

# Adaptive learning in two-player Stackelberg games with continuous action sets

Guosong Yang, Radha Poovendran, and João P. Hespanha

**Abstract**—We study a two-player Stackelberg game in which the follower’s best response depends on parameters that are unknown to the leader. An adaptive learning algorithm is designed to simultaneously estimate these unknown parameters and minimize the leader’s cost, based on adaptive control techniques and hysteresis-type switching. The algorithm guarantees that the leader’s estimated cost will be indistinguishable from its actual cost in finite time, up to an error no larger than a preselected, arbitrarily small threshold, and that the first-order necessary condition for optimality holds asymptotically for the estimates cost. Provided that an additional persistent excitation condition holds, the difference between the parameter estimate and the actual parameter can also be bounded by a preselected, arbitrarily small threshold in finite time. The algorithm and the convergence results are illustrated via a simple simulation example in the domain of network security.

## I. INTRODUCTION

Modern engineering systems often involve multiple self-interested parties whose actions have mutual consequences. Examples include communications over a shared network with limited capacity, and computer programs sharing a limited computational resource. Game theory provides a systematic framework for modeling cooperation and conflict between these so-called strategic players [1], and has been widely used in areas such as robust design, resource allocation, and network security [2]–[4].

A fundamental question in game theory is whether the players can converge to a Nash equilibrium—a tuple of strategies for which no one has a unilateral incentive to change—if they play the game repeatedly and adjust their strategies based on historical outcomes. A primary example of such a learning process is fictitious play [5], in which each player believes that its opponents are playing constant mixed strategies in agreement with the empirical distributions of their past actions, and plays the corresponding best response. Fictitious play has attracted significant research interests [6], [7] and has been used for prescribing behavior in distributed control [8]. Another well-known example is the gradient response method developed in [9].

In this paper, we propose an adaptive learning approach for a hierarchical game model proposed by Stackelberg [10]. In a two-player Stackelberg game, one player (called the *leader*) selects its action first, and then the other player (called the *follower*), informed of the leader’s choice, selects its

own action. Therefore, a follower’s strategy in a Stackelberg game is a function that specifies a response to each leader’s possible action.

Stackelberg games provide a natural framework for understanding systems with asymmetrical information structure, which is common in network routing [11] and scheduling [12]. This feature is especially useful for modeling security problems, where the defender (leader) is unaware of the attack objective ahead of time, whereas the attacker (follower) is able to observe the defense strategy and attack after careful planning. Stackelberg Security Games have been applied to various real-world security domains, and have lead to practical implementations such as the ARMOR program at the Los Angeles International Airport and the IRIS program used by the US Federal Air Marshals [13].

Asymmetrical information often leads to scenarios where a Nash equilibrium does not exist but a Stackelberg equilibrium does, as the sufficient conditions for existence of the former is much stronger than those of the latter [2, p. 181]. For these scenarios, learning options like fictitious play and gradient response cannot be applied, and novel approaches are needed to achieve convergence to a Stackelberg equilibrium. Existing results on learning in Stackelberg games are limited to linear and quadratic costs and finite action sets [14]–[16], which are too restrictive for many applications including network security.

We study a Stackelberg game between two players, each with a continuous action set. We consider the scenario where the leader only has partial knowledge of the follower’s cost function. Specifically, the follower’s actual strategy (best-response function) admits a known functional form with unknown parameters. Our main contribution is an adaptive learning algorithm that simultaneously estimates these unknown parameters based on the follower’s past actions and minimizes the leader’s current cost, designed based on adaptive control techniques and hysteresis-type switching. The algorithm guarantees that the leader’s cost predicted using the parameter estimate will be indistinguishable from its actual cost in finite time, up to an error no larger than a preselected, arbitrarily small threshold, and that the first-order necessary condition for optimality holds asymptotically for the former. If an additional persistent excitation condition holds, then the difference between the parameter estimate and the actual parameter can also be bounded by a preselected, arbitrarily small threshold in finite time. Furthermore, we consider the case where the function used by the leader to predict the follower’s response does not perfectly match the follower’s actual strategy. In this case, our adaptive learning

This work was supported by the ONR MURI grant N00014-16-1-2710. G. Yang and J. P. Hespanha are with the Center for Control, Dynamical Systems, and Computation, University of California, Santa Barbara, CA 93106, USA. Email: {guosongyang, hespanha}@ucsb.edu.

R. Poovendran is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA. Email: rp3@uw.edu.

algorithm can be adjusted to guarantee the same convergence results with preselected error thresholds larger than the size of the mismatch. The algorithm and convergence results are illustrated via a simple simulation example motivated by link-flooding denial-of-service (DoS) attacks such as the Crossfire attack [17].

*Notations:* Let  $\mathbb{R}_+ := [0, \infty)$  and  $\mathbb{N} := \{0, 1, \dots\}$ . Let  $I_n$  be the identity matrix in  $\mathbb{R}^{n \times n}$ ; the subscript is omitted when the dimension is implicit. Denote by  $\|\cdot\|$  the Euclidean norm for vectors and the (induced) Euclidean norm for matrices. For a set  $\mathcal{S} \subset \mathbb{R}^n$ , denote by  $\partial\mathcal{S}$ ,  $\overline{\mathcal{S}}$ , and  $\overline{\text{conv}}\mathcal{S}$  its boundary, closure, and closed convex hull, respectively. A signal  $u : [t_0, \infty) \rightarrow \mathbb{R}^n$  is of class  $\mathcal{L}_\infty$  if  $\sup_{t \geq t_0} \|u(t)\|$  is finite. A function is of class  $\mathcal{C}^1$  if it is continuously differentiable. Denote by  $\mathcal{B}_r(x)$  the closed ball of radius  $r > 0$  centered at  $x \in \mathbb{R}^n$ , that is,  $\mathcal{B}_r(x) := \{y \in \mathbb{R}^n : \|y - x\| \leq r\}$ .

## II. PROBLEM FORMULATION

Consider a two-player game where  $\mathcal{U} \subset \mathbb{R}^{n_u}$  and  $\mathcal{A} \subset \mathbb{R}^{n_a}$  are the *action sets* of the first and the second players, respectively, and  $J : \mathcal{U} \times \mathcal{A} \rightarrow \mathbb{R}$  and  $H : \mathcal{A} \times \mathcal{U} \rightarrow \mathbb{R}$  are the corresponding *cost functions*. We are interested in a hierarchical game model proposed by Stackelberg [10], where the first player (called the *leader*) selects its action  $u \in \mathcal{U}$  first, and then the second player (called the *follower*), informed of the leader's choice, selects its action  $a \in \mathcal{A}$ . The corresponding notion of equilibrium is defined as follows:

**Definition 1** ([1], [2, Def. 4.6]). An action  $u^* \in \mathcal{U}$  is a *Stackelberg equilibrium action* for the leader if

$$\sup_{a \in \beta_a(u^*)} J(u^*, a) = \inf_{u \in \mathcal{U}} \sup_{a \in \beta_a(u)} J(u, a),$$

where  $\beta_a(u) \subset \mathcal{A}$  denotes the set of best responses against  $u$ , that is,  $\beta_a(u) := \{a \in \mathcal{A} : H(a, u) = \inf_{a' \in \mathcal{A}} H(a', u)\}$ .

We consider games with perfect but incomplete information, where the leader only has partial knowledge of the follower's action set and cost function. Specifically, for a leader's choice  $u$ , the follower selects an action  $a = f(\theta^*, u) \in \beta_a(u)$ , where the function  $f : \Theta \times \mathcal{U} \rightarrow \mathbb{R}^{n_a}$  and the parameter set  $\Theta \subset \mathbb{R}^{n_\theta}$  are known to the leader, but the specific value of the parameter  $\theta^* \in \Theta$  is unknown.

In practice, assuming that the functional form of the follower's best-response is known introduces little loss of generality, as it can always be approximated on a compact set up to an arbitrary precision as a finite weighted sum of a preselected class of basis functions. An example of such an approximation is the *radial basis function (RBF)* model [18], in which the leader assumes

$$f(\theta^*, u) = \sum_{j=1}^{n_\theta} \theta_j^* F_j(u) = \sum_{j=1}^{n_\theta} \theta_j^* \phi(\|u - u_j^c\|),$$

where  $\theta^* = (\theta_1^*, \dots, \theta_{n_\theta}^*)$  is the unknown parameter, and each  $F_j : \mathcal{U} \rightarrow \mathbb{R}^{n_a}$  is an RBF centered at  $u_j^c$ . Note that in the RBF model, the approximation is affine with respect to the unknown parameter, which is a common feature in many widely-used approximation models such as *orthogonal*

*polynomials* and *multivariate splines* [18]. This motivates restricting our attention to affine maps  $\theta \mapsto f(\theta, u)$ . The following assumption captures this and additional regularity conditions that we use to guarantee existence of a Stackelberg equilibrium.

**Assumption 1** (Regularity). The leader's action set  $\mathcal{U}$  and the parameter set  $\Theta$  are convex and compact; the leader's cost function  $J$  is  $\mathcal{C}^1$  with locally Lipschitz Jacobian; the function  $f$  is  $\mathcal{C}^1$  with locally Lipschitz gradient; the map  $\theta \mapsto f(\theta, u)$  is affine for each fixed  $u \in \mathcal{U}$ .

Under Assumption 1, existence of a Stackelberg equilibrium follows from standard results, e.g., [2, Th. 4.8]. In particular, these conditions are much weaker than the sufficient conditions required for existence of a Nash equilibrium [2, p. 181], which is consistent with our interest in games where a Nash equilibrium does not exist but a Stackelberg equilibrium does; see also the example in Section VI.

We denote by  $\nabla_u J(u, a)$  and  $\nabla_a J(u, a)$  the gradients of the maps  $u \mapsto J(u, a)$  and  $a \mapsto J(u, a)$ , respectively, and by  $\nabla_\theta f(u)$  and  $\nabla_u f(\theta, u)$  the Jacobian matrices of the maps  $\theta \mapsto f(\theta, u)$  and  $u \mapsto f(\theta, u)$ , respectively. (To be consistent with the definition of Jacobian matrix, we take gradients as row vectors.) In particular, the Jacobian matrix  $\nabla_\theta f(u)$  is independent of  $\theta$  following the affine condition in Assumption 1.

Our goal is to adjust the leader's action  $u$  to minimize its cost:

$$\begin{aligned} & \text{minimize} && J(u, f(\theta^*, u)), \\ & \text{subject to} && u \in \mathcal{U}, \end{aligned}$$

based on past observations of the follower's action  $a = f(\theta^*, u)$  and the leader's cost  $J(u, a)$ , but without knowing the actual value of the unknown parameter  $\theta^*$ . Our approach to solve this problem consists of two components:

- 1) Construct a *parameter estimate*  $\hat{\theta}$  that approaches the actual parameter  $\theta^*$ .
- 2) Adjust the leader's action  $u$  based on a gradient descent method to minimize its *estimated cost*:

$$\begin{aligned} & \text{minimize} && \hat{J}(u, \hat{\theta}) := J(u, f(\hat{\theta}, u)), \\ & \text{subject to} && u \in \mathcal{U}. \end{aligned} \tag{1}$$

In this paper, our analysis and design are formulated using continuous-time dynamics, which is common in the literature of learning in game theory [6], [7].

## III. ESTIMATION AND MINIMIZATION

To specify the adaptive algorithm for estimating the unknown parameter  $\theta^*$  that is known to lie in the compact convex set  $\Theta$  and optimizing the leader's action  $u$  within the compact convex set  $\mathcal{U}$ , we recall the following notions and basic properties from convex analysis [19, Sec. 4.7].

For a convex set  $\mathcal{S} \subset \mathbb{R}^n$ , we denote by  $N_{\mathcal{S}}(x)$  the *normal cone* at a point  $x \in \mathcal{S}$ , that is,

$$N_{\mathcal{S}}(x) := \{v \in \mathbb{R}^n : \forall y \in \mathcal{S}, v^\top (y - x) \leq 0\}, \tag{2}$$

and by  $T_{\mathcal{S}}(x)$  the *tangent cone* at  $x$ , that is,

$$T_{\mathcal{S}}(x) := \{w \in \mathbb{R}^n : \forall v \in N_{\mathcal{S}}(x), v^\top w \leq 0\}. \tag{3}$$

In particular,  $T_{\mathcal{S}}(x) = \mathbb{R}^n$  and  $N_{\mathcal{S}}(x) = \{0\}$  for all interior points  $x \in \mathcal{S} \setminus \partial\mathcal{S}$ . For a function  $g : \mathcal{S} \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ , denote by  $[g(x, d)]_{T_{\mathcal{S}}}$  its *tangent projection* with respect to  $\mathcal{S}$  at  $(x, d) \in \mathcal{S} \times \mathbb{R}^m$ , that is,

$$[g(x, d)]_{T_{\mathcal{S}}} := \arg \min_{v \in T_{\mathcal{S}}(x)} \|v - g(x, d)\|.$$

The convexity of  $\mathcal{S}$  ensures that  $[g(x, d)]_{T_{\mathcal{S}}}$  is unique for each  $(x, d) \in \mathcal{S} \times \mathbb{R}^m$ . Moreover, for all  $x \in \mathcal{S}$  and  $d \in \mathbb{R}^m$ , the tangent projection satisfies

$$[g(x, d)]_{T_{\mathcal{S}}} \in T_{\mathcal{S}}(x), \quad g(x, d) - [g(x, d)]_{T_{\mathcal{S}}} \in N_{\mathcal{S}}(x), \quad (4)$$

and

$$(g(x, d) - [g(x, d)]_{T_{\mathcal{S}}})^\top [g(x, d)]_{T_{\mathcal{S}}} = 0. \quad (5)$$

#### A. Parameter estimation

We construct the parameter estimate  $\theta$  by comparing past observations of the follower's action  $a = f(\theta^*, u)$  and the leader's cost  $J(u, a)$  with the corresponding estimated values  $f(\theta, u)$  and  $\hat{J}(u, \theta) = J(u, f(\theta, u))$ . Their difference is defined as the following *observation error*

$$e_{\text{obs}} := \begin{bmatrix} f(\theta, u) - a \\ \hat{J}(u, \theta) - J(u, a) \end{bmatrix}.$$

In the following, we develop an estimation law in terms of the observation error  $e_{\text{obs}}$ , so that the norm of the *estimation error*  $\|\theta - \theta^*\|$  is monotonically nonincreasing, regardless of how the leader's action  $u$  is adjusted.

First, we establish a relation between the observation error  $e_{\text{obs}}$  and the estimation error  $\theta - \theta^*$ ; the proof can be found in [20].

**Lemma 1.** *The observation error  $e_{\text{obs}}$  satisfies*

$$e_{\text{obs}} = K(u, a, \theta)(\theta - \theta^*) \quad (6)$$

with the gain matrix

$$K(u, a, \theta) := \begin{bmatrix} I \\ \int_0^1 \nabla_a J(u, \lambda f(\theta, u) + (1 - \lambda)a) d\lambda \end{bmatrix} \nabla_\theta f(u). \quad (7)$$

Following Lemma 1, the observation error  $e_{\text{obs}}$  would be zero if the current estimate  $\theta$  of the unknown parameter  $\theta^*$  was correct. However, in most interesting scenarios, the dimension  $n_\theta$  of the parameter vector  $\theta^*$  is much larger than the dimension  $n_a + 1$  of the observation error  $e_{\text{obs}}$ ; thus the gain matrix  $K(u, a, \theta)$  cannot be invertible, and a zero value for the observation error  $e_{\text{obs}}$  does not imply that the estimate  $\theta$  is correct.

We propose the following estimation law to drive the parameter estimate  $\theta$  towards the actual parameter  $\theta^*$ :

$$\dot{\theta} = [-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}]_{T_\Theta} \quad (8)$$

with the gain matrix  $K(u, a, \theta)$  defined by (7) and the *switching signal*  $\lambda_e : \mathbb{R}_+ \rightarrow \{0, \lambda_1\}$  defined by

$$\lambda_e(t) := \begin{cases} \lambda_1 & \text{if } \|e_{\text{obs}}(t)\| \geq \varepsilon_{\text{obs}}; \\ \lim_{s \nearrow t} \lambda_e(s) & \text{if } \|e_{\text{obs}}(t)\| \in (\varepsilon_{\text{obs}}/2, \varepsilon_{\text{obs}}); \\ 0 & \text{if } \|e_{\text{obs}}(t)\| \leq \varepsilon_{\text{obs}}/2 \end{cases} \quad (9)$$

and  $\lambda_e(0) := \lambda_1$  if  $\|e_{\text{obs}}(0)\| \in (\varepsilon_{\text{obs}}/2, \varepsilon_{\text{obs}})$ , where  $\varepsilon_{\text{obs}}, \lambda_1 > 0$  are preselected constants. Several comments are in order: First, the gain matrix  $K(u, a, \theta)$  depends on the estimate  $\theta$  but not on the actual parameter  $\theta^*$ , so (8) can be implemented without knowing  $\theta^*$ . Second, the tangent projection  $[\cdot]_{T_\Theta}$  is used to guarantee that the parameter estimate  $\theta$  remains inside the compact convex set  $\Theta$ . Finally, the right-continuous, piecewise constant switching signal  $\lambda_e$  is designed so that the adaption is on when  $\|e_{\text{obs}}\| \geq \varepsilon_{\text{obs}}$  and off when  $\|e_{\text{obs}}\| \leq \varepsilon_{\text{obs}}/2$ , with a hysteresis-type switching rule that avoids chattering. The key feature of (8) is that the estimation error  $\theta - \theta^*$  satisfies

$$\begin{aligned} \frac{d\|\theta - \theta^*\|^2}{dt} &= 2(\theta - \theta^*)^\top [-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}]_{T_\Theta} \\ &\leq 2(\theta - \theta^*)^\top (-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}) = -2\lambda_e \|e_{\text{obs}}\|^2, \end{aligned}$$

where the inequality follows from (4) and (2). We thus conclude that the estimation law (8) with the switching signal (9) guarantees

$$\frac{d\|\theta - \theta^*\|^2}{dt} \leq -2\lambda_e \|e_{\text{obs}}\|^2 \leq 0, \quad (10)$$

which implies that  $\|\theta - \theta^*\|$  is monotonically nonincreasing and will not stop approaching zero unless  $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ . In the convergence results in Section IV, we will show that the adaption of the parameter estimate  $\theta$  stops in finite time, and the observation error  $e_{\text{obs}}$  satisfies  $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$  afterwards.

#### B. Cost minimization

Several options are available to adjust the leader's action  $u$ , but in this paper our analysis will focus on a gradient descent method, which is fairly robust for a wide range of problems. Our ultimate goal is to minimize the leader's cost  $J(u, a) = J(u, f(\theta^*, u))$ . However, computing the gradient descent direction of the actual cost requires knowledge of the actual parameter  $\theta^*$ . Therefore, we minimize instead the leader's estimated cost  $\hat{J}(u, \theta) = J(u, f(\theta, u))$ , which only depends on the estimate  $\theta$ . This change in objective is justified by the fact that  $\|\hat{J}(u, \theta) - J(u, a)\| \leq \|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$  holds in finite time, which will be established in Section IV.

The time derivative of the estimated cost  $\hat{J}(u, \theta)$  is

$$\dot{\hat{J}}(u, \theta) = \nabla_u \hat{J}(u, \theta) \dot{u} + \nabla_\theta \hat{J}(u, \theta) \dot{\theta}, \quad (11)$$

where

$$\begin{aligned} \nabla_u \hat{J}(u, \theta) &= \nabla_u J(u, f(\theta, u)) + \nabla_a J(u, f(\theta, u)) \nabla_u f(\theta, u), \\ \nabla_\theta \hat{J}(u, \theta) &= \nabla_a J(u, f(\theta, u)) \nabla_\theta f(u) \end{aligned}$$

are the gradients of the maps  $u \mapsto \hat{J}(u, \theta)$  and  $\theta \mapsto \hat{J}(u, \theta)$ , respectively. (Note that here  $\nabla_u J(u, f(\theta, u))$  denotes the gradient of the map  $u \mapsto J(u, \hat{a})$  at  $\hat{a} = f(\theta, u)$ .) As we will establish that the adaption of the parameter estimate  $\theta$  stops in finite time, we neglect the second term in (11) and focus exclusively in adjusting  $u$  along the gradient descent direction of  $u \mapsto \hat{J}(u, \theta)$ . This motivates the following minimization law to adjust the leader's action:

$$\dot{u} = [-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top]_{T_{U_a}} \quad (12)$$

where  $\lambda_2 > 0$  is a preselected constant. The tangent projection  $[\cdot]_{T_{\mathcal{U}}}$  is used to guarantee that the leader's action  $u$  remains inside the compact convex set  $\mathcal{U}$ . Consequently,

$$\begin{aligned}\dot{J}(u, \theta) &= \nabla_u \hat{J}(u, \theta) [-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top]_{T_{\mathcal{U}}} + \nabla_\theta \hat{J}(u, \theta) \dot{\theta} \\ &= -\| [-\lambda_2 \nabla_u \hat{J}(u, \theta)^\top]_{T_{\mathcal{U}}} \|^2 / \lambda_2 + \nabla_\theta \hat{J}(u, \theta) \dot{\theta} \\ &= -\|\dot{u}\|^2 / \lambda_2 + \nabla_\theta \hat{J}(u, \theta) \dot{\theta},\end{aligned}$$

where the second equality follows from (5). We thus conclude that the minimization law (12) guarantees

$$\dot{\theta} = 0 \implies \dot{J}(u, \theta) \leq -\|\dot{u}\|^2 / \lambda_2 \leq 0.$$

In the convergence results in Section IV, we will show that the leader's action  $u$  converges asymptotically to the set of points for which the first-order necessary condition for optimality holds for the optimization problem (1).

#### IV. CONVERGENCE ANALYSIS

We now state the main result of this paper:

**Theorem 1.** *Suppose that Assumption 1 holds. Given any threshold  $\varepsilon_{\text{obs}} > 0$  in (9), the estimation and minimization algorithm (8) and (12) with the switching signal (9) guarantees the following properties:*

1) *There exists a time  $T_1 \geq 0$  such that*

$$\theta(t) = \theta(T_1) \quad \forall t \geq T_1, \quad (13)$$

and

$$\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}} \quad \forall t \geq T_1. \quad (14)$$

2) *The first-order necessary condition for optimality holds asymptotically for the optimization problem (1), that is,*

$$\lim_{t \rightarrow \infty} [-\nabla_u \hat{J}(u(t), \theta(T_1))^\top]_{T_{\mathcal{U}}} = 0. \quad (15)$$

Essentially, item 1) guarantees that the parameter estimate  $\theta$  converges in finite time to a point which is indistinguishable from the actual parameter  $\theta^*$  based on observations of the follower's action  $a = f(\theta^*, u)$  and the leader's cost  $J(u, a)$ , up to an error no larger than the threshold  $\varepsilon_{\text{obs}}$ . Regarding item 2), we recall the first-order necessary condition for optimality from [19, Prop. 4.7.1]:

**Lemma 2.** *If  $u^*$  is a local minimum point of the map  $u \mapsto \hat{J}(u, \theta(T_1))$  on  $\mathcal{U}$ , then there exists a  $\delta > 0$  such that for all  $u \in \mathcal{U}$  satisfying  $\|u - u^*\| < \delta$ , it holds that*

$$\nabla_u \hat{J}(u^*, \theta(T_1))(u - u^*) \geq 0,$$

or equivalently,

$$[-\nabla_u \hat{J}(u^*, \theta(T_1))^\top]_{T_{\mathcal{U}}} = 0.$$

*Proof of Theorem 1.* As the right hand-sides of (8) and (12) are potentially discontinuous due to the tangent projections, the proof of Theorem 1 requires results from differential inclusions theory; see the Appendix for the necessary preliminaries.

First, we establish existence, boundedness, and uniqueness of solutions for the system defined by (8) and (12).

**Lemma 3.** *For each  $(\theta_0, u_0) \in \Theta \times \mathcal{U}$ , there exists a unique Carathéodory solution to the system defined by (8) and (12) on  $\mathbb{R}_+$  with  $(\theta(0), u(0)) = (\theta_0, u_0)$ , that is, there exist unique absolutely continuous functions  $\theta : \mathbb{R}_+ \rightarrow \mathbb{R}^{n_\theta}$  and  $u : \mathbb{R}_+ \rightarrow \mathbb{R}^{n_u}$  with  $(\theta(0), u(0)) = (\theta_0, u_0)$  such that (8) and (12) hold almost everywhere on  $\mathbb{R}_+$ . Moreover,  $(\theta(t), u(t)) \in \Theta \times \mathcal{U}$  for all  $t \geq 0$ , and*

$$\theta, \dot{\theta}, u, \dot{u}, e_{\text{obs}}, \dot{e}_{\text{obs}} \in \mathcal{L}_\infty. \quad (16)$$

*Proof.* Lemma 3 is established by modeling the system defined by (8) and (12) using the project dynamical system (27) in the Appendix with the state  $x := (\theta, u)$  and the set  $\mathcal{S} := \Theta \times \mathcal{U}$ , and applying the results on hysteresis switching from [21] and the results on existence, boundedness, and uniqueness of solutions for projected dynamical systems from Proposition 5 in the Appendix; the complete proof can be found in [20]  $\square$

Second, we establish item 1) of Theorem 1 via arguments along the lines of the proof of Barbalat's lemma [22, Lemma 3.2.6]. We cannot use Barbalat's lemma directly since the switching signal  $\lambda_e$  in (8) is not continuous but only piecewise continuous. Following (10), we see that  $\|\theta - \theta^*\|^2$  is monotonically nonincreasing. Therefore  $\lim_{t \rightarrow \infty} \|\theta(t) - \theta^*\|^2$ , and thus

$$\lim_{t \rightarrow \infty} \int_0^t \lambda_e(s) \|e_{\text{obs}}(s)\|^2 ds, \quad (17)$$

exists and is finite. On the other hand, (8) and (9) imply that (13) and (14) hold if there exists a time  $T_1 \geq 0$  such that

$$\lambda_e(t) = 0 \quad \forall t \geq T_1. \quad (18)$$

Assume (18) does not hold for any  $T_1 \geq 0$ . Then (9) implies that there exists an unbounded increasing sequence  $(t_k)_{k \in \mathbb{N}}$  with  $t_0 > 0$  such that  $\lambda_e(t_k) = \lambda_1$  and  $\|e_{\text{obs}}(t_k)\| > \varepsilon_{\text{obs}}/2$  for all  $k \in \mathbb{N}$ . Now we show that there exists an unbounded sequence  $(s_k)_{k \in \mathbb{N}}$  with  $s_k \in [t_k - \delta, t_k]$  such that

$$\|e_{\text{obs}}(t)\| \geq \varepsilon_{\text{obs}}/2, \quad \lambda_e(t) = \lambda_1 \quad (19)$$

for all  $k \in \mathbb{N}$  and  $t \in [s_k, s_k + \delta]$  with the constant

$$\delta := \min \left\{ t_0, \frac{\varepsilon_{\text{obs}}}{2 \sup_{s \geq 0} \|\dot{e}_{\text{obs}}(s)\|} \right\} > 0,$$

where  $\delta > 0$  following  $\dot{e}_{\text{obs}} \in \mathcal{L}_\infty$  in (16),  $\varepsilon_{\text{obs}} > 0$ , and  $t_0 > 0$ . Indeed, for each  $k \in \mathbb{N}$ , consider the following two possibilities:

- 1) If  $\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}}$  for all  $t \in [t_k - \delta, t_k]$ , then (9) and  $\lambda_e(t_k) = \lambda_1$  imply that (19) holds with  $s_k = t_k - \delta$ .
- 2) Otherwise, there exists an  $s_k \in [t_k - \delta, t_k]$  such that  $\|e_{\text{obs}}(s_k)\| = \varepsilon_{\text{obs}}$ , and (19) follows from the definition of  $\delta$  and (9).

Moreover,  $(s_k)_{k \in \mathbb{N}}$  is unbounded as  $(t_k)_{k \in \mathbb{N}}$  is unbounded. Following (19), we see that

$$\int_{s_k}^{s_k + \delta} \lambda_e(s) \|e_{\text{obs}}(s)\|^2 ds \geq \frac{\lambda_1 \varepsilon_{\text{obs}}^2 \delta}{4} > 0$$

for the unbounded sequence  $(s_k)_{k \in \mathbb{N}}$ , which contradicts the property that (17) exists and is finite. Therefore, there exists a time  $T_1 \geq 0$  such that (18), and thus (13) and (14), holds.

Finally, we establish item 2) of Theorem 1 based on an invariance principle for Filippov solutions [23, Th. 3.2]. After  $T_1$ , the system (12) becomes

$$\dot{u} = [-\lambda_2 \nabla_u \hat{J}(u, \theta(T_1))^\top]_{T_{\mathcal{U}}},$$

which can be modeled using the projected dynamical system (27) in the Appendix with the state  $x := u$  and the set  $\mathcal{S} := \mathcal{U}$ . The corresponding function  $g$  in (27) is defined by  $g(x) := -\lambda_2 \nabla_u \hat{J}(x, \theta(T_1))^\top$ . Note that every Filippov solution to (27) is a Carathéodory solution to the differential inclusion (29). Then Proposition 5 in the Appendix implies that for each  $u_0 \in \mathcal{U}$ , there exists a unique Filippov solution  $u$  to (27) on  $[T_1, \infty)$  with  $u(T_1) = u_0$  which is also a Carathéodory solution, and  $u(t) \in \mathcal{U}$  for all  $t \geq T_1$ . Consider the function  $V : \mathcal{U} \rightarrow \mathbb{R}$  defined by

$$V(u) := \hat{J}(u, \theta(T_1)).$$

Then for all  $u \in \mathcal{U}$  and  $v \in G(u)$  with the corresponding set-valued function  $G$  defined by (28), it holds that

$$\begin{aligned} \nabla V(u) v &= \nabla_u \hat{J}(u, \theta(T_1)) v \\ &= -g(u)^\top v / \lambda_2 \leq -\|g(u)\|_{T_{\mathcal{U}}}^2 / \lambda_2 \leq 0, \end{aligned}$$

where the first inequality follows from (32). Consequently, the invariance principle [23, Th. 3.2] implies that all Filippov solutions, and therefore all Carathéodory solutions, to (27) converge to the largest invariant set in  $\{u \in \mathcal{U} : \exists v \in G(u) \text{ s.t. } \nabla V(u) v = 0\} \subset \{u \in \mathcal{U} : [g(u)]_{T_{\mathcal{U}}} = 0\}$ . Therefore,  $\lim_{t \rightarrow \infty} [g(u(t))]_{T_{\mathcal{U}}} = 0$ , that is, (15) holds.  $\square$

In Theorem 1, there is no claim that the parameter estimate  $\theta$  necessarily converges to the actual parameter  $\theta^*$ . However, this can be guaranteed if we assume that the following *persistent excitation (PE)* condition holds.

**Assumption 2** (Persistent excitation). There exist constants  $\tau_0, \alpha_0 > 0$  such that the gain matrix  $K(u, a, \theta)$  defined by (7) satisfies

$$\int_t^{t+\tau_0} K(u(s), a(s), \theta(s))^\top K(u(s), a(s), \theta(s)) ds \geq \alpha_0 I \quad (20)$$

for all  $t \geq 0$ .

**Theorem 2.** Suppose that Assumptions 1 and 2 hold. Then by setting the threshold

$$\varepsilon_{\text{obs}} := \varepsilon_\theta \sqrt{\alpha_0 / \tau_0} \quad (21)$$

in (9) for any given constant  $\varepsilon_\theta > 0$ , the estimation and minimization algorithm (8) and (12) with the switching signal (9) guarantees the following properties:

1) There exists a time  $T_1 \geq 0$  such that (13) and (14) hold, and

$$\|\theta(T_1) - \theta^*\| < \varepsilon_\theta. \quad (22)$$

2) The first-order necessary condition for optimality holds asymptotically for the optimization problem (1), that is, (15) holds.

*Proof.* As (13), (14) and (15) are established in Theorems 1, it remains to prove (22).

To this effect, we note that (14) implies

$$\int_{T_1}^{T_1+\tau_0} \|e_{\text{obs}}(s)\|^2 ds < \varepsilon_{\text{obs}}^2 \tau_0 = \alpha_0 \varepsilon_\theta^2,$$

where the equality follows from (21). On the other hand, (6) and (13) imply

$$\begin{aligned} &\int_{T_1}^{T_1+\tau_0} \|e_{\text{obs}}(t)\|^2 dt \\ &= \int_{T_1}^{T_1+\tau_0} \|K(u(t), a(t), \theta(T_1))(\theta(T_1) - \theta^*)\|^2 ds \\ &\geq \alpha_0 \|\theta(T_1) - \theta^*\|^2, \end{aligned}$$

where the inequality follows from the PE condition (20). Combining the inequalities above yields (22).  $\square$

*Remark 1.* In view of (7), a sufficient condition for (20) is

$$\int_t^{t+\tau_0} \nabla_\theta f(u(s))^\top \nabla_\theta f(u(s)) ds \geq \alpha_0 I \quad \forall t \geq 0. \quad (23)$$

The PE condition (23) is more restrictive than (20), but has the advantage that it can be checked without knowing the estimate  $\theta$ . Moreover, from the proof of Theorem 2, we see that (22) only requires (20) or (23) to hold at  $t = T_1$ . Consequently, in practice it suffices to enforce that (20) or (23) holds when  $\lambda_e$  in (9) has been set to zero.

## V. MODEL MISMATCH

In this section, we study the case where a bounded mismatch exists between the follower's actual best response  $a$  against a leader's action  $u$  and the leader's belief  $f(\theta^*, u)$ . Denoting by  $f^* : \mathcal{U} \rightarrow \mathcal{A}$  the unknown follower's strategy (best-response function), that is,  $a = f^*(u) \in \beta_a(u)$ , up to now we assumed that there was some unknown parameter  $\theta^* \in \Theta$  such that

$$f^*(u) = f(\theta^*, u) \quad \forall u \in \mathcal{U}$$

for the known function  $f$ . We now consider the scenario where such perfect matching between the follower's actual best response  $f^*(u)$  and some  $f(\theta^*, u)$  may not exist.

**Assumption 3** (Mismatch). The unknown follower's strategy  $f^*$  is locally Lipschitz and satisfies

$$\begin{aligned} &\left\| \int_0^1 \nabla_a J(u, \lambda f(\theta, u) + (1-\lambda)f^*(u)) d\lambda \right. \\ &\quad \times (f(\theta^*, u) - f^*(u)) \Big\| \leq \varepsilon_f \quad \forall \theta \in \Theta, \forall u \in \mathcal{U} \end{aligned}$$

for the known function  $f$  and some known constant  $\varepsilon_f > 0$ .

Similar arguments to those in the proof of Lemma 1 show that the observation error  $e_{\text{obs}}$  now satisfies

$$e_{\text{obs}} = K(u, a, \theta)(\theta - \theta^*) + e_f \quad (24)$$

with the gain matrix  $K(u, a, \theta)$  defined by (7) and the mismatch error

$$e_f := \left[ \int_0^1 \nabla_a J(u, \lambda f(\theta, u) + (1 - \lambda)a) d\lambda \right] (f(\theta^*, u) - a).$$

Then Assumption 3 implies

$$\|e_f(t)\| \leq \varepsilon_f \quad \forall t \geq 0. \quad (25)$$

The effect of the mismatch error  $e_f$  in (24) can be mitigated by keeping the same estimation law (8) while adjusting the definition of the switching signal  $\lambda_e$  by

$$\lambda_e(t) := \begin{cases} \lambda_1 & \text{if } \|e_{\text{obs}}(t)\| \geq \varepsilon_{\text{obs}}; \\ \lim_{s \nearrow t} \lambda_e(s) & \text{if } \|e_{\text{obs}}(t)\| \in ((\varepsilon_{\text{obs}} + \varepsilon_f)/2, \varepsilon_{\text{obs}}); \\ 0 & \text{if } \|e_{\text{obs}}(t)\| \leq (\varepsilon_{\text{obs}} + \varepsilon_f)/2 \end{cases} \quad (26)$$

and  $\lambda_e(0) = \lambda_1$  if  $\|e_{\text{obs}}(0)\| \in ((\varepsilon_{\text{obs}} + \varepsilon_f)/2, \varepsilon_{\text{obs}})$ , where  $\varepsilon_{\text{obs}} > \varepsilon_f$  and  $\lambda_1 > 0$  are preselected constants. Under the estimation law (8), the estimation error  $\theta - \theta^*$  now satisfies

$$\begin{aligned} \frac{d\|\theta - \theta^*\|^2}{dt} &= 2(\theta - \theta^*)^\top [-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}]_{T_\Theta} \\ &\leq 2(\theta - \theta^*)^\top (-\lambda_e K(u, a, \theta)^\top e_{\text{obs}}) \\ &= -2\lambda_e (e_{\text{obs}} - e_f)^\top e_{\text{obs}}, \end{aligned}$$

where the inequality follows from (4) and (2). In the following, we prove that

$$\frac{d\|\theta - \theta^*\|^2}{dt} \leq -2\lambda_e (e_{\text{obs}} - e_f)^\top e_{\text{obs}} \leq 0,$$

where  $d\|\theta - \theta^*\|^2/dt = 0$  if and only if  $\lambda_e = 0$ . Indeed, consider the following two possibilities:

- 1) If  $\lambda_e = 0$ , then  $d\|\theta - \theta^*\|^2/dt = 0$ .
- 2) Otherwise  $\lambda_e = \lambda_1$ , and thus  $\|e_{\text{obs}}\| > (\varepsilon_{\text{obs}} + \varepsilon_f)/2 > \varepsilon_f \geq \|e_f\|$  following (25) and (26). Hence

$$\begin{aligned} \frac{d\|\theta - \theta^*\|^2}{dt} &\leq -2\lambda_1 (e_{\text{obs}} - e_f)^\top e_{\text{obs}} \\ &\leq -\lambda_1 (\|e_{\text{obs}}\|^2 + \|e_{\text{obs}} - e_f\|^2 - \|e_f\|^2) < 0. \end{aligned}$$

The following two results extend Theorems 1 and 2 to the current case without perfect matching between the follower's actual best response  $f^*(u)$  and some  $f(\theta^*, u)$ ; the proofs can be found in [20].

**Theorem 3.** *Suppose that Assumptions 1 and 3 hold. Given any threshold  $\varepsilon_{\text{obs}} > \varepsilon_f$  in (26), the estimation and minimization algorithm (8) and (12) with the switching signal (26) guarantees the following properties:*

- 1) *There exists a time  $T_1 \geq 0$  such that (13) and (14) hold.*
- 2) *The first-order necessary condition for optimality holds asymptotically for the optimization problem (1), that is, (15) holds.*

**Theorem 4.** *Suppose that Assumptions 1–3 hold. Then by setting the threshold*

$$\varepsilon_{\text{obs}} := \sqrt{\alpha_0 \varepsilon_\theta^2 / (2\tau_0) - \varepsilon_f^2}$$

in (26) for any given constant  $\varepsilon_\theta > 2\varepsilon_f \sqrt{\tau_0/\alpha_0}$ , the estimation and minimization algorithm (8) and (12) with the switching signal (26) guarantees the following properties:

- 1) *There exists a time  $T_1 \geq 0$  such that (13), (14) and (22) hold.*
- 2) *The first-order necessary condition for optimality holds asymptotically for the optimization problem (1), that is, (15) holds.*

## VI. SIMULATION EXAMPLE

We illustrate the estimation and minimization algorithm via a simple example motivated by link-flooding denial-of-service (DoS) attacks such as the Crossfire attack [17].

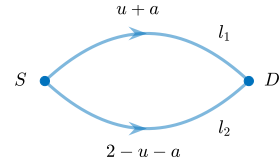


Fig. 1. A simple network with one source  $S$ , one destination  $D$ , and two links  $l_1$  and  $l_2$ .

Consider the communication network in Fig. 1, and suppose there is a router (leader) that distributes 1 unit of legitimate traffic between the two links  $l_1$  and  $l_2$ , and an attacker (follower) that disrupts communication by injecting 1 unit of malicious traffic on the two links. Denote by  $u, a \in [0, 1]$  the amounts of legitimate and malicious traffic on the link  $l_1$ , respectively. Then the total traffic on the links  $l_1$  and  $l_2$  are given by  $f_1 = u + a$  and  $f_2 = 2 - u - a$ , respectively. We assume that a communication delay is incurred on each link, which is a quadratic function of the corresponding total traffic, that is,  $p_1(f_1) := f_1^2$  and  $p_2(f_2) := f_2^2$ , and that the router aims to minimize the average delay for legitimate traffic, which corresponds to the cost function defined by

$$\begin{aligned} J(u, a) &:= up_1(f_1) + (1 - u)p_2(f_2) \\ &= 5u^2 + 6ua + a^2 - 8u - 4a + 4, \end{aligned}$$

whereas the attacker aims to disrupt communication by maximizing the average delay for legitimate traffic, which corresponds to the cost function defined by

$$H(a, u) := -J(u, a).$$

Standard convex analysis shows that the attacker's best response against  $u$  is given by

$$\beta_a(u) := \begin{cases} 0 & u \leq 1/2; \\ 1 & u > 1/2, \end{cases}$$

and the router's best response against  $a$  is given by

$$\beta_u(a) := (4 - 3a)/5.$$

Then it is straightforward to see that a Nash equilibrium does not exist for this game, but there is a Stackelberg equilibrium action  $u^* = 1/2$  for the router.

If the router knew that the attacker's cost function was indeed  $H$ , it could select the Stackelberg equilibrium action

$u^* = 1/2$ . However, we consider a scenario where it does not and, instead, will use the approach proposed in this paper to construct its best action. To this effect, the router will assume the attacker's strategy (best-response function) to be of the form<sup>1</sup>

$$f(\theta, u) := \sum_{j=1}^4 \theta_j F_j(u),$$

where the RBF  $F_j : [0, 1] \rightarrow \mathbb{R}$  are defined by

$$F_j(u) := \mathbb{1}_{(-1/8, 1/8]}(\|u - (2j - 1)/8\|), \quad j = 1, \dots, 4,$$

and  $\theta$  is the parameter estimate in the parameter set  $\Theta := [0, 1]^4$ . For the specific cost function  $H$ , the attacker's actual best response is given by  $f(\theta^*, u)$  with  $\theta^* = (0, 0, 1, 1)$ .

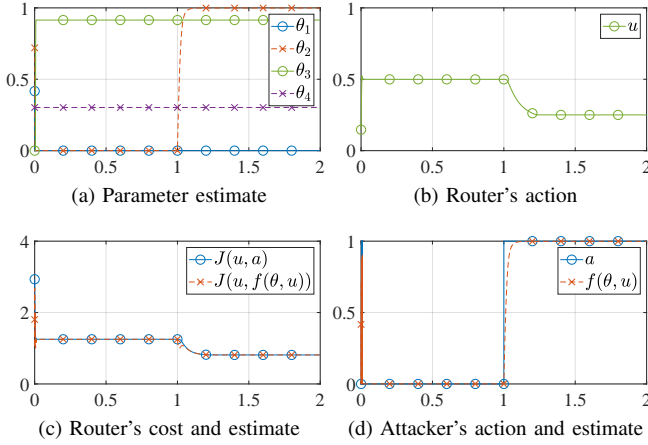


Fig. 2. Simulation results w/o PE (horizontal axis unit:  $\times 10^4$  iterations). In the first  $10^4$  iterations, the router's action  $u$  converges to the optimum  $u^* = 1/2$ ; in the second  $10^4$  iterations, the attacker's cost function changes, and the router's action  $u$  converges to the new optimum  $u^* = 1/4$ ; the parameter estimate  $\theta$  does not converge to the actual parameter  $\theta^*$  in either case.

In simulation shown in Fig. 2 and 3, the threshold  $\varepsilon_{\text{obs}}$  is set to  $10^{-3}$ , and the initial values of the parameter estimate  $\theta$  and the router's action  $u$  are randomly generated. For the case without enforcing PE in Fig. 2, in the first  $10^4$  iterations the router's action  $u$  converges to the optimum  $u^* = 1/2$ , despite that the parameter estimate  $\theta$  does not converge to the actual parameter  $\theta^*$ . In Fig. 3, we enforce PE by adding some random noise to  $u$  for a short interval when the observation error  $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ . In this case, in the first  $10^4$  iterations the router's action  $u$  converges to the optimum  $u^* = 1/2$ , and the parameter estimate  $\theta$  converges to the actual parameter  $\theta^*$ . In both cases, we also simulate the scenario where after  $10^4$  iterations, the attacker starts focusing more on disrupting the link  $l_1$ , so that the new value of the unknown parameter is  $\theta^* = (0, 1, 1, 1)$ , and the new router's Stackelberg equilibrium action is  $u^* = 1/4$ . The

<sup>1</sup>The function  $f$  used here actually violates the regularity conditions in Assumption 1 as it is discontinuous in  $u$ . The continuity requirement of  $f$  in Assumption 1 is only needed so that the gradient descent is well-defined and does not lead to chattering. In simulation, these issues can be handled by using generalized subgradients at discontinuities [24] and setting  $\dot{u} = 0$  when the right-hand side of (12) becomes very small.

corresponding simulation results show that our estimation and minimization algorithm is able to identify this switch in the attack, as the router's action converges to the new optimum  $u^* = 1/4$  in both Fig. 2 and 3, and the parameter estimate  $\theta$  converges to the new parameter  $\theta^*$  in Fig. 3.

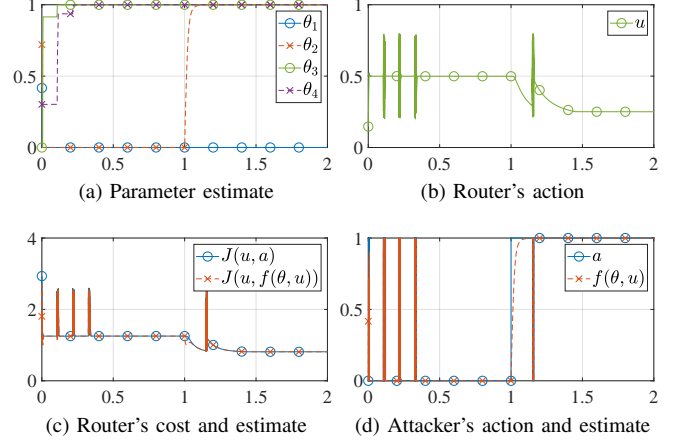


Fig. 3. Simulation results w/ PE (horizontal axis unit:  $\times 10^4$  iterations). In the first  $10^4$  iterations, the router's action  $u$  converges to the optimum  $u^* = 1/2$ ; in the second  $10^4$  iterations, the attacker's cost function changes, and the router's action  $u$  converges to the new optimum  $u^* = 1/4$ ; the parameter estimate  $\theta$  converges to the actual parameter  $\theta^*$  in both cases.

## VII. FUTURE RESEARCH TOPICS

An advantage of our estimation law (8) is that the norm of the estimation error  $\|\theta - \theta^*\|$  will be monotonically non-increasing, and the observation error  $e_{\text{obs}}$  will be bounded by the preselected, arbitrarily small threshold  $\varepsilon_{\text{obs}}$  in finite time, regardless how the leader's action is adjusted. A future research topic is to integrate our estimation law with more efficient optimization methods for minimizing the leader's cost. Other future research topics include to relax the affine condition in Assumption 1, and to extend the current results to Stackelberg games on distributed networks.

### APPENDIX

#### PROJECTED DYNAMICAL SYSTEMS

Let  $\mathcal{S} \subset \mathbb{R}^n$  be a compact convex set, and  $g : \mathcal{S} \rightarrow \mathbb{R}^n$  a locally Lipschitz function. In this section, we prove existence, boundedness, and uniqueness of solutions for the *projected dynamical system*

$$\dot{x} = [g(x)]_{T_{\mathcal{S}}}. \quad (27)$$

The difficulty in analyzing (27) lies in that fact that its right-hand side is potentially discontinuous due to the tangent projection. Therefore, we extend (27) to a differential inclusion that satisfies a suitable notion of semicontinuity. Consider the set-valued function  $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$  defined by

$$G(x) := \bigcap_{\varepsilon > 0} \overline{\text{conv}}\{[g(y)]_{T_{\mathcal{S}}} : y \in \mathcal{B}_{\varepsilon}(x)\}, \quad (28)$$

where  $[y]_{\mathcal{S}} := \arg \min_{z \in \mathcal{S}} \|z - y\|$  denotes the *projection* of a point  $y \in \mathbb{R}^n$  onto  $\mathcal{S}$ . Existence, boundedness, and



uniqueness of Carathéodory solutions for (27) is established by examining the corresponding properties for the differential inclusion

$$\dot{x} \in G(x). \quad (29)$$

**Proposition 5.** *For each  $(t_0, x_0) \in \mathbb{R} \times \mathcal{S}$ , there exists a unique Carathéodory solution to (29) on  $[t_0, \infty)$  with  $x(t_0) = x_0$ , that is, there exists a unique absolutely continuous function  $x : [t_0, \infty) \rightarrow \mathbb{R}^n$  with  $x(t_0) = x_0$  such that (29) holds almost everywhere on  $[t_0, \infty)$ . Moreover,  $x(t) \in \mathcal{S}$  for all  $t \geq t_0$ , and  $x$  is also the unique Carathéodory solution to (27) on  $[t_0, \infty)$  with  $x(t_0) = x_0$ .*

Before proving Proposition 5, we provide some useful properties of the functions  $g$  and  $G$ . As the set  $\mathcal{S}$  is compact and the function  $g$  is locally Lipschitz, there exists a constant  $\gamma \geq 0$  such that

$$\|g(x) - g(y)\| \leq \gamma \|x - y\| \quad \forall x, y \in \mathcal{S}. \quad (30)$$

The following properties follow from the definition of tangent cone (2), the definition (28) of the function  $G$ , and the properties (3), (5) and (30); the proof can be found in [20].

**Lemma 4.** *The set-valued function  $G$  defined by (28) satisfies*

$$g(x) - v \in N_{\mathcal{S}}(x) \quad \forall x \in \mathcal{S}, \forall v \in G(x) \quad (31)$$

and

$$g(x)^\top v \geq \|[g(x)]_{T_{\mathcal{S}}}\|^2 \quad \forall x \in \mathcal{S}, \forall v \in G(x). \quad (32)$$

Following (2), (30) and (31), the set-valued function  $G$  satisfies the following *one-sided Lipschitz condition*

$$(v - w)^\top (x - y) \leq \gamma \|x - y\|^2 \quad (33)$$

for all  $x, y \in \mathcal{S}$ ,  $v \in G(x)$  and  $w \in G(y)$ .

*Proof of Proposition 5.* Consider an arbitrary  $(t_0, x_0) \in \mathbb{R} \times \mathcal{S}$ .

First, we establish existence of Carathéodory solutions for (29). The function  $G$  defined by (28) is upper semicontinuous on  $\mathbb{R}^n$ , and  $G(x)$  is convex and compact for every  $x \in \mathbb{R}^n$ . Also, (30), together with the triangle inequality, implies

$$\|g(x)\| \leq \alpha(1 + \|x\|) \quad \forall x \in \mathcal{S}$$

with the constant  $\alpha := \max_{y \in \mathcal{S}} \{\|g(y)\| + \gamma \|y\|, \gamma\}$ . Consider an arbitrary  $T > t_0$ . Then a standard result on existence of Carathéodory solutions for differential inclusions (see, e.g., [25, Appendix]) implies that there exists at least one absolutely continuous function  $x : [t_0, T] \rightarrow \mathbb{R}^n$  with  $x(t_0) = x_0$  such that (29) holds almost everywhere on  $[t_0, T]$ .

Next, we establish boundedness and uniqueness of Carathéodory solutions for (29). Consider an arbitrary Carathéodory solution  $x$  to (29) on  $[t_0, T]$  with  $x(t_0) = x_0$ . Then [25, Lemma 1] implies  $x(t) \in \mathcal{S}$  for all  $t \in [t_0, T]$ . As  $G$  satisfies the one-sided Lipschitz condition (33), [26, Cor. 2.4] implies that  $x$  is the unique Carathéodory solution to (29) on  $[t_0, T]$  with  $x(t_0) = x_0$ .

Finally, [25, Lemma 2] implies that the solution  $x$  satisfies (27) almost everywhere on  $[t_0, T]$ . Therefore,  $x$  is also the unique Carathéodory solution to (27) on  $[t_0, T]$  with  $x(t_0) = x_0$ . The proof is completed by noticing that  $T > t_0$  is arbitrary.  $\square$

## REFERENCES

- [1] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.
- [2] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. SIAM, 1999.
- [3] T. Alpcan and T. Başar, *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press, 2010.
- [4] J. P. Hespanha, *Noncooperative Game Theory*. Princeton University Press, 2017.
- [5] G. W. Brown, "Iterative solution of games by fictitious play," in *Act. Anal. Prod. Alloc.* John Wiley & Sons, 1951, pp. 374–376.
- [6] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. MIT Press, 1998.
- [7] S. Hart, "Adaptive heuristics," *Econometrica*, vol. 73, no. 5, pp. 1401–1430, 2005.
- [8] J. R. Marden and J. S. Shamma, "Game theory and distributed control," in *Handb. Game Theory with Econ. Appl.* Elsevier, 2015, vol. 4, pp. 861–899.
- [9] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.
- [10] H. von Stackelberg, *Market Structure and Equilibrium*. Springer, 2011.
- [11] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 161–173, 1997.
- [12] T. Roughgarden, "Stackelberg scheduling strategies," *SIAM J. Comput.*, vol. 33, no. 2, pp. 332–350, 2004.
- [13] M. Jain, J. Tsai, J. Pita, C. Kiekintveld, S. Rath, M. Tambe, and F. Ordóñez, "Software assistants for randomized patrol planning for the LAX Airport Police and the Federal Air Marshal Service," *Interfaces*, vol. 40, no. 4, pp. 267–290, 2010.
- [14] M. Brückner and T. Scheffer, "Stackelberg games for adversarial prediction problems," in *17th ACM Int. Conf. Knowl. Discov. Data Min.*, 2011, pp. 547–555.
- [15] J. Marecki, G. Tesauro, and R. Segal, "Playing repeated Stackelberg games with unknown opponents," in *11th Int. Conf. Auton. Agents Multiagent Syst.*, vol. 2, 2012, pp. 821–828.
- [16] A. Blum, N. Haghtalab, and A. D. Procaccia, "Learning optimal commitment to overcome insecurity," in *Neural Inf. Process. Syst.* 2014, 2014, pp. 1826–1834.
- [17] M. S. Kang, S. B. Lee, and V. D. Gligor, "The Crossfire attack," in *2013 IEEE Symp. Secur. Priv.*, 2013, pp. 127–141.
- [18] K.-T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. Chapman & Hall/CRC, 2005.
- [19] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2009.
- [20] G. Yang, R. Poovendran, and J. P. Hespanha, "Adaptive learning in two-player Stackelberg games with continuous action sets," <https://guosong-yang.github.io/resources/YPH-StbgAdptCtrl-Report.pdf>, University of California, Santa Barbara, Tech. Rep., 2019.
- [21] A. S. Morse, D. Q. Mayne, and G. C. Goodwin, "Applications of hysteresis switching in parameter adaptive control," *IEEE Trans. Automat. Contr.*, vol. 37, no. 9, pp. 1343–1354, 1992.
- [22] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Prentice Hall, 1996.
- [23] D. Shevitz and B. Paden, "Lyapunov stability theory of nonsmooth systems," *IEEE Trans. Automat. Contr.*, vol. 39, no. 9, pp. 1910–1914, 1994.
- [24] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Springer, 1998, vol. 317.
- [25] C. Henry, "An existence theorem for a class of differential equations with multivalued right-hand side," *J. Math. Anal. Appl.*, vol. 41, no. 1, pp. 179–186, 1973.
- [26] A. Kastner-Maresch, "Implicit Runge–Kutta methods for differential inclusions," *Numer. Funct. Anal. Optim.*, vol. 11, no. 9–10, pp. 937–958, 1990.