, Teneketzis D. 2013. Decentralized stochastic control with partial history sharing: a common information approach. *IEEE Trans. Autom. Control* 58:1644–58
- Nayyar A, Mahajan A, Teneketzis D. 2014. The common-information approach to decentralized stochastic control. In *Information and Control in Networks*, ed. G Como, B Bernhardsson, A Rantzer, pp. 123–56. Cham, Switz.: Springer
- Kurtaran BZ, Sivan R. 1974. Linear-quadratic-Gaussian control with one-step-delay sharing pattern. IEEE Trans. Autom. Control 19:571–74
- Sandell N, Athans M. 1974. Solution of some nonclassical LQG stochastic decision problems. IEEE Trans. Autom. Control 19:108–16
- Nayyar N, Kalathil D, Jain R. 2018. Optimal decentralized control with asymmetric one-step delayed information sharing. *IEEE Trans. Control Netw. Syst.* 5:653–63
- Bamieh B, Voulgaris PG. 2005. A convex characterization of distributed control problems in spatially invariant systems with communication constraints. Syst. Control Lett. 54:575–83

- Rotkowitz M, Lall S. 2006. A characterization of convex problems in decentralized control. IEEE Trans. Autom. Control 51:274–86
- 63. Roughgarden T. 2005. Selfish Routing and the Price of Anarchy. Cambridge, MA: MIT Press
- Brown PN, Marden JR. 2017. Studies on robust social influence mechanisms, incentives for efficient network routing in uncertain settings. IEEE Control Syst. Mag. 37:98–115
- Lasaulce S, Jimenez T, Solan E, eds. 2017. Network Games, Control, and Optimization: Proceedings of NETGCOOP 2016, Avignon, France. Basel, Switz.: Birkhaüser
- Ozdaglar A, Menache I. 2011. Network Games: Theory, Models, and Dynamics. San Rafael, CA: Morgan & Claypool
- Marden JR, Ruben S, Pao L. 2013. A model-free approach to wind farm control using game theoretic methods. IEEE Trans. Control Syst. Technol. 21:1207–14
- Johnson KE, Thomas N. 2009. Wind farm control: addressing the aerodynamic interaction among wind turbines. In 2009 American Control Conference, pp. 2104

 –9. New York: IEEE
- Steinbuch M, de Boer WW, Bosgra OH, Peters S, Ploeg J. 1998. Optimal control of wind power plants.
 Wind Eng. Ind. Aerodyn. 27:237–46
- Pao L, Johnson KE. 2009. A tutorial on the dynamics and control of wind turbines and wind farms. In 2009 American Control Conference, pp. 35–36. New York: IEEE
- Cortes J, Martinez S, Karatas T, Bullo F. 2004. Coverage control for mobile sensing networks. IEEE Trans. Robot. Autom. 20:243–55
- Marden JR, Arslan G, Shamma JS. 2009. Cooperative control and potential games. IEEE Trans. Syst. Man Cybernet. B 39:1393

 –407
- Yazicioglu AY, Egerstedt M, Shamma JS. 2016. Communication-free distributed coverage for networked systems. IEEE Trans. Control Netw. Syst. 4:499–510
- 74. Fudenberg D, Levine D. 1998. The Theory of Learning in Games. Cambridge, MA: MIT Press
- Hart S, Mas-Colell A. 2003. Uncoupled dynamics do not lead to Nash equilibrium. Am. Econ. Rev. 93:1830–36
- Marden JR, Young HP, Arslan G, Shamma JS. 2009. Payoff based dynamics for multi-player weakly acyclic games. SIAM 7. Control Optim. 48:373–96
- Babichenko Y. 2010. Completely uncoupled dynamics and Nash equilibria. Discuss. Pap., Cent. Study Ration., Hebrew Univ., Ierusalem
- 78. Hart S. 2005. Adaptive heuristics. Econometrica 73:1401-30
- 79. Young HP. 2005. Strategic Learning and Its Limits. Oxford, UK: Oxford Univ. Press
- Shamma JS, Arslan G. 2005. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Trans. Autom. Control* 50:312–27
- 81. Fox MJ, Shamma JS. 2013. Population games, stable games, and passivity. Games 4:561-83
- Frihauf P, Krstic M, Başar T. 2012. Nash equilibrium seeking in noncooperative games. IEEE Trans. Autom. Control 57:1192–207
- 83. Monderer D, Shapley L. 1996. Potential games. Games Econ. Behav. 14:124-43
- 84. Blume L. 1993. The statistical mechanics of strategic interaction. Games Econ. Behav. 5:387-424
- Blume L. 1997. Population games. In The Economy as an Evolving Complex System II, ed. B Arthur, S Durlauf, D Lane, pp. 425–60. Reading, MA: Addison-Wesley
- Marden JR, Shamma JS. 2012. Revisiting log-linear learning: asynchrony, completeness and a payoff-based implementation. *Games Econ. Behav.* 75:788–808
- 87. Alos-Ferrer C, Netzer N. 2010. The logit-response dynamics. Games Econ. Behav. 68:413-27
- 88. Koutsoupias E, Papadimitriou C. 1999. Worst-case equilibria. In STACS 99: 16th Annual Symposium on Theoretical Aspects of Computer Science, Trier, Germany, March 4–6, 1999: Proceedings, ed. C Meinel, S Tison, pp. 404–13. Berlin: Springer
- Wolpert D. 2004. Theory of collective intelligence. In Collectives and the Design of Complex Systems, ed. K Tumer, D Wolpert, pp. 43–106. New York: Springer
- Arslan G, Marden JR, Shamma JS. 2007. Autonomous vehicle-target assignment: a game theoretical formulation. ASME 7. Dyn. Syst. Meas. Control 129:584–96
- 91. Marden JR, Wierman A. 2013. Distributed welfare games. Oper. Res. 61:155-68

- Gopalakrishnan R, Marden JR, Wierman A. 2011. An architectural view of game theoretic control. ACM SIGMETRICS Perf. Eval. Rev. 38:31–36
- 93. Phillips M, Shalaby Y, Marden JR. 2016. The importance of budget in efficient utility design. In 2016 55th Annual IEEE Conference on Decision and Control (CDC), pp. 6117–22. New York: IEEE
- Marden J, Phillips M. 2016. Optimizing the price of anarchy in concave cost sharing games. In 2017 American Control Conference (ACC), pp. 5237–42. New York: IEEE
- Marden J, Roughgarden T. 2014. Generalized efficiency bounds in distributed resource allocation. IEEE Trans. Autom. Control 59:571–84
- Vetta A. 2002. Nash equilibria in competitive societies with applications to facility location, traffic routing, and auctions. In 43rd Annual IEEE Symposium on Foundations of Computer Science, pp. 416–25. New York: IEEE
- Marden JR, Effros M. 2012. The price of selfishness in network coding. IEEE Trans. Inform. Theory 58:2349–61
- Gairing M. 2009. Covering games: approximation through noncooperation. In *Internet and Network Economics: 5th International Workshop, WINE 2009, Rome, Italy, December 14–18, 2009: Proceedings*, ed. S Leonardi, pp. 184–95. Berlin: Springer
- 99. Brown PN, Marden JR. 2016. Optimal mechanisms for robust coordination in congestion games. In 2015 54th Annual IEEE Conference on Decision and Control (CDC), pp. 2283–88. New York: IEEE
- Chen HL, Roughgarden T, Valiant G. 2008. Designing networks with good equilibria. In Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '08), pp. 854–63. New York: IEEE
- Gkatzelis V, Kollias K, Roughgarden T. 2016. Optimal cost-sharing in general resource selection games. Oper. Res. 64:1230–38
- Dobzinski S, Mehta A, Roughgarden T, Sundararajan M. 2017. Is Shapley cost sharing optimal? Games Econ. Behav. In press
- Gopalakrishnan R, Marden JR, Wierman A. 2014. Potential games are necessary to ensure pure Nash equilibria in cost sharing games. Math. Oper. Res. 39:1252–96
- 104. Marden JR. 2012. State based potential games. Automatica 48:3075–88
- Li N, Marden JR. 2013. Designing games for distributed optimization. IEEE J. Sel. Top. Signal Process.
 7:230–42
- 106. Li N, Marden JR. 2010. Designing games to handle coupled constraints. In 2010 49th Annual IEEE Conference on Decision and Control (CDC), pp. 250–55. New York: IEEE
- 107. Solan E, Vieille N. 2015. Stochastic games. PNAS 112:13743-46
- 108. Dudebout N, Shamma JS. 2012. Empirical evidence equilibria in stochastic games. In 2012 51st Annual IEEE Conference on Decision and Control (CDC), pp. 5780–85. New York: IEEE
- Dudebout N, Shamma JS. 2014. Exogenous empirical-evidence equilibria in perfect-monitoring repeated games yield correlated equilibria. In 2014 53rd Annual IEEE Conference on Decision and Control (CDC), pp. 1167–72. New York: IEEE
- Arslan G, Yksel S. 2017. Decentralized Q-learning for stochastic teams and games. IEEE Trans. Autom. Control 62:1545–58
- Yu CK, van der Schaar M, Sayed AH. 2017. Distributed learning for stochastic generalized Nash equilibrium problems. IEEE Trans. Signal Process. 65:3893–908
- 112. Sandholm W. 2010. Population Games and Evolutionary Dynamics. Cambridge, MA: MIT Press
- 113. Quijano N, Ocampo-Martinez C, Barreiro-Gomez J, Obando G, Pantoja A, Mojica-Nava E. 2017. The role of population games and evolutionary dynamics in distributed control systems: the advantages of evolutionary game theory. *IEEE Control Syst.* 37:70–97
- 114. Caines PE. 2013. Mean field games. In Encyclopedia of Systems and Control, ed. J Baillieul, T Samad, pp. 1–6. London: Springer
- Hamza D, Shamma JS. 2017. BLMA: a blind matching algorithm with application to cognitive radio networks. IEEE 7. Sel. Areas Commun. 35:302–16

- Saad W, Han Z, Debbah M, Hjorungnes A, Basar T. 2009. Coalitional game theory for communication networks. IEEE Signal Process. Mag. 26:77–97
- Saad W, Han Z, Poor HV. 2011. Coalitional game theory for cooperative micro-grid distribution networks. In 2011 IEEE International Conference on Communications Workshops (ICC). New York: IEEE. https://doi.org/10.1109/iccw.2011.5963577
- Fele F, Maestre JM, Camacho EF. 2017. Coalitional control: cooperative game theory and control. IEEE Control Syst. 37:53–69



Annual Review of Control, Robotics, and Autonomous Systems

Volume 1, 2018

Contents

Toward Robotic Manipulation Matthew T. Mason	1
Autonomous Flight Sarah Tang and Vijay Kumar	29
Soft Micro- and Nanorobotics Chengzhi Hu, Salvador Pané, and Bradley J. Nelson	53
Distributed Optimization for Control Angelia Nedić and Ji Liu	77
Game Theory and Control Jason R. Marden and Jeff S. Shamma	105
The Bountiful Intersection of Differential Geometry, Mechanics, and Control Theory Andrew D. Lewis	135
Sampling-Based Methods for Motion Planning with Constraints Zachary Kingston, Mark Moll, and Lydia E. Kavraki	159
Planning and Decision-Making for Autonomous Vehicles Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus	187
Synthesis for Robots: Guarantees and Feedback for Robot Behavior Hadas Kress-Gazit, Morteza Lahijanian, and Vasumathi Raman	211
Invariant Kalman Filtering Axel Barrau and Silvère Bonnabel	237
Data-Driven Predictive Control for Autonomous Systems Ugo Rosolia, Xiaojing Zhang, and Francesco Borrelli	259
A Separation Principle for Control in the Age of Deep Learning Alessandro Achille and Stefano Soatto	287
Privacy in Control and Dynamical Systems Shuo Han and George J. Pappas	309

Hamilton–Jacobi Reachability: Some Recent Theoretical Advances and Applications in Unmanned Airspace Management	
Mo Chen and Claire J. Tomlin	333
Design of Materials and Mechanisms for Responsive Robots Elliot W. Hawkes and Mark R. Cutkosky	359
Haptics: The Present and Future of Artificial Touch Sensation Heather Culbertson, Samuel B. Schorr, and Allison M. Okamura	385
Programming Cells to Work for Us Yili Qian, Cameron McBride, and Domitilla Del Vecchio	411
Autonomy in Rehabilitation Robotics: An Intersection Brenna D. Argall	441
Medical Technologies and Challenges of Robot-Assisted Minimally Invasive Intervention and Diagnostics	
Nabil Simaan, Rashid M. Yasin, and Long Wang	465

Errata

An online log of corrections to *Annual Review of Control, Robotics, and Autonomous Systems* articles may be found at http://www.annualreviews.org/errata/control