Adaptive learning in two-player Stackelberg games with application to network security

Guosong Yang, Radha Poovendran, and João P. Hespanha

***Abstract*—An adaptive learning approach is proposed for solving two-player Stackelberg games with incomplete information. Specifically, the follower's cost is unknown to the leader, who knows only that the follower's response to its own action belongs to some parametric family of functions, but not the actual parameter value. The proposed approach is capable of simultaneously estimating the unknown parameter and optimizing the leader's action. It ensures that the estimated follower's action and leader's cost become indistinguishable from their actual values in finite time, up to a preselected, arbitrarily small error bound; moreover, the first-order necessary condition for optimality holds asymptotically in time for the estimated leader's cost. Under a persistent excitation condition, the parameter estimation error can be bounded by a preselected, arbitrarily small constant in finite time as well. When the parametric function known to the leader does not match the follower's response function perfectly, the same convergence results can be ensured for preselected error bounds proportional to the size of the mismatch. The approach and convergence results are also extended to the case where the follower's actions cannot be observed, and are illustrated by simulation examples in the domain of network security.**

I. Introduction

In many complex engineering systems, one can find multiple decision-making agents, each of which aims to minimize an individual cost that also depends on the others' decisions. Examples of such agents and systems include processes in a multitasking central processing unit, vehicles on a public highway, and routers in a shared telecommunication network such as the Internet. Originated from economics, game theory provides a rich set of mathematical tools for understanding the interplay between *rational* agents [1], which proves to be crucial for optimizing both individual and overall performance. Over the past decades, game-theoretic tools have become prevalent in numerous engineering fields such as circuit design, congestion control, and network security [2]–[4].

Most research in game theory has focused on equilibrium in games, especially Nash equilibrium—a tuple consisting of one strategy for each player from which no one has an incentive to deviate by itself [1]. This raises the question of when and how players can reach an equilibrium without complete information of the game. Such a question is especially important in network security, where it is usually difficult to gather information about an attack before it actually takes place. See [5] for an overview of standard learning rules for Nash games.

Another common issue in many engineering applications of game theory is that players have asymmetric information. For example, in a link-flooding distributed denial-of-service (DDoS) attack such as the Crossfire attack [6], the attacker is able to monitor changes in routing and make rapid adjustments to sustain a high congestion level, whereas information about the attack is usually unavailable or slow to be acquired. This motivates us to consider Stackelberg games—a hierarchical game model in which one player (called the *follower*) is able to observe the action of the other player (called the *leader*) before making its own selection [7]. For the example of link-flooding DDoS attack above, it is natural to model the attacker as the follower in a Stackelberg game in which the leader is a router that tries to minimize the effect of the attack.

A Stackelberg equilibrium corresponds to a leader's action that minimizes its own cost, assuming that the follower selects its action in response to the advertised leader's action. Conditions for the existence of a Stackelberg equilibrium are generally much weaker than those for a Nash equilibrium [2, p. 181], and there are many games with information asymmetry where the former exists but the latter does not. For such games, one cannot rely on standard learning rules for Nash equilibrium, but requires a novel approach for reaching a Stackelberg equilibrium.

In this paper, we propose an adaptive learning approach for solving two-player Stackelberg games with incomplete information about the follower, which provides provable guarantees for both convergence and optimality. In Section II, we present a problem formulation in which neither the follower's cost function nor its action set is known to the leader. Instead, the follower's response to a leader's action is assumed to belong to a known parametric family of functions, but the actual value of the parameter vector is unknown. Our approach for simultaneously estimating the unknown parameter and optimizing the leader's action is formulated in Section III. The estimation dynamics are constructed by comparing past observations of follower's action and leader's cost with their estimated values. The optimization dynamics are constructed based on the current parameter estimate rather than the unknown actual parameter value. Our design utilizes adaptive control tools including projected gradient descent and hysteresis switching to ensure feasibility, convergence, and optimality.

In Section IV, we prove that the estimated follower's action and leader's cost reach in finite time values that are indistinguishable from their actual values, up to a preselected, arbitrarily small error bound; moreover, the first-order necessary condition for optimality holds asymptotically in time for the estimated leader's cost. Provided that a persistent excitation

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 and Computer Engineering, University of Washington, Seattle, WA 98195, USA. Email: rp3@uw.edu.

condition holds, the parameter estimation error can be bounded by a preselected, arbitrarily small constant in finite time as well. Our proof provides a rigorous treatment for the existence and convergence of solutions to the discontinuous dynamics resulting from projection and switching. To achieve this, we establish an invariance theorem for projected gradient descent in continuous time using tools from differential inclusions theory, which is of independent interest and novel to the best of our knowledge.

In Section V, we analyze the proposed approach in the more complicated scenario where the parametric function known to the leader does not match the follower's response function perfectly. It is shown that the same convergence results can be guaranteed for preselected error bounds that are proportional to the size of the mismatch between the two functions.

In Section VI, we consider the more challenging case where the leader cannot observe the follower's actions. By constructing parametric estimation for the leader's cost (as a function of only its own action) rather than the follower's response, we are able to adjust the proposed approach and ensure similar convergence properties to those in Sections IV and V, in both scenarios with and without mismatch in parameterization.

In Section VII, the approach and convergence results are illustrated through simulation examples motivated by link-flooding DDoS attacks. Section VIII concludes the paper with a brief summary and an outlook on future research directions.

A preliminary version of some of these results was presented in the paper [8]. The current paper improves upon [8] by adopting a notion of "practical" Stackelberg equilibrium (Definition 1), which enables us to remove unnecessary assumptions (see Assumptions 1 and 4 and Appendix A). Moreover, we provide a more detailed analysis in the scenario with mismatch in parameterization (Section V), new results in the case with unobservable follower's actions (Section VI), and more realistic and elaborate simulation examples (Section VII).

Related work

Stackelberg game: Introduced in 1934 as a model to explain oligopoly [7], Stackelberg games have been frequently used to model engineering problems with information asymmetry such as routing [9], scheduling [10], and channel allocation [11]. In the national security domain, mixed-integer algorithms developed for solving Stackelberg games formed the basis of several real-world defense programs, including the ARMOR program deployed at the Los Angeles International Airport [12], the IRIS program used by the US Federal Air Marshals [13], and several counterterrorism programs for crucial infrastructures such as power grid and oil reserves [14], [15]. More recently, Stackelberg semi-Markov games were used in detecting advanced persistent threats in network security [16].

Learning in games: A well-known learning rule for Nash games is fictitious play [17], [18]. In fictitious play, a game with discrete action sets is played repeatedly toward the goal of reaching a Nash equilibrium. All players assume that the opponents are playing stationary mixed strategies, and continuously estimate these strategies by computing historical frequencies of actions. Then they each selects an action that is optimal against the estimated opponents' strategies. Another common learning approach is to rely on gradient dynamics, namely, each player takes a step along the gradient descent direction of its cost function, computed with the current opponents' strategies [19], [20] or their empirical estimates [21]. Fictitious play and gradient dynamics have drawn significant research interests [5], [22] and have been frequently used in multi-agent reinforcement learning [23] and distributed control [24]. The replicator equation, a prototype of evolutionary games, often generates the same asymptotic behaviors as those by the best response dynamics, a continuous-time variation of fictitious play [5], [25]. For Stackelberg games, several learning rules have been proposed in Bayesian setups [26]–[28]. However, these results are limited to games with discrete action sets, which are often too restrictive for engineering applications such as network security.

Notations: Let $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{Z}_+ = \{0, 1, \ldots\}$. Denote by I_n the identity matrix in $\mathbb{R}^{n\times n}$; the subscript is omitted when the dimension is clear from context. For a vector $v \in \mathbb{R}^n$, denote by r_i its i-th scalar component and write $r = (r_1, \ldots, r_n)$. For a set $\mathcal{S} \subset \mathbb{R}^n$, denote by $\operatorname{cl}\mathcal{S}$, $\partial\mathcal{S}$, and $\operatorname{co}\mathcal{S}$ its closure, boundary, and convex hull, respectively. Denote by $\|v\|$ the Euclidean norm of a vector $v \in \mathbb{R}^n$, and by $\|A\|$ the corresponding induced norm (also called the spectral norm) of a matrix $A \in \mathbb{R}^{n\times n}$. Denote by $sB(x)$ the closed ball of radius $s \geq 0$ centered at a point x in $\mathbb{R}^n$, namely, $sB(x) := \{z \in \mathbb{R}^n : \|z-x\| \leq s\}$. For a function $f(x, z)$ from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^k$, denote by $\nabla_x f(v, w)$ its Jacobian matrix with respect to x at (v, w); if $k = 1$ then $\nabla_x f(v, w)$ is the gradient with respect to x at (v, w), taken as a row vector for consistency.

II. Problem formulation

A two-player game can be defined by a tuple $(\mathcal{R}, \mathcal{A}, J, H)$, in which the first and the second player select actions $r \in \mathcal{R} \subset \mathbb{R}^{n_r}$ and $a \in \mathcal{A} \subset \mathbb{R}^{n_a}$ and pay costs $J(r, a)$ and $H(r, a)$, respectively. In a Stackelberg model, the second player (the follower) can observe the choice of the first one (the leader) before making its own selection [7]. Hence a rational follower will always select an action from the *best-response set*

$$\beta_a(r) := \operatorname*{arg\,min}_{a\in\mathcal{A}} H(r, a) \tag{1}$$

against a leader's action r. We restrict our attention to compact action sets and continuous cost functions; hence the set $\beta_a(r)$ is nonempty and compact for each $r \in \mathcal{R}$. The notion of Stackelberg equilibrium is defined as follows (see, e.g., [2, Def. 4.6 and 4.7, pp. 179–180]):

Definition 1 (Stackelberg equilibrium)**.** Given a two-player game $(\mathcal{R}, \mathcal{A}, J, H)$, an action $r^* \in \mathcal{R}$ is called a *Stackelberg equilibrium action* for the leader if

$$\max_{a\in\beta_a(r^*)} J(r^*, a) = J^* := \inf_{r\in\mathcal{R}} \max_{a\in\beta_a(r)} J(r, a), \tag{2}$$

where the follower's best-response set $\beta_a(r)$ is defined by (1), and J^* is known as the *Stackelberg cost* for the leader. For a constant $\varepsilon > 0$, an action $r^*_\varepsilon \in \mathcal{R}$ is called an ε *Stackelberg action* for the leader if

$$\max_{a\in\beta_a(r^*_\varepsilon)} J(r^*_\varepsilon, a) \leq J^* + \varepsilon.$$

The Stackelberg cost provides a cost that the leader is able to guarantee against a rational follower. However, the leader may receive a better (smaller) cost while playing a Stackelberg equilibrium action, since the follower's actual action does not necessarily maximize the leader's cost over the best-response set; see [29], [30] and [31, Sec. 15.3] for more discussions. In practice, it is possible that no Stackelberg equilibrium action exists, but the leader is able to guarantee essentially the Stackelberg cost by playing an ε Stackelberg action for a sufficiently small $\varepsilon > 0$ (see the discussion after Assumption 1 and the numerical examples in Section VII).

We are interested in games with incomplete information in which the leader does not know the follower's cost function or action set and thus cannot simply compute a Stackelberg equilibrium action or an ε Stackelberg action. Specifically, the follower selects its action by an unknown function $f : \mathcal{R} \to \mathcal{A}$ such that

$$f(r) \in \beta_a(r) \qquad \forall\, r \in \mathcal{R}.$$

To avoid confusion, we use the terminology *follower's strategy* for the function $f(\cdot)$, and *follower's action* for the value $f(r)$ obtained for a given leader's action $r \in \mathcal{R}$. While the specific follower's strategy f is unknown, we do know that it belongs to a parametric family of functions $\{r \mapsto \hat{f}(\theta, r) : \theta \in \Theta \subset \mathbb{R}^{n_\theta}\}$, namely, there is a parameter value $\theta \in \Theta$ such that

$$f(r) = \hat{f}(\theta, r) \qquad \forall\, r \in \mathcal{R}. \tag{3}$$

The parametric form $\hat{f} : \Theta \times \mathcal{R} \to \mathbb{R}^{n_a}$, including the set Θ, is known to the leader, but the actual parameter value θ is unknown. The following generic regularity conditions are imposed to ensure that an ε Stackelberg action exists, and the gradients and projections needed for constructing our estimation and optimization dynamics are well-defined.

Assumption 1 (Regularity)**.** The leader's action set $\mathcal{R}$ and the parameter set Θ are compact and convex, and the follower's action set $\mathcal{A}$ is compact; the leader's cost function J and the parametric function $\hat{f}$ are continuously differentiable, and the follower's cost function H is continuous.

As a result of Assumption 1, an ε Stackelberg action for the leader exists for every $\varepsilon > 0$ [2, Prop. 4.2, p. 180]. The conditions in Assumption 1 are much weaker than standard conditions for the existence of a Stackelberg equilibrium action (see, e.g., [2, Th. 4.8, p. 180]), which are in turn much weaker than those for a Nash equilibrium (see, e.g., [2, Th. 4.3, p. 173]). Therefore, they are in line with our interest in games where no Nash equilibrium exists but a "practical" Stackelberg equilibrium does (see the numerical examples in Section VII).

In practice, there is little loss of generality in assuming that the unknown follower's strategy f belongs to a known parametric family of functions, as one can always construct an arbitrarily accurate approximation for f on the compact set $\mathcal{R}$ using a finite weighted sum of a preselected class of basis functions. A well-known method for constructing such approximation is the *radial basis function (RBF)* method [32, Sec. 3], in which the parametric function takes the form

$$\hat{f}(\theta, r) := \sum_{j=1}^{n_\theta} \theta_j b_j(r) := \sum_{j=1}^{n_\theta} \theta_j \phi(\|r - r_j^c\|/\mu_j),$$

where $\phi : \mathbb{R}_+ \to \mathbb{R}^{n_a}$ is a fixed kernel function, and each $b_j : \mathcal{R} \to \mathbb{R}^{n_a}$ is an RBF with center $r_j^c \in \mathbb{R}^{n_r}$ and scaling factor $\mu_j > 0$. Note that the parametric function $\hat{f}(\theta, r)$ in the RBF method is affine in θ for each fixed $r \in \mathcal{R}$. This property also holds for many other widely-used function approximation methods such as those based on polynomials and splines [32, Sec. 3], which motivates us to focus exclusively on such parametric functions in this paper.

Assumption 2 (Affinity)**.** The map $\theta \mapsto \hat{f}(\theta, r)$ is affine for each fixed $r \in \mathcal{R}$.

The leader's goal is to minimize its cost $J(r, a)$ for the follower's action $a = f(r) = \hat{f}(\theta, r)$, namely, to solve the optimization problem

$$\min_{r \in \mathcal{R}} J(r, a) = \min_{r \in \mathcal{R}} J\big(r, \hat{f}(\theta, r)\big), \tag{4}$$

using past observations of follower's action a and leader's cost $J(r, a)$, but without knowing the actual parameter value θ. To solve this problem, we propose an adaptive learning approach that consists of the following two components:

1) Constructing a *parameter estimate* $\hat{\theta} \in \Theta$ that approaches the unknown value θ.
2) Adjusting the leader's action r to minimize its *estimated cost*

$$\hat{J}(\hat{\theta}, r) := J\big(r, \hat{f}(\hat{\theta}, r)\big),$$

namely, to solve the optimization problem

$$\min_{r \in \mathcal{R}} \hat{J}(\hat{\theta}, r) = \min_{r \in \mathcal{R}} J\big(r, \hat{f}(\hat{\theta}, r)\big). \tag{5}$$

Our approach is designed based on continuous-time dynamical systems, which is a common practice in the literature of learning in games [5], [22].

III. Estimation and optimization

In this section, we construct dynamical systems for estimating the unknown parameter value and optimizing the leader's action. First, we recall some notions and basic properties from convex analysis; more details can be found in standard textbooks, e.g, [33, Ch. 6] or [34, Sec. 5.1].

We denote by $[x]_{\mathcal{S}}$ the *projection* of a point $x \in \mathbb{R}^n$ onto a closed convex set $\mathcal{S} \subset \mathbb{R}^n$, namely,

$$[x]_{\mathcal{S}} \in \arg\min_{z \in \mathcal{S}} \|z - x\|,$$

which is unique and satisfies $[x]_{\mathcal{S}} = x$ for all $x \in \mathcal{S}$.

We denote by $T_{\mathcal{S}}(x)$ and $N_{\mathcal{S}}(x)$ the *tangent cone* and the *normal cone* to a convex set $\mathcal{S} \subset \mathbb{R}^n$ at a point $x \in \mathcal{S}$, respectively, namely,

$$T_{\mathcal{S}}(x) := \mathrm{cl}\{s(z - x) : z \in \mathcal{S},\ s > 0\},$$
$$N_{\mathcal{S}}(x) := \{v \in \mathbb{R}^n : \forall\, z \in \mathcal{S},\ v^\top (z - x) \le 0\},$$

which are closed and convex sets that satisfy $T_{\mathcal{S}}(x) = \mathbb{R}^n$ and $N_{\mathcal{S}}(x) = \{0\}$ for all $x \in \mathcal{S} \backslash \partial\mathcal{S}$. For arbitrary convex set $\mathcal{S} \subset \mathbb{R}^n$ and point $x \in \mathcal{S}$, we have

$$z - [z]_{T_{\mathcal{S}}(x)} \in N_{\mathcal{S}}(z),\ \ [z]_{T_{\mathcal{S}}(x)}^\top z = \big\|[z]_{T_{\mathcal{S}}(x)}\big\|^2 \qquad \forall\, z \in \mathbb{R}^n \tag{6}$$

and

$$v^\top w \le 0,\ \ [w]_{T_{\mathcal{S}}(x)} = 0 \qquad \forall\, v \in T_{\mathcal{S}}(x), w \in N_{\mathcal{S}}(x). \tag{7}$$

A. Parameter estimation

Our goal is to design a parameter estimate $\hat{\theta}$ for the unknown value θ so that the *estimation error* $\hat{\theta}-\theta$ decreases monotonically in norm, regardless of how the leader's action r is adjusted. Since neither $\hat{\theta}-\theta$ nor its norm can be observed directly, we construct $\hat{\theta}$ in terms of the *observation error*

$$e_{\text{obs}} := \begin{bmatrix} \hat{J}(\hat{\theta}, r) - J(r, a) \\ \hat{a} - a \end{bmatrix}, \tag{8}$$

which includes the difference between the observed follower's action $a = \hat{f}(\theta, r)$ and its current estimate $\hat{a} := \hat{f}(\hat{\theta}, r)$, as well as the difference between the known leader's cost $J(r, a)$ and the estimate $\hat{J}(\hat{\theta}, r) = J(r, \hat{a})$. The following lemma establishes a relation between e_{obs} and $\hat{\theta}-\theta$:

Lemma 1. *For every $r \in \mathcal{R}$ and $\theta, \hat{\theta} \in \Theta$, we have*

$$e_{\text{obs}} = K(r, a, \hat{\theta})(\hat{\theta}-\theta), \tag{9}$$

where the gain matrix $K(r, a, \hat{\theta})$ is defined by

$$K(r, a, \hat{\theta}) := \begin{bmatrix} \int_0^1 \nabla_a J(r, \rho\hat{a} + (1-\rho)a)\,\mathrm{d}\rho \\ I \end{bmatrix} \nabla_\theta \hat{f}(r). \tag{10}$$

With a slight abuse of notation, we denote by $\nabla_\theta \hat{f}(r)$ the Jacobian matrix of $\hat{f}(\theta, r)$ with respect to θ, as it is independent of θ under the affine condition in Assumption 2.

Proof of Lemma 1. Given $a = \hat{f}(\theta, r)$ and $\hat{a} = \hat{f}(\hat{\theta}, r)$, consider the function $h : [0, 1] \to \mathbb{R}$ defined by

$$h(\rho) := J(r, \rho\hat{a} + (1-\rho)a).$$

As the function J is continuously differentiable, so is h. Hence

$$\begin{aligned} \hat{J}(\hat{\theta}, r) - J(r, a) &= h(1) - h(0) = \int_0^1 \frac{\mathrm{d}h(\rho)}{\mathrm{d}\rho}\,\mathrm{d}\rho \\ &= \left(\int_0^1 \nabla_a J(r, \rho\hat{a} + (1-\rho)a)\,\mathrm{d}\rho\right)(\hat{a} - a). \end{aligned}$$

Moreover, as the Jacobian matrix $\nabla_\theta \hat{f}(r)$ is independent of θ following the affine condition in Assumption 2, we have

$$\hat{a} - a = \hat{f}(\hat{\theta}, r) - \hat{f}(\theta, r) = \nabla_\theta \hat{f}(r)(\hat{\theta} - \theta).$$

Therefore, (9) holds. □

As a result of Lemma 1, we would have an observation error $e_{\text{obs}} = 0$ if the current parameter estimate was correct, namely, $\hat{\theta} = \theta$. However, in most applications, the dimension n_θ of the parameter vector θ is much higher than the dimension $n_a + 1$ of the observables, in which case the gain matrix $K(r, a, \hat{\theta})$ cannot be injective, and a zero observation error e_{obs} does not imply a correct parameter estimate $\hat{\theta}$.

To ensure that the estimation error $\hat{\theta}-\theta$ decreases monotonically in norm, we propose the following estimation dynamics:

$$\dot{\hat{\theta}} = \left[-\lambda_e K(r, a, \hat{\theta})^\top e_{\text{obs}}\right]_{T_\Theta(\hat{\theta})}, \tag{11}$$

where the gain matrix $K(r, a, \hat{\theta})$ is defined by (10), and $\lambda_e : \mathbb{R}_+ \to \{0,\ \lambda_\theta\}$ is a *switching signal* defined by

$$\lambda_e(t) := \begin{cases} \lambda_\theta & \text{if } \|e_{\text{obs}}(t)\| \geq \varepsilon_{\text{obs}}; \\ 0 & \text{if } \|e_{\text{obs}}(t)\| \leq \varepsilon'_{\text{obs}}; \\ \lim_{s \nearrow t} \lambda_e(s) & \text{otherwise} \end{cases} \tag{12}$$

and $\lambda_e(0) := \lambda_\theta$ if $\|e_{\text{obs}}(0)\| \in (\varepsilon'_{\text{obs}}, \varepsilon_{\text{obs}})$ with preselected constants $\varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > 0$ and $\lambda_\theta > 0$. The estimation dynamics (11) has the following features: First, its implementation only requires variables known to the leader; specifically, computing $K(r, a, \hat{\theta})$ does not require knowing the actual parameter value θ. Second, the right-hand side of (11) is projected onto the tangent cone $T_\Theta(\hat{\theta})$ to ensure that the parameter estimate $\hat{\theta}$ is always inside the feasible set Θ [34]. However, this operation introduces discontinuities and requires tools from differential inclusions theory to ensure global existence and convergence of solutions. Finally, the switching signal λ_e is designed so that the adaptation of $\hat{\theta}$ is on when $\|e_{\text{obs}}\| \geq \varepsilon_{\text{obs}}$ and off when $\|e_{\text{obs}}\| \leq \varepsilon'_{\text{obs}}$, using a hysteresis switching rule to prevent chattering.

Under (11), the estimation error $\hat{\theta}-\theta$ satisfies

$$\begin{aligned} \frac{\mathrm{d}\|\hat{\theta}-\theta\|^2}{\mathrm{d}t} &= 2(\hat{\theta}-\theta)^\top \left[-\lambda_e K(r, a, \hat{\theta})^\top e_{\text{obs}}\right]_{T_\Theta(\hat{\theta})} \\ &\leq 2(\hat{\theta}-\theta)^\top \left(-\lambda_e K(r, a, \hat{\theta})^\top e_{\text{obs}}\right) = -2\lambda_e \|e_{\text{obs}}\|^2, \end{aligned}$$

where the inequality follows from the fact that $\theta - \hat{\theta} \in T_\Theta(\hat{\theta})$ and the first properties in (6) and (7), and the last equality follows from (9). Hence we conclude that the estimation dynamics (11) ensures that

$$\frac{\mathrm{d}\|\hat{\theta}-\theta\|^2}{\mathrm{d}t} \leq -2\lambda_e \|e_{\text{obs}}\|^2 \leq 0, \tag{13}$$

which shows that $\|\hat{\theta}-\theta\|$ decreases monotonically. Moreover, the definition of λ_e in (12) implies that $\|\hat{\theta}-\theta\|$ does not stop approaching 0 as long as $\|e_{\text{obs}}\| \geq \varepsilon_{\text{obs}}$. In Section IV, we will prove that the adaptation of $\hat{\theta}$ is guaranteed to terminate in finite time, after which $\|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ will always hold.

B. Cost minimization

The estimation dynamics (11) ensures that the estimation error $\hat{\theta}-\theta$ decreases monotonically in norm regardless of the leader's actions $r \in \mathcal{R}$. This enables a wide choice of algorithms to adjust r toward a Stackelberg equilibrium action. Our analysis in this paper is focused on adjusting r via a gradient descent method, which is easy to implement and fairly robust for a broad range of applications. The leader's ultimate goal is to minimize its cost $J(r, a) = J(r, \hat{f}(\theta, r))$, for which the gradient descent direction depends on the unknown parameter value θ. To overcome this, we change the optimization objective to the estimated leader's cost $\hat{J}(\hat{\theta}, r) = J(r, \hat{f}(\hat{\theta}, r))$, which depends instead on the parameter estimate $\hat{\theta}$. This change is justified by the property that the difference between the actual and estimated leader's costs will be bounded by $\|\hat{J}(\hat{\theta}, r) - J(r, a)\| \leq \|e_{\text{obs}}\| < \varepsilon_{\text{obs}}$ in finite time, as we shall prove in Section IV.

The time derivative of the estimated leader's cost $\hat{J}(\hat{\theta}, r)$ is given by

$$\dot{\hat{J}}(\hat{\theta}, r) = \nabla_\theta \hat{J}(\hat{\theta}, r)\,\dot{\hat{\theta}} + \nabla_r \hat{J}(\hat{\theta}, r)\,\dot{r}, \tag{14}$$

where

$$\begin{aligned} \nabla_\theta \hat{J}(\hat{\theta}, r) &= \nabla_a J(r, \hat{a}) \nabla_\theta \hat{f}(r), \\ \nabla_r \hat{J}(\hat{\theta}, r) &= \nabla_r J(r, \hat{a}) + \nabla_a J(r, \hat{a}) \nabla_r \hat{f}(\hat{\theta}, r) \end{aligned}$$

with $\hat{a} = \hat{f}(\hat{\theta}, r)$. Since we will prove that the adaptation of $\hat{\theta}$ terminates in finite time, we can neglect the term with $\dot{\hat{\theta}}$ in (14) and focus exclusively on adjusting r along the gradient descent direction of $\hat{J}(\hat{\theta}, r)$ in r, which leads to the following optimization dynamics:

$$\dot{r} = \left[-\lambda_r \nabla_r \hat{J}(\hat{\theta}, r)^\top\right]_{T_{\mathcal{R}}(r)} \tag{15}$$

with a preselected constant $\lambda_r > 0$. Note that the right-hand side of (15) is projected onto the tangent cone $T_{\mathcal{R}}(r)$ to ensure that the leader's action r is always inside its action set $\mathcal{R}$ [34].

Substituting (15) into (14) yields

$$\begin{aligned}\dot{\hat{J}}(\hat{\theta}, r) &= \nabla_r \hat{J}(\hat{\theta}, r)\left[-\lambda_r \nabla_r \hat{J}(\hat{\theta}, r)^\top\right]_{T_{\mathcal{R}}(r)} + \nabla_\theta \hat{J}(\hat{\theta}, r)\,\dot{\hat{\theta}} \\ &= -\left\|\left[-\lambda_r \nabla_r \hat{J}(\hat{\theta}, r)^\top\right]_{T_{\mathcal{R}}(r)}\right\|^2 \Big/ \lambda_r + \nabla_\theta \hat{J}(\hat{\theta}, r)\,\dot{\hat{\theta}} \\ &= -\|\dot{r}\|^2/\lambda_r + \nabla_\theta \hat{J}(\hat{\theta}, r)\,\dot{\hat{\theta}},\end{aligned}$$

where the second equality follows from the second property in (6). Hence we conclude that the optimization dynamics (15) ensures that

$$\dot{\hat{\theta}} = 0 \implies \dot{\hat{J}}(\hat{\theta}, r) \le -\|\dot{r}\|^2/\lambda_r \le 0,$$

which shows that $\hat{J}(\hat{\theta}, r)$ decreases monotonically when the adaptation of $\hat{\theta}$ stops. In Section IV, we will prove that the leader's action r is guaranteed to converge asymptotically to a set where the *first-order necessary condition (FONC) for optimality* holds for the optimization problem (5).

IV. Convergence analysis

In this section, we prove convergence properties for the estimation and optimization dynamics introduced in Section III.

Theorem 1. *Under Assumptions 1 and 2, given arbitrary $\varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > 0$ in* (12)*, the estimation and optimization dynamics* (11) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \ge 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} satisfy*
$$\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}}, \quad \hat{\theta}(t) = \hat{\theta}(T) \qquad \forall\, t \ge T. \tag{16}$$
2) *The estimated leader's cost $\hat{J}(\hat{\theta}, r)$ satisfies*
$$\lim_{t\to\infty} \left[-\nabla_r \hat{J}(\hat{\theta}(T), r(t))^\top\right]_{T_{\mathcal{R}}(r(t))} = 0. \tag{17}$$

Specifically, item 1) ensures that the adaptation of $\hat{\theta}$ terminates in finite time, after which one cannot distinguish $\hat{\theta}$ from the actual parameter value θ by observing the follower's action $a = \hat{f}(\theta, r)$ and the leader's cost $J(r, a)$, up to an error e_{obs} bounded in norm by the preselected constant ε_{obs}. Item 2) means that the FONC for optimality holds asymptotically for the optimization problem (5), as justified by the next lemma:

Lemma 2. *For each fixed $\hat{\theta} \in \Theta$, at a local optimum $\hat{r}^*$ of the optimization problem* (5)*, we have*

$$\left[-\nabla_r \hat{J}(\hat{\theta}, \hat{r}^*)^\top\right]_{T_{\mathcal{R}}(\hat{r}^*)} = 0. \tag{18}$$

Proof. It is a standard result in constrained optimization that $-\nabla_r \hat{J}(\hat{\theta}, \hat{r}^*)^\top \in N_{\mathcal{R}}(\hat{r}^*)$; see, e.g., [33, Th. 6.12, p. 207]. Then (18) follows from the second property in (7). □

Proof of Theorem 1. Due to projection and switching, there may be discontinuities in the right-hand sides of (11) and (15). Hence we prove Theorem 1 using tools from differential inclusions theory; see Appendix A for the required preliminaries.

First, we establish existence of solutions for the projected dynamical system defined by (11) and (15).

Lemma 3. *For each initial value $(\hat{\theta}_0, r_0) \in \Theta \times \mathcal{R}$, there exists a solution to* (11) *and* (15) *over $\mathbb{R}_+$; specifically, there exist absolutely continuous functions $\hat{\theta} : \mathbb{R}_+ \to \Theta$ and $r : \mathbb{R}_+ \to \mathcal{R}$ such that* (11) *and* (15) *hold almost everywhere in $\mathbb{R}_+$, with $(\hat{\theta}(0), r(0)) = (\hat{\theta}_0, r_0)$. Moreover, $\hat{\theta}$, r, and e_{obs} defined by* (8)*, and their time derivatives $\dot{\hat{\theta}}$, $\dot{r}$, and $\dot{e}_{\text{obs}}$, are essentially bounded over $\mathbb{R}_+$.*

Proof. Lemma 3 can be proved using results on hysteresis switching [35] and projected differential inclusions [34]; see Appendix B for the complete proof. □

We are now ready to prove item 1) of Theorem 1 using similar arguments to those used in the proof of Barbalat's lemma [36, Lemma 3.2.6, p. 76]. Note that Barbalat's lemma cannot be applied directly becasue $\dot{e}_{\text{obs}}$ may be piecewise continuous instead of continuous, due to projection and switching. Recall that $\|\hat{\theta}-\theta\|$ decreases monotonically following (13). As $\|\hat{\theta}-\theta\|$ is bounded from below by 0, the limit $\lim_{t\to\infty} \|\hat{\theta}(t)-\theta\|$ exists and is finite; hence

$$\lim_{t\to\infty} \int_0^t \lambda_e(s)\|e_{\text{obs}}(s)\|^2 \,\mathrm{d}s \tag{19}$$

exists and is finite. Meanwhile, (11) and (12) imply that (16) holds if there is a time $T \ge 0$ for which

$$\lambda_e(t) = 0 \qquad \forall\, t \ge T, \tag{20}$$

which we will prove by contradiction. Assume that (20) does not hold for any $T \ge 0$. Then (12) implies that there exists a strictly increasing, unbounded time sequence $(t_k)_{k\in\mathbb{Z}_+}$ with $t_0 > 0$ such that for all $k \in \mathbb{Z}_+$, we have $\lambda_e(t_k) = \lambda_\theta$ and thus $\|e_{\text{obs}}(t_k)\| > \varepsilon'_{\text{obs}}$. Recall that $\dot{e}_{\text{obs}}$ is essentially bounded over $\mathbb{R}_+$, and let

$$\delta := \min\left\{t_0,\ \frac{\varepsilon_{\text{obs}} - \varepsilon'_{\text{obs}}}{\operatorname{ess\,sup}_{s\ge 0} \|\dot{e}_{\text{obs}}(s)\|}\right\} > 0.$$

For each $k \in \mathbb{Z}_+$, consider the following two possibilities:

1) If there is a time $s_k \in [t_k - \delta, t_k]$ such that $\|e_{\text{obs}}(s_k)\| = \varepsilon_{\text{obs}}$, then (12) and the definition of δ imply that
$$\|e_{\text{obs}}(t)\| > \varepsilon'_{\text{obs}}, \ \ \lambda_e(t) = \lambda_\theta \quad \forall\, t \in [s_k, s_k + \delta). \tag{21}$$
2) Otherwise $\|e_{\text{obs}}(t)\| < \varepsilon_{\text{obs}}$ for all $t \in [t_k - \delta, t_k]$; hence (12) with $\lambda_e(t_k) = \lambda_\theta$ implies that (21) holds with $s_k = t_k - \delta$.

In summary, there is an unbounded time sequence $(s_k)_{k\in\mathbb{Z}_+}$ such that (21) holds for all $k \in \mathbb{Z}_+$. Then we have

$$\int_{s_k}^{s_k+\delta} \lambda_e(s)\|e_{\text{obs}}(s)\|^2 \,\mathrm{d}s > \lambda_\theta(\varepsilon'_{\text{obs}})^2\delta > 0 \qquad \forall\, k \in \mathbb{Z}_+,$$

which contradicts the property that (19) exists and is finite. Hence there is a time $T \ge 0$ for which (20), and therefore (16), are both satisfied.

Finally, we prove item 2) of Theorem 1 using the invariance theorem for projected gradient descent from Appendix A: After the time T in (16), the optimization dynamics (15) become

$$\dot{r} = \left[-\lambda_r \nabla_r \hat{J}(\hat{\theta}(T), r)^\top\right]_{T_{\mathcal{R}}(r)},$$

which can be modeled as a projected dynamical system (45) with the state $x := r$, the set $\mathcal{S} := \mathcal{R}$, and the function

$$g(r) := -\lambda_r \nabla_r \hat{J}(\hat{\theta}(T), r)^\top.$$

Hence (46) in Proposition 1 holds with $V(x) := \lambda_r \hat{J}(\hat{\theta}(T), x)$. Then (17) follows from (47). □

Theorem 1 makes no claim regarding the size of the final estimation error $\hat{\theta}(T) - \theta$. However, a small $\|\hat{\theta}(T) - \theta\|$ can be guaranteed under a *persistent excitation (PE)* condition.

Assumption 3 (PE)**.** The gain matrix defined by (10), written as $K(t) := K(r(t), a(t), \hat{\theta}(t))$ for brevity, satisfies

$$\int_t^{t+\tau_0} K(s)^\top K(s)\,\mathrm{d}s - \alpha_0 I \geq 0 \qquad \forall\, t \geq 0 \tag{22}$$

for some constants $\tau_0, \alpha_0 > 0$ (i.e., the matrix on the left-hand side of the first inequality in (22) is positive semidefinite).

Theorem 2. *Under Assumptions 1–3, given an arbitrary $\varepsilon_\theta > 0$, if ε_{obs} and $\varepsilon'_{\text{obs}}$ in* (12) *are selected so that*

$$\varepsilon_\theta \sqrt{\alpha_0/\tau_0} \geq \varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > 0, \tag{23}$$

then the estimation and optimization dynamics (11) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \geq 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} satisfy* (16) *and*

$$\|\hat{\theta}(T) - \theta\| < \varepsilon_\theta. \tag{24}$$

2) *The asymptotic FONC for optimality* (17) *holds for the estimated leader's cost $\hat{J}(\hat{\theta}, r)$.*

Proof. The properties (16) and (17) follow from Theorem 1. To prove (24), we note that

$$\int_T^{T+\tau_0} \|e_{\text{obs}}(s)\|^2\,\mathrm{d}s < \varepsilon_{\text{obs}}^2 \tau_0 \leq \alpha_0 \varepsilon_\theta^2 \tag{25}$$

following (16) and (23). Meanwhile, (9), (16), and the PE condition (22) imply that

$$\begin{aligned}\int_T^{T+\tau_0} \|e_{\text{obs}}(s)\|^2\,\mathrm{d}s &= \int_T^{T+\tau_0} \big\|K(s)(\hat{\theta}(T) - \theta)\big\|^2\,\mathrm{d}s \\ &= (\hat{\theta}(T) - \theta)^\top \bigg(\int_T^{T+\tau_0} K(s)^\top K(s)\,\mathrm{d}s\bigg)(\hat{\theta}(T) - \theta) \\ &\geq \alpha_0 \|\hat{\theta}(T) - \theta\|^2,\end{aligned}$$

which, when compared with (25), yields (24). □

Remark 1. Following (10), the PE condition (22) holds if

$$\int_t^{t+\tau_0} \nabla_\theta \hat{f}(r(s))^\top \nabla_\theta \hat{f}(r(s))\,\mathrm{d}s - \alpha_0 I \geq 0 \quad \forall\, t \geq 0. \tag{26}$$

While (26) is more restrictive than (22), the former can be verified without observing the follower's action a (or even the parameter estimate $\hat{\theta}$).

Remark 2. We can see from the proof of Theorem 2 that, to obtain (24), the PE condition (22) (or (26)) only needs to hold at the time $t = T$ in (16). Therefore, to ensure (24) in practice, it suffices to enforce (22) (or (26)) whenever the switching signal λ_e in (12) is set to 0, instead of at all times.

V. Mismatch in parameterization

The results so far assumed that the unknown follower's strategy f satisfies the matching condition (3) for some unknown value θ in the parameter set Θ. In this section we show that, in the scenario without such perfect matching, the estimation and optimization dynamics introduced in Section III can still ensure an error bound in proportion to the size of the mismatch between $f(r)$ and $\hat{f}(\theta, r)$.

Assumption 4 (Mismatch)**.** The follower's strategy f is continuous, and there is a parameter value $\theta \in \Theta$ and a constant $\varepsilon_f \geq 0$ such that

$$\|\hat{f}(\theta, r) - f(r)\| \leq \varepsilon_f \qquad \forall\, r \in \mathcal{R}. \tag{27}$$

The upper bound ε_f is known to the leader, but the parameter value θ is unknown.

The following lemma establishes a bound for the portion of observation error due to mismatch in parameterization

$$e^f_{\text{obs}} := e_{\text{obs}} - K(r, a, \hat{\theta})(\hat{\theta} - \theta), \tag{28}$$

where $a = f(r)$, $\hat{a} = \hat{f}(\hat{\theta}, r)$, and the gain matrix $K(r, a, \hat{\theta})$ is defined by (10).

Lemma 4. *For every $r \in \mathcal{R}$, $\theta \in \Theta$ such that* (27) *holds, and $\hat{\theta} \in \Theta$, we have*

$$\|e^f_{\text{obs}}\| \leq \varepsilon^f_{\text{obs}} := \varepsilon_f \sqrt{1 + \kappa^2}, \tag{29}$$

where

$$\kappa := \max_{\bar{a} \in \operatorname{co} \hat{\mathcal{A}}} \|\nabla_a J(r, \bar{a})\|, \qquad \hat{\mathcal{A}} := \bigcup_{\theta \in \Theta,\, r \in \mathcal{R}} \varepsilon_f B(\hat{f}(\theta, r)).$$

Proof. Using similar arguments to those in the proof of Lemma 1, we can show that

$$e_{\text{obs}} = \begin{bmatrix} \int_0^1 \nabla_a J(r, \rho\hat{a} + (1-\rho)a)\,\mathrm{d}\rho \\ I \end{bmatrix} (\hat{a} - a),$$

and thus

$$e^f_{\text{obs}} = \begin{bmatrix} \int_0^1 \nabla_a J(r, \rho\hat{a} + (1-\rho)a)\,\mathrm{d}\rho \\ I \end{bmatrix} (\hat{f}(\theta, r) - a).$$

As $\hat{a} \in \hat{\mathcal{A}}$ and (27) implies that $a = f(r) \in \hat{\mathcal{A}}$, we have $\rho\hat{a} + (1-\rho)a \in \operatorname{co}\hat{\mathcal{A}}$ and thus $\|\nabla_a J(r, \rho\hat{a} + (1-\rho)a)\| \leq \kappa$. Then (29) follows from (27) and the definition of the matrix norm induced by the Euclidean norm. □

We now generalize Theorems 1 and 2 to the current scenario without perfect matching between $f(r)$ and $\hat{f}(\theta, r)$ for some $\theta \in \Theta$.

Theorem 3. *Under Assumptions 1, 2 and 4, if ε_{obs} and $\varepsilon'_{\text{obs}}$ in* (12) *are selected so that*

$$\varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > \varepsilon^f_{\text{obs}}, \tag{30}$$

$$\begin{aligned}\|e_{\mathrm{obs}}(s)\|^2 &= \|K(s)(\hat{\theta}(T)-\theta)+e^f_{\mathrm{obs}}(s)\|^2 \geq \left(1-\frac{\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0}}{\varepsilon_\theta\sqrt{\alpha_0}}\right)\|K(s)(\hat{\theta}(T)-\theta)\|^2+\left(1-\frac{\varepsilon_\theta\sqrt{\alpha_0}}{\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0}}\right)\|e^f_{\mathrm{obs}}(s)\|^2 \\ &= (\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0})\left(\frac{\|K(s)(\hat{\theta}(T)-\theta)\|^2}{\varepsilon_\theta\sqrt{\alpha_0}}-\frac{\|e^f_{\mathrm{obs}}(s)\|^2}{\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0}}\right).\end{aligned} \tag{31}$$

$$\begin{aligned}\int_T^{T+\tau_0}\|e_{\mathrm{obs}}(s)\|^2\,\mathrm{d}s &\geq (\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0})\left(\int_T^{T+\tau_0}\frac{\|K(s)(\hat{\theta}(T)-\theta)\|^2}{\varepsilon_\theta\sqrt{\alpha_0}}\,\mathrm{d}s-\int_T^{T+\tau_0}\frac{\|e^f_{\mathrm{obs}}(s)\|^2}{\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0}}\,\mathrm{d}s\right) \\ &\geq (\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0})\left((\hat{\theta}(T)-\theta)^\top\left(\int_T^{T+\tau_0}\frac{K(s)^\top K(s)}{\varepsilon_\theta\sqrt{\alpha_0}}\,\mathrm{d}s\right)(\hat{\theta}(T)-\theta)-\int_T^{T+\tau_0}\frac{\varepsilon^f_{\mathrm{obs}}}{\sqrt{\tau_0}}\,\mathrm{d}s\right) \\ &\geq (\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0})\left(\frac{\sqrt{\alpha_0}}{\varepsilon_\theta}\|\hat{\theta}(T)-\theta\|^2-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0}\right).\end{aligned} \tag{32}$$

then the estimation and optimization dynamics (11) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \geq 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} satisfy* (16)*.*
2) *The asymptotic FONC for optimality* (17) *holds for the estimated leader's cost $\hat{J}(\hat{\theta}, r)$.*

Proof. First, Lemma 3 still holds as the function f is continuous, and item 2) here is the same as item 2) of Theorem 1 because the optimization dynamics are the same after the adaptation of $\hat{\theta}$ stops. Then it remains to prove item 1) here, which will be done using similar arguments to those in Section III-A and the proof of item 1) of Theorem 1. Under the estimation dynamics (11) with (30) in (12), the estimation error $\hat{\theta}-\theta$ now satisfies

$$\begin{aligned}\frac{\mathrm{d}\|\hat{\theta}-\theta\|^2}{\mathrm{d}t} &= 2(\hat{\theta}-\theta)^\top\big[-\lambda_e K(r,a,\hat{\theta})^\top e_{\mathrm{obs}}\big]_{T_\Theta(\hat{\theta})} \\ &\leq 2(\hat{\theta}-\theta)^\top\big(-\lambda_e K(r,a,\hat{\theta})^\top e_{\mathrm{obs}}\big) \\ &= -2\lambda_e(e_{\mathrm{obs}}-e^f_{\mathrm{obs}})^\top e_{\mathrm{obs}} \\ &\leq -2\lambda_e(\|e_{\mathrm{obs}}\|-\|e^f_{\mathrm{obs}}\|)\|e_{\mathrm{obs}}\|,\end{aligned}$$

where the first inequality follows from the fact that $\theta-\hat{\theta} \in T_\Theta(\hat{\theta})$ and the first properties in (6) and (7), and the last equality follows from the definition of e^f_{obs} in (28). Note that if $\lambda_e = 0$ then $\mathrm{d}\|\hat{\theta}-\theta\|^2/\,\mathrm{d}t = 0$. Otherwise (12) implies that $\lambda_e = \lambda_\theta$; hence

$$\|e_{\mathrm{obs}}\| > \varepsilon'_{\mathrm{obs}} > \varepsilon^f_{\mathrm{obs}} \geq \|e^f_{\mathrm{obs}}\| \tag{33}$$

following also (29) and (30). Consequently, we have

$$\frac{\mathrm{d}\|\hat{\theta}-\theta\|^2}{\mathrm{d}t} \leq -2\lambda_\theta(\|e_{\mathrm{obs}}\|-\|e^f_{\mathrm{obs}}\|)\|e_{\mathrm{obs}}\| < 0.$$

Hence we conclude that the estimation dynamics (11) now ensures that

$$\frac{\mathrm{d}\|\hat{\theta}-\theta\|^2}{\mathrm{d}t} \leq -2\lambda_e(\|e_{\mathrm{obs}}\|-\|e^f_{\mathrm{obs}}\|)\|e_{\mathrm{obs}}\| \leq 0, \tag{34}$$

and the equality in the last inequality holds if and only if $\lambda_e = 0$. Therefore, $\|\hat{\theta}-\theta\|$ decreases monotonically, and the definition of λ_e in (12) implies that $\|\hat{\theta}-\theta\|$ does not stop approaching 0 as long as $\|e_{\mathrm{obs}}\| \geq \varepsilon_{\mathrm{obs}}$. As $\|\hat{\theta}-\theta\|$ is bounded from below by 0, the limit $\lim_{t\to\infty}\|\hat{\theta}(t)-\theta\|$ exists and is finite; hence

$$\lim_{t\to\infty}\int_0^t \lambda_e(s)(\|e_{\mathrm{obs}}(s)\|-\|e^f_{\mathrm{obs}}(s)\|)\|e_{\mathrm{obs}}(s)\|\,\mathrm{d}s \tag{35}$$

exists and is finite. Meanwhile, (11) and (12) imply that (16) holds if there is a time $T \geq 0$ for which (20) in the proof of Theorem 1 holds, which we will prove by contradiction. Assume that (20) does not hold for any $T \geq 0$. Then the arguments in the second step of the proof of Theorem 1 show that there is an unbounded time sequence $(s_k)_{k\in\mathbb{Z}_+}$ such that (21) holds for all $k \in \mathbb{Z}_+$. Consequently, (33) implies that

$$\begin{aligned}\int_{s_k}^{s_k+\delta} &\lambda_e(s)(\|e_{\mathrm{obs}}(s)\|-\|e^f_{\mathrm{obs}}(s)\|)\|e_{\mathrm{obs}}(s)\|\,\mathrm{d}s \\ &> \lambda_\theta(\varepsilon'_{\mathrm{obs}}-\varepsilon^f_{\mathrm{obs}})\,\varepsilon'_{\mathrm{obs}}\delta > 0 \qquad \forall\, k \in \mathbb{Z}_+,\end{aligned}$$

which, together with (34), contradicts the property that (35) exists and is finite. Hence there is a time $T \geq 0$ for which (20), and therefore (16), are both satisfied. □

Theorem 4. *Under Assumptions 1–4, given an arbitrary $\varepsilon_\theta > 2\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0/\alpha_0}$, if $\varepsilon_{\mathrm{obs}}$ and $\varepsilon'_{\mathrm{obs}}$ in* (12) *are selected so that*

$$\varepsilon_\theta\sqrt{\alpha_0/\tau_0}-\varepsilon^f_{\mathrm{obs}} \geq \varepsilon_{\mathrm{obs}} > \varepsilon'_{\mathrm{obs}} > \varepsilon^f_{\mathrm{obs}}, \tag{36}$$

then the estimation and optimization dynamics (11) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \geq 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} satisfy* (16) *and* (24)*.*
2) *The asymptotic FONC for optimality* (17) *holds for the estimated leader's cost $\hat{J}(\hat{\theta}, r)$.*

Proof. The properties (16) and (17) follow from Theorem 3. To prove (24), we note that

$$\int_T^{T+\tau_0}\|e_{\mathrm{obs}}(s)\|^2\,\mathrm{d}s < \varepsilon^2_{\mathrm{obs}}\tau_0 \leq (\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0})^2 \tag{37}$$

following (16) and (36), in which $\varepsilon_\theta\sqrt{\alpha_0}-\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0} > 0$ as $\varepsilon_\theta > 2\varepsilon^f_{\mathrm{obs}}\sqrt{\tau_0/\alpha_0}$. Meanwhile, (16), (28), and Young's inequality imply that (31) above holds for all $s \geq T$, which, combined with the PE condition (22) and (29), yields (32) above. Comparing (32) with (37), we see that (24) holds. □

Note that the matching condition (3) is a special case of the condition (27) with $\varepsilon_f = 0$. Therefore, Theorems 3 and 4 generalize Theorems 1 and 2 to the scenario where ε_f is not necessarily 0, respectively.

VI. Unobservable follower's actions

In this section, we consider the case where the follower's actions cannot be observed. By making a few adjustments to the proposed adaptive learning approach, we are able to ensure similar convergence properties to those in Section IV and V, using only past observations of leader's cost.

We start by assuming that the leader's cost given the follower's strategy, which is an unknown function $J(r, f(r))$ of only its own action $r \in \mathcal{R}$, belongs to a known parametric family of functions $\{r \mapsto \hat{J}(\theta, r) : \theta \in \Theta \subset \mathbb{R}^{n_\theta}\}$, namely, there is an unknown parameter value $\theta \in \Theta$ such that

$$J(r, f(r)) = \hat{J}(\theta, r) \qquad \forall\, r \in \mathcal{R}. \tag{38}$$

The regularity and affinity conditions in Assumptions 1 and 2 are assumed to hold with $\hat{J}$ in place of $\hat{f}$.

The observation error is now

$$e_{\text{obs}} := \hat{J}(\hat{\theta}, r) - J(r, a) \tag{39}$$

and satisfies

$$e_{\text{obs}} = \nabla_\theta \hat{J}(r)(\hat{\theta} - \theta) \qquad \forall\, r \in \mathcal{R}, \theta, \hat{\theta} \in \Theta,$$

where $\nabla_\theta \hat{J}(r)$ is the Jacobian matrix of $\hat{J}(\theta, r)$ with respect to θ which is independent of θ. Consequently, the estimation dynamics are adjusted to

$$\dot{\hat{\theta}} = \left[-\lambda_e \nabla_\theta \hat{J}(r)^\top e_{\text{obs}}\right]_{T_\Theta(\hat{\theta})} \tag{40}$$

with the same switching signal λ_e defined by (12). Using similar arguments to those in Section III-A and the proofs of Theorems 1 and 2, we obtain the convergence results below.

Theorem 5. *Under Assumptions 1 and 2 with $\hat{J}$ in place of $\hat{f}$, given arbitrary $\varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > 0$ in* (12)*, the estimation and optimization dynamics* (40) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \geq 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} in* (39) *satisfy* (16).
2) *The asymptotic FONC for optimality* (17) *holds for the estimated leader's cost $\hat{J}(\hat{\theta}, r)$.*
3) *Suppose that*

$$\int_t^{t+\tau_0} \nabla_\theta \hat{J}(r(s))^\top \nabla_\theta \hat{J}(r(s))\,\mathrm{d}s - \alpha_0 I \geq 0 \qquad \forall\, t \geq 0 \tag{41}$$

for some constants $\tau_0, \alpha_0 > 0$. Given an arbitrary $\varepsilon_\theta > 0$, if ε_{obs} and $\varepsilon'_{\text{obs}}$ in (12) *are selected so that* (23) *holds, then the parameter estimate $\hat{\theta}$ in* (39) *also satisfies* (24).

In the scenario without the perfect matching (38) between $J(r, f(r))$ and $\hat{J}(\theta, r)$ for some $\theta \in \Theta$, the adjusted approach can still ensure an error bound in proportion to the size of the mismatch between the two functions.

Assumption 5 (Mismatch in cost parameterization)**.** The follower's strategy f is continuous, and there is a parameter value $\theta \in \Theta$ and a constant $\varepsilon_J \geq 0$ such that

$$\|\hat{J}(\theta, r) - J(r, f(r))\| \leq \varepsilon_J \qquad \forall\, r \in \mathcal{R}. \tag{42}$$

The upper bound ε_J is known to the leader, but the parameter value θ is unknown.

The portion of observation error due to mismatch in parameterization is now

$$e^J_{\text{obs}} := e_{\text{obs}} - \nabla_\theta \hat{J}(r)(\hat{\theta} - \theta)$$

and satisfies

$$\|e^J_{\text{obs}}\| = \|\hat{J}(\theta, r) - J(r, f(r))\| \leq \varepsilon_J$$

for every $r \in \mathcal{R}$, $\theta \in \Theta$ such that (42) holds, and $\hat{\theta} \in \Theta$. Using similar arguments to those in the proofs of Theorems 3 and 4, we obtain the following generalization of Theorem 5 to the scenario where ε_J is not necessarily 0.

Theorem 6. *Under Assumptions 1, 2, and 5 with $\hat{J}$ in place of $\hat{f}$, if ε_{obs} and $\varepsilon'_{\text{obs}}$ in* (12) *are selected so that*

$$\varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > \varepsilon_J,$$

then the estimation and optimization dynamics (40) *and* (15) *ensure that the following properties hold:*

1) *There is a time $T \geq 0$ after which the parameter estimate $\hat{\theta}$ and the observation error e_{obs} in* (39) *satisfy* (16).
2) *The asymptotic FONC for optimality* (17) *holds for the estimated leader's cost $\hat{J}(\hat{\theta}, r)$.*
3) *Suppose that* (41) *holds for some constants $\tau_0, \alpha_0 > 0$. Given an arbitrary $\varepsilon_\theta > 2\varepsilon_J\sqrt{\tau_0/\alpha_0}$, if ε_{obs} and $\varepsilon'_{\text{obs}}$ in* (12) *are selected so that*

$$\varepsilon_\theta\sqrt{\alpha_0/\tau_0} - \varepsilon_J \geq \varepsilon_{\text{obs}} > \varepsilon'_{\text{obs}} > \varepsilon_J,$$

then the parameter estimate $\hat{\theta}$ in (39) *also satisfies* (24).

Note that, while the convergence properties in Theorems 5 and 6 are similar to those in Theorems 1–4, it usually takes a considerably longer time for the adaptation of $\hat{\theta}$ to terminate in the current case, due to the low dimension of the observable.

VII. Simulation examples

In this section, we illustrate the proposed approach and convergence results through simulation examples motivated by link-flooding DDoS attacks such as the Crossfire attack [6].

Consider a communication network consisting of L parallel links connecting a source to a destination. The set of links is denoted by $\mathcal{L} := \{1, \dots, L\}$. Suppose that a router (the leader) distributes a total of R units of legitimate traffic among the parallel links, and an attacker (the follower) disrupts communication by injecting superfluous traffic with a budget of A units. The router's action is represented by an L-vector $r \in \mathcal{R} := \{r \in \mathbb{R}^L_+ : \sum_{l \in \mathcal{L}} r_l = R\}$ of the *desired* legitimate traffic on each link, and the attacker's action is represented by an L-vector $a \in \mathcal{A} := \{a \in \mathbb{R}^L_+ : \sum_{l \in \mathcal{L}} a_l = A\}$ of the attack traffic. Each link $l \in \mathcal{L}$ is characterized by a fixed capacity

$c_0 > 0$ that limits the total traffic on l. When $r_l + a_l > c_0$, the *actual* legitimate traffic on l is decreased to

$$u_l := \min\{r_l,\, \max\{c_0 - a_l,\, 0\}\}$$

to meet the capacity. The router aims to maximize the total actual legitimate traffic; hence its cost is defined by

$$J(r,a) := -\sum_{l\in\mathcal{L}} u_l.$$

We start by considering the case where the attacker aims to minimize the total actual legitimate traffic; hence its cost is defined by

$$H(r,a) := \sum_{l\in\mathcal{L}} u_l = -J(r,a). \tag{43}$$

Clearly, neither the router nor the attacker has an incentive to assign more traffic on a link than the capacity c_0. Hence we assume that $r_l, a_l \le c_0$ for all $l \in \mathcal{L}$. More details about the network and attack models can be found in [31]. For most nontrivial cases, the game defined by $(\mathcal{R}, \mathcal{A}, J, H)$ has no (pure) Nash equilibrium. On the other hand, in [31, Cor. 15.2] it was shown that there is a Stackelberg equilibrium action r^* for the router defined by $r_l^* := R/L$ for all $l \in \mathcal{L}$.

If the router knew that the attacker's cost function was indeed given by (43), it could play the Stackelberg equilibrium action r^*. However, we are interested in the more challenging scenario where it does not, and instead adopts the proposed adaptive learning approach to optimize its action. We could use any sufficiently rich parametric family of functions $\{r \mapsto \hat{f}(\theta, r) : \theta \in \Theta\}$, but the structure of the problem (and the results in [31]) enables us to select a parameterization that is accurate for a reasonably small number of parameters: Following [31, Cor. 15.1], the attacker's action $a = f(r)$ depends on the order of the desired legitimate traffic r_l on each link $l \in \mathcal{L}$, which can be seen as a function of the ratio r_{l_1}/r_{l_2} of desired legitimate traffic on each pair of links $(l_1, l_2) \in \mathcal{L} \times \mathcal{L}$. Therefore, we partition the router's action set $\mathcal{R}$ according to these ratios. Specially, we take $\bar{n}_\theta$ partition cells for each of the $\bar{L} := \binom{L}{2}$ pairs of links, and the indicator function for each partition cell is defined by

$$b_{j_1,\dots,j_{\bar{L}}}(r) := \begin{cases} 1 & \text{if } \forall\, i \in \{1,\dots,\bar{L}\},\ \text{either} \\ & \left\lceil \frac{\bar{n}_\theta \arctan(r_{l_{i1}}/r_{l_{i2}})}{\pi/2} \right\rceil = j_i, \\ & \text{or } r_{l_{i1}} = 0,\ r_{l_{i2}} \neq 0,\ \text{and } j_i = 1; \\ 0 & \text{otherwise} \end{cases}$$

for $j_1, \dots, j_{\bar{L}} \in \{1, \dots, \bar{n}_\theta\}$, where $(l_{i1}, l_{i2}) \in \mathcal{L} \times \mathcal{L}$ is the i-th pair of links, and the arctangent function and division by $\pi/2$ are used to normalize the ratios to the unit interval $[0, 1]$. For $L = 2$ and $L = 3$, this partition with $\bar{n}_\theta = 10$ is illustrated by dashed lines in Fig. 3(a), 3(b), 5(a), and 5(b) below. Using this partition, we estimate the attacker's strategy $f = (f_1, \dots, f_L)$ using the parametric function $\hat{f} = (\hat{f}_1, \dots, \hat{f}_L)$ defined by

$$\hat{f}_l(\theta, r) := \sum_{j_1=1}^{\bar{n}_\theta} \cdots \sum_{j_{\bar{L}}=1}^{\bar{n}_\theta} \theta_{l,j_1,\dots,j_{\bar{L}}} b_{j_1,\dots,j_{\bar{L}}}(r), \quad l \in \mathcal{L}. \tag{44}$$

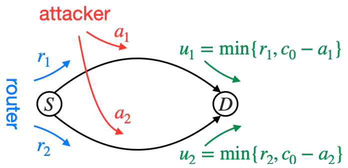

attacker
a_1
r_1
$u_1 = \min\{r_1, c_0 - a_1\}$
router
S
D
a_2
r_2
$u_2 = \min\{r_2, c_0 - a_2\}$

Fig. 1. A network with one source S, one destination D, and two parallel links (assuming that $a_1, a_2 \le c_0$).

Then the dimension of parameter vector is given by $n_\theta = L\bar{n}_\theta^{\bar{L}}$. For large networks, one can adopt clustering techniques [32, Sec. 16] to reduce the number of partition cells and thus n_θ, or estimate the attacker's strategy using neural networks; see [31] for some preliminary results based on the latter approach.

Remark 3. In this model, the functions J and f and the map $r \mapsto \hat{f}(\theta, r)$ actually violate the smoothness conditions in Assumptions 1 and 4, as they are only piecewise continuously differentiable and the last two are only piecewise continuous. However, these conditions are only needed to ensure that the estimation and optimization dynamics (11) and (15) are well-defined and continuous everywhere. In practice, the set of non-differentiable points has measure zero and does not affect the simulation. The set of discontinuous points also has measure zero but can lead to situations where, near this set, the optimization cost does not decrease along the steepest descent direction even for small step sizes. In that case, we project the gradient descent direction along the hyperplanes defining the boundary of the current partition cell; from a theoretical viewpoint, this means following a Filippov solution [37] to the optimization dynamics (15).

In the following, we simulate the estimation and optimization dynamics (11) and (15) for networks with $L = 2$ and $L = 3$ parallel links. In these examples, we set $\varepsilon_{\text{obs}} = 0.02$, $\varepsilon'_{\text{obs}} = 0.01$, $\lambda_\theta = 0.2$, and $\lambda_r = 0.1$, and use randomly generated initial values of the parameter estimate $\hat{\theta}$ and the routers' action r.

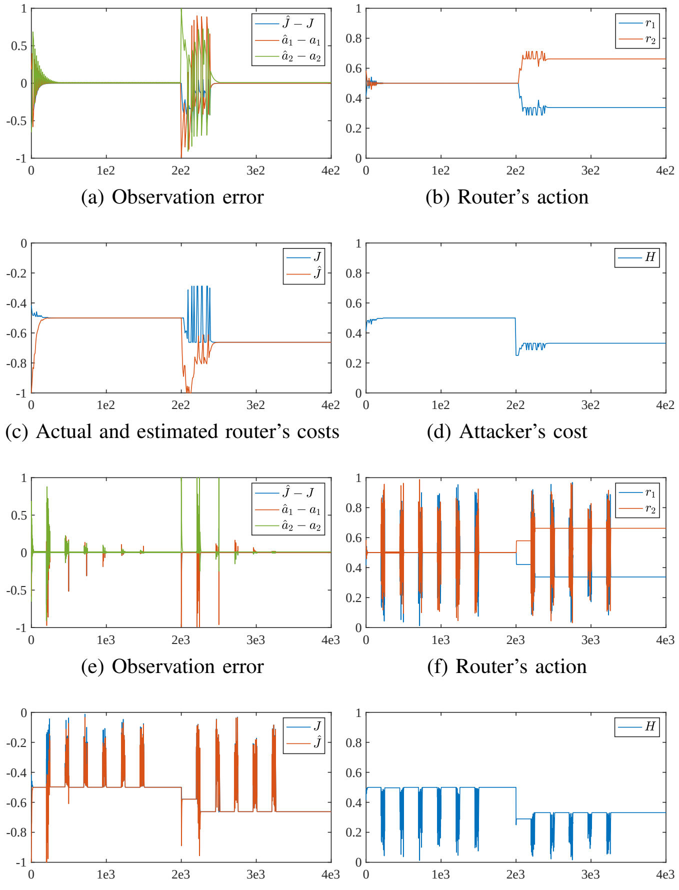
A. A network with two parallel links

Consider the network with $L = 2$ parallel links in Fig. 1, link capacity $c_0 = 1$, total desired legitimated traffic $R = Lc_0/2 = 1$, and attack budget $A = \lceil Lc_0/2 \rceil = 1$. We set $\bar{n}_\theta = 10$ in (44); hence the dimension of parameter vector is given by $n_\theta = 20$. Following [31, Cor. 15.1], the attacker's best response to a router's action r is to set $a_l = 1$ on the link l with the larger r_l. Hence the actual parameter value θ in (3), written in the tensor form in (44), is given by

$$\theta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

As established in [31, Cor. 15.2], the Stackelberg equilibrium action for the router is given by $r^* = (1/2, 1/2)$. The simulation results are plotted in Fig. 2 and 3 below, with their main properties summarized as follows:

First, in Fig. 2(a)–2(d), the PE condition is not enforced. In the first half of the simulation, the observation error e_{obs} converges to 0 and the router's action r converges to the Stackelberg equilibrium action r^*, despite the fact that the



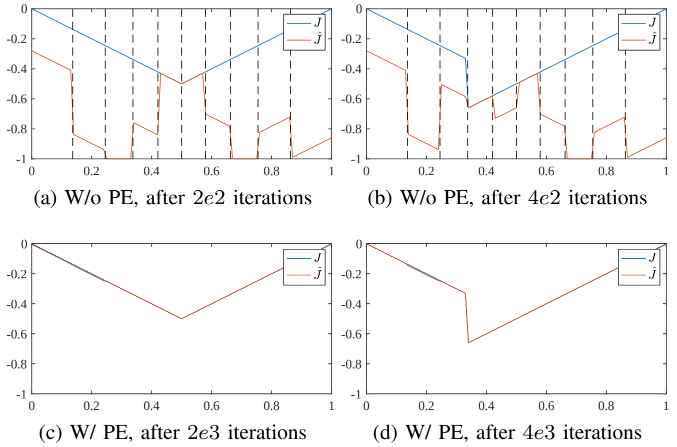
(a) Observation error (b) Router's action

(c) Actual and estimated router's costs (d) Attacker's cost

(e) Observation error (f) Router's action

(g) Actual and estimated router's costs (h) Attacker's cost

Fig. 2. Simulation results for $L = 2$ (horizontal axis: number of iterations). (a)–(d): The case w/o PE. (e)–(h): The case w/ PE. For both cases, in the first half of the simulation, the observation error e_{obs} converges to 0, the router's action r converges to the Stackelberg equilibrium action $r^* = (1/2, 1/2)$, the actual and estimated router's costs J and $\hat{J}$ converge to $-1/2$, and the attacker's cost H converges to $1/2$; in the second half of the simulation, the attacker switches to the new cost function $\bar{H}$, the observation error e_{obs} converges again to 0, the router's action r converges to an ε Stackelberg action near $\bar{r}^* = (1/3, 2/3)$, the actual and estimated router's costs J and $\hat{J}$ converge to $-2/3$, while the attacker's cost H converges to $1/3$.



(a) W/o PE, after $2e2$ iterations (b) W/o PE, after $4e2$ iterations

(c) W/ PE, after $2e3$ iterations (d) W/ PE, after $4e3$ iterations

Fig. 3. Actual and estimated router's cost functions $r_1 \mapsto J(r, f(r))$ and $r_1 \mapsto \hat{J}(\hat{\theta}(T), r)$ for $L = 2$. (a) and (b): W/o PE, the estimation is only accurate near the asymptotic router's action. (c) and (d): W/ PE, the estimation is accurate everywhere.

parameter estimate $\hat{\theta}$ does not converge to the actual value θ as shown by Fig. 3(a). These results illustrate Theorem 1.

Second, in Fig. 2(e)–2(h), we enforce the PE condition by monitoring the observation error e_{obs} (see also Remark 2). Whenever $\|e_{\mathrm{obs}}\|$ has been continuously smaller than a threshold 0.05 for 200 iterations, we set the router's action r for each of the next 50 iterations to a randomly generated L-vector of sum R. In the first half of the simulation, in addition to the convergence of e_{obs} and r, the parameter estimate $\hat{\theta}$ also converges to the actual value θ as shown by Fig. 3(c). These results illustrate Theorem 2.

Finally, we also consider the scenario where, after half of the simulation, the attacker starts to focus more on disrupting link 1 by switching to a new cost function defined by

$$\bar{H}(r, a) := u_1 + u_2/2.$$

Following [31, Cor. 15.1], the attacker's best response to a router's action r is then to set $a_l = 1$ on the link $l \in \{1, 2\}$ that corresponds to the larger one in $\{r_1, r_2/2\}$. In this scenario, the parametric function $\hat{f}$ defined by (44) cannot match the attacker's strategy f perfectly. Nevertheless, their mismatch is bounded by (27) with the parameter value

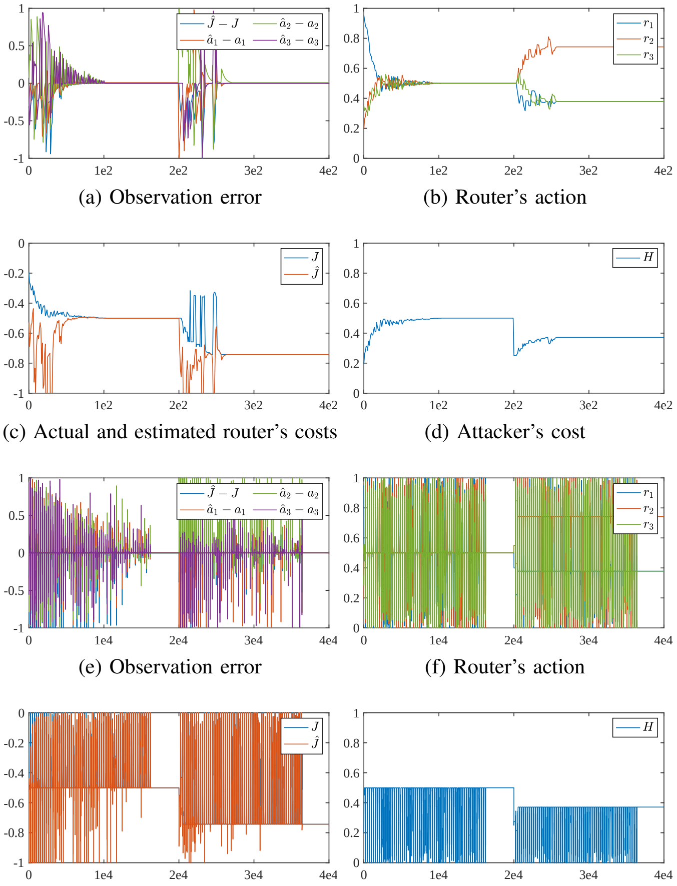
$$\theta = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the constant $\varepsilon_f = 1$. The results in [31, Th. 15.3] allow us to conclude that there is no Stackelberg equilibrium action for the router as defined by Definition 1 for the non-zero-sum game $(\mathcal{R}, \mathcal{A}, J, \bar{H})$; however, there are ε Stackelberg actions for the router near $\bar{r}^* := (1/3, 2/3)$ for a sufficiently small $\varepsilon > 0$. The simulation results in Fig. 2 show that the proposed approach is able to identify this switch in the attack, as the observation error e_{obs} converges again to 0 and the router's action r converges to an ε Stackelberg action near $\bar{r}^*$ for both cases with and without enforcing the PE condition; when the PE condition is enforced, the parameter estimate $\hat{\theta}$ also converges to the new parameter value θ as shown by Fig. 3(d). These results illustrate Theorems 3 and 4.

B. A network with three parallel links

Consider a network with $L = 3$ parallel links, link capacity $c_0 = 1$, total desired legitimated traffic $R = Lc_0/2 = 1.5$, and attack budget $A = \lceil Lc_0/2 \rceil = 2$. We set $\bar{n}_\theta = 10$ in (44); hence the dimension of parameter vector is given by $n_\theta = 3000$. Following [31, Cor. 15.1], the attacker's best response to a router action r is to set $a_{l_1} = a_{l_2} = 1$ on the two links l_1 and l_2 with the two largest r_l (see also Fig. 5(a)). Hence the actual parameter value θ in (3), written in the tensor form in (44), is given by

$$\begin{cases} \theta_{1,j_1,j_2,j_3} = 1 & \text{if } j_1 \geq 6 \text{ or } j_2 \geq 6, \\ \theta_{2,j_1,j_2,j_3} = 1 & \text{if } j_1 \leq 5 \text{ or } j_3 \geq 6, \\ \theta_{3,j_1,j_2,j_3} = 1 & \text{if } j_2 \leq 5 \text{ or } j_3 \leq 5, \\ \theta_{l,j_1,j_2,j_3} = 0 & \text{otherwise.} \end{cases}$$



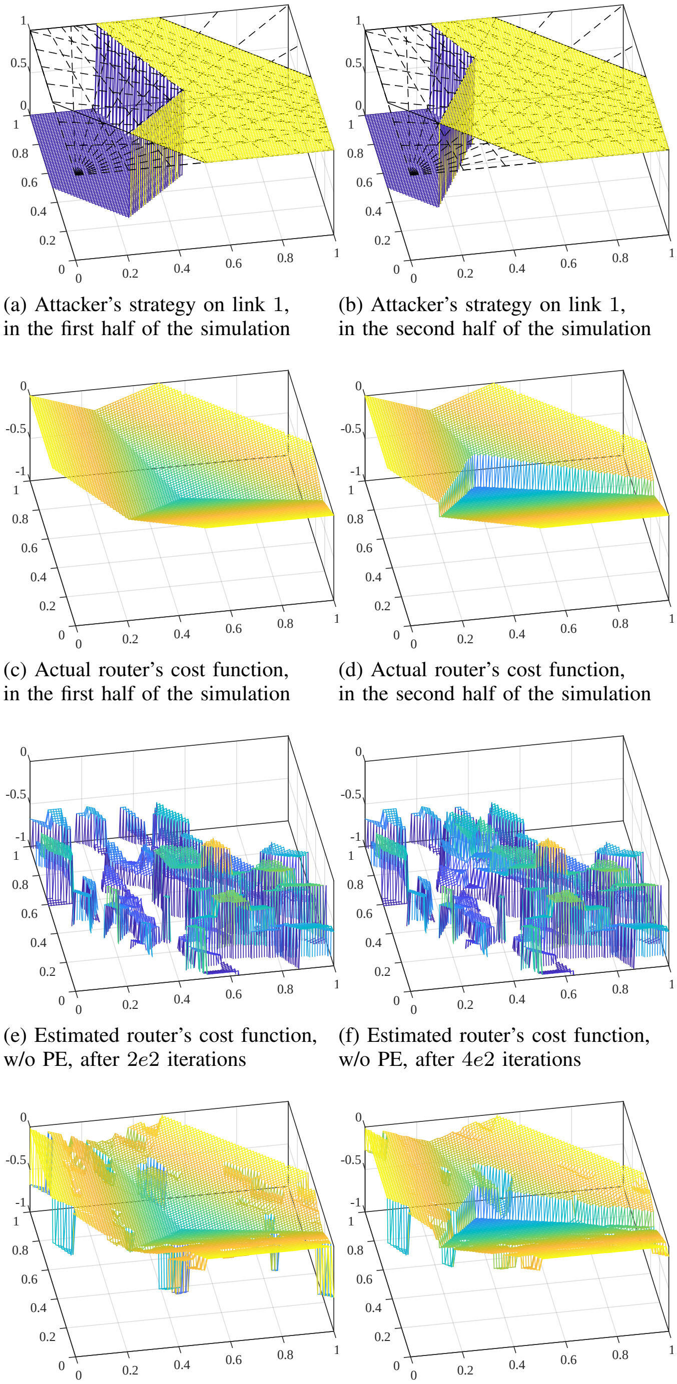
(a) Observation error (b) Router's action

(c) Actual and estimated router's costs (d) Attacker's cost

(e) Observation error (f) Router's action

(g) Actual and estimated router's costs (h) Attacker's cost

Fig. 4. Simulation results for $L = 3$ (horizontal axis: number of iterations). (a)–(d): The case w/o PE. (e)–(h): The case w/ PE. For both cases, in the first half of the simulation, the observation error e_{obs} converges to 0, the router's action r converges to the Stackelberg equilibrium action $r^* = (1/2, 1/2, 1/2)$, the actual and estimated router's costs J and $\hat{J}$ converge to $-1/2$, and the attacker's cost H converges to $1/2$; in the second half of the simulation, the attacker switches to the new cost function $\bar{H}$, the observation error e_{obs} converges again to 0, the router's action r converges to an ε Stackelberg action near $\bar{r}^* = (3/8, 3/4, 3/8)$, the actual and estimated router's costs J and $\hat{J}$ converge to $-3/4$, while the attacker's cost H converges to $3/8$.


(a) Attacker's strategy on link 1, in the first half of the simulation (b) Attacker's strategy on link 1, in the second half of the simulation

(c) Actual router's cost function, in the first half of the simulation (d) Actual router's cost function, in the second half of the simulation

(e) Estimated router's cost function, w/o PE, after $2e2$ iterations (f) Estimated router's cost function, w/o PE, after $4e2$ iterations

(g) Estimated router's cost function, w/ PE, after $2e4$ iterations (h) Estimated router's cost function, w/ PE, after $4e4$ iterations

Fig. 5. Attacker's strategy $(r_1, r_2) \mapsto f_1(r)$ and actual and estimated router's cost functions $(r_1, r_2) \mapsto J(r, f(r))$ and $(r_1, r_2) \mapsto \hat{J}(\hat{\theta}(T), r)$ for $L = 3$. (e) and (f): W/o PE, the estimation is only accurate near the asymptotic router's action. (g) and (h): W/ PE, the estimation is accurate nearly everywhere.

As established in [31, Cor. 15.2], the Stackelberg equilibrium action for the router is given by $r^* = (1/2, 1/2, 1/2)$. The simulation results are plotted in Fig. 4 and 5, with their main properties summarized as follows:

First, similar to the previous example with $L = 2$, in the first half of the simulation results in Fig. 4, the observation error e_{obs} converges to 0 and the router's action r converges to the Stackelberg equilibrium action r^* for both cases with and without enforcing the PE condition, despite the fact that the parameter estimation is only quite accurate when the PE condition is enforced as shown by Fig. 5(e) and 5(g). These results illustrate Theorems 1 and 2.

Second, we also consider the scenario where, after half of the simulation, the attacker starts to focus more on disrupting links 1 and 3 by switching to a new cost function defined by

$$\bar{H}(r, a) := u_1 + u_2/2 + u_3.$$

Following [31, Cor. 15.1], the attacker's best response to a router's action r is then to set $a_{l_1} = a_{l_2} = 1$ on the two link $l_1, l_2 \in \{1,\ 2,\ 3\}$ that correspond to the two largest ones in $\{r_1,\ r_2/2,\ r_3\}$. In this scenario, the parametric function $\hat{f}$ defined by (44) cannot match the attacker's strategy f perfectly. Nevertheless, their mismatch is bounded by (27) with the parameter value θ given by

$$\begin{cases} \theta_{1,j_1,j_2,j_3} = 1 & \text{if } j_1 \geq 4 \text{ or } j_2 \geq 6, \\ \theta_{2,j_1,j_2,j_3} = 1 & \text{if } j_1 \leq 3 \text{ or } j_3 \geq 8, \\ \theta_{3,j_1,j_2,j_3} = 1 & \text{if } j_2 \leq 5 \text{ or } j_3 \leq 7, \\ \theta_{l,j_1,j_2,j_3} = 0 & \text{otherwise.} \end{cases}$$

and the constant $\varepsilon_f = 1$. The results in [31, Th. 15.3] allow us to conclude that there is no Stackelberg equilibrium action for the router as defined by Definition 1 for the non-zero-sum game $(\mathcal{R}, \mathcal{A}, J, \bar{H})$; however, there are ε Stackelberg actions for the router near $\bar{r}^* := (3/8, 3/4, 3/8)$ for a sufficiently small $\varepsilon > 0$. The simulation results in Fig. 4 show that the proposed approach is able to identify this switch in the attack, as the observation error e_{obs} converges again to 0 and the router's action r converges to an ε Stackelberg action near $\bar{r}^*$ for both cases with and without enforcing the PE condition; when the PE condition is enforced, the parameter estimation is also quite accurate as shown by Fig. 5(h). These results illustrate Theorems 3 and 4.

Note that, in the second half of all examples, the mismatch between $f(r)$ and $\hat{f}(\theta, r)$ is bounded by (27) with the constant $\varepsilon_f = 1$. Hence the lower bound in (30) and (36) are given by $\varepsilon^f_{\text{obs}} = \sqrt{3}$ for $L = 2$ and $\varepsilon^f_{\text{obs}} = 2$ for $L = 3$. However, in the simulation results, the convergence properties in Theorems 3 and 4 are achieved with the constant ε_{obs} in (11) set to a much smaller value 0.02, which indicates that the theoretical results are fairly conservative.

VIII. Conclusion

This paper studied the problem of solving two-player Stackelberg games with incomplete information about the follower. An adaptive learning approach was proposed for estimating the follower's strategy using a parametric family of functions based on past observations of follower's action and leader's cost, and simultaneously optimizing the leader's action for its estimated cost. Our approach ensured that a preselected, arbitrarily small error bound could be achieved in finite time for the estimation error of observables (and the parameter estimation error as well if a PE condition was enforced), and the FONC for optimality held asymptotically in time for the estimated leader's cost. Moreover, it was shown that these convergence properties were robust with respect to a bounded mismatch between the actual follower's strategy and the parametric family of follower's strategies assumed by the leader. The approach and convergence results were also extended to a scenario with unobservable follower's action, and were illustrated through simulation examples motivated by link-flooding DDoS attacks.

We focused on Stackelberg games with one leader and one follower in this paper, but plan to extend the proposed adaptive learning approach to games with multiple leaders and followers in future research. Such results will be useful for understanding and developing systems with information asymmetry among decentralized agents, for example, designing mitigation measures against DDoS attacks on communication networks with decentralized routers. Other future research directions include adopting neural networks for efficient estimation in high dimensional problems (some preliminary results can be found in [31]), and incorporating more sophisticated optimization methods such as simulated annealing [32, Sec. 10] to ensure a globally optimal leader's action. On applications to network security, we plan to adapt the proposed approach for networks with more complex topology and time-varying topology.

Appendix A
Projected dynamical systems

Here we provide some preliminaries on existence, boundedness, and convergence of solutions for the *projected dynamical system*

$$\dot{x} = [g(x)]_{T_{\mathcal{S}}(x)} \tag{45}$$

defined by a continuous function $g : \mathcal{S} \to \mathbb{R}^n$ on a compact convex set $\mathcal{S} \subset \mathbb{R}^n$. Analyzing solutions to (45) is difficult as its right-hand side is only defined over the compact set $\mathcal{S}$ and may be discontinuous in the boundary $\partial\mathcal{S}$ due to projection. Therefore, we consider the notion of *viable Carathéodory solution* [34] to (45) over a time interval $L \subset \mathbb{R}_+$, which is an absolutely continuous function $x : L \to \mathcal{S}$ such that (45) holds almost everywhere in L; in particular, it requires that $x(t) \in \mathcal{S}$ for all $t \in L$. The following lemma establishes existence of such solutions for projected dynamical systems:

Lemma 5. *For each $x_0 \in \mathcal{S}$, there exists a solution x to* (45) *over $\mathbb{R}_+$ with $x(0) = x_0$.*

Proof. On the compact set $\mathcal{S}$, the continuous function g is bounded and thus a Marchaud map [34, Def. 2.2.4, p. 62]. Then Lemma 5 follows from [34, Th. 10.1.1, p. 354]. □

Next, we establish an invariance theorem for (45) when the function g is defined by gradient descent.

Proposition 1. *If the function g in* (45) *satisfies*

$$g(x) = -\nabla V(x)^\top \qquad \forall\, x \in \mathcal{S} \tag{46}$$

for some function $V : \mathcal{S} \to \mathbb{R}$, then every solution x to (45) *satisfies*

$$\lim_{t\to\infty} [g(x(t))]_{T_{\mathcal{S}}(x(t))} = 0. \tag{47}$$

To establish Proposition 1, we extend the projected differential equation (45) to the differential inclusion

$$\dot{x} \in G(x), \tag{48}$$

where $G : \mathcal{S} \rightrightarrows \mathbb{R}^n$ is a set-valued map defined by

$$\begin{aligned} G(x) := \{&g(x) - v : v \in N_{\mathcal{S}}(x)\} \\ &\cap \big\|g(x) - [g(x)]_{T_{\mathcal{S}}(x)}\big\| B(g(x)). \end{aligned} \tag{49}$$

This extension is inspired by similar ones from [38] and [34, p. 354], but is specifically designed to simplified the proof of Proposition 1. As $[g(x)]_{T_{\mathcal{S}}(x)} \in G(x)$ for all $x \in \mathcal{S}$, a solution to (45) is also a solution to (48). We prove Proposition 1 by applying an invariance theorem for differential inclusions to (48), which requires the following continuity property.

Lemma 6. *The set-valued map G defined by* (49) *is upper semicontinuous on $\mathcal{S}$.*

Proof. The set-valued map $x \mapsto T_{\mathcal{S}}(x)$ is lower semicontinuous on the compact convex set $\mathcal{S}$ [34, Th. 5.1.7, p. 162]. Also, as g is continuous on $\mathcal{S}$, the map $(x, v) \mapsto \|g(x) - v\|$ is continuous on $\mathcal{S} \times \mathbb{R}^n$. Hence the map $x \mapsto \inf_{v \in T_{\mathcal{S}}(x)} \|g(x) - v\| = \|g(x) - [g(x)]_{T_{\mathcal{S}}(x)}\|$ is upper semicontinuous on $\mathcal{S}$ [34, Th. 2.1.6, p. 59]. Moreover, the map $x \mapsto \{g(x) - v : v \in N_{\mathcal{S}}(x)\}$ is closed. Hence the map G is upper semicontinuous on $\mathcal{S}$ [34, Cor. 2.2.3, p. 61]. □

Proof of Proposition 1. Suppose that

$$g(x)^\top z \geq \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \qquad \forall\, x \in \mathcal{S}, z \in G(x). \tag{50}$$

Then the function V satisfies

$$\nabla V(x)\, z \leq -\big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \qquad \forall\, x \in \mathcal{S}, z \in G(z).$$

Note that the set-valued maps G is upper semicontinuous, and for each $x \in \mathcal{S}$, the set $G(x)$ is nonempty, compact, and convex. Hence the invariance theorem [39, Th. 2.11] implies that every solution to (48), including every solution to (45), converges to the largest invariant subset of $\big\{x \in \mathcal{S} : \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\| = 0\big\}$. Then (47) holds as g is continuous on the compact set $\mathcal{S}$.

It remains to show that (50) holds. Consider arbitrary $x \in \mathcal{S}$ and $z \in G(z)$. First, we have

$$\big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 - [g(x)]_{T_{\mathcal{S}}(x)}^\top z = [g(x)]_{T_{\mathcal{S}}(x)}^\top (g(x) - z) \leq 0,$$

where the equality follows from the second property in (6), and the inequality follows from the fact that $g(x) - z \in N_{\mathcal{S}}(x)$ and the first property in (7). Hence

$$\begin{aligned}\|z\|^2 &\geq \|z\|^2 - 2[g(x)]_{T_{\mathcal{S}}(x)}^\top z + 2\big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \\ &= \big\|z - [g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 + \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \\ &\geq \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2.\end{aligned}$$

Next, we have

$$\begin{aligned}\|z - g(x)\|^2 &\leq \big\|g(x) - [g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \\ &= \|g(x)\|^2 - 2[g(x)]_{T_{\mathcal{S}}(x)}^\top g(x) + \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \\ &= \|g(x)\|^2 - \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2,\end{aligned}$$

where the inequality follows from the fact that $z \in \|g(x) - [g(x)]_{T_{\mathcal{S}}(x)}\| B(g(x))$, and the last equality follows from the second property in (6). Hence

$$\begin{aligned}2g(x)^\top z &= \|z\|^2 + \|g(x)\|^2 - \|z - g(x)\|^2 \\ &\geq \|z\|^2 + \big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2 \geq 2\big\|[g(x)]_{T_{\mathcal{S}}(x)}\big\|^2,\end{aligned}$$

that is, (50) holds. □

Appendix B
Proof of Lemma 3

Given an initial value $(\hat{\theta}_0, r_0) \in \Theta \times \mathcal{R}$, we construct a solution to (11) and (15) over $\mathbb{R}_+$ recursively. Our procedure assumes that $\|e_{\text{obs}}(0)\| > \varepsilon'_{\text{obs}}$; hence $\lambda_e(0) = \lambda_\theta$ due to (12). If $\|e_{\text{obs}}(0)\| \leq \varepsilon'_{\text{obs}}$, a solution can be constructed using the same procedure while starting with Step 2.

Step 1: Consider (11) and (15) with $\lambda_e \equiv \lambda_\theta$, namely,

$$\begin{aligned}\dot{\hat{\theta}} &= \big[-\lambda_\theta K(r, f(r), \hat{\theta})^\top K(r, f(r), \hat{\theta})(\hat{\theta} - \theta)\big]_{T_\Theta(\hat{\theta})}, \\ \dot{r} &= \big[-\lambda_r \nabla_r \hat{J}(\hat{\theta}, r)^\top\big]_{T_{\mathcal{R}}(r)},\end{aligned} \tag{51}$$

which can be modeled as a projected dynamical system (45) in Appendix A with the state $x := (\hat{\theta}, r)$ and the set $\mathcal{S} := \Theta \times \mathcal{R}$; in particular, the resulting function g in (45) is continuous following (3) and Assumption 1. Then Lemma 5 ensures that there exists a solution $(\hat{\theta}_1, r_1)$ to (51) over $\mathbb{R}_+$ with $(\hat{\theta}_1(0), r_1(0)) = (\hat{\theta}_0, r_0)$. Consider the corresponding observation error $e_{\text{obs},1}$ and switching signal $\lambda_{e,1}$ defined by (8) and (12), respectively, and let

$$t_1 := \inf\{t > 0 : \|e_{\text{obs},1}(t)\| \leq \varepsilon'_{\text{obs}}\}.$$

Then $(\hat{\theta}_1, r_1)$ is a solution to (11) and (15) over $[0, t_1)$ with $(\hat{\theta}_1(0), r_1(0)) = (\hat{\theta}_0, r_0)$. The proof is complete if $t_1 = \infty$. Otherwise $e_{\text{obs},1}(t_1) = \varepsilon'_{\text{obs}}$; hence $\lambda_{e,1}(t_1) = 0$ due to (12), and we proceed with Step 2 below.

Step 2: Consider (11) and (15) with $\lambda_e \equiv 0$, namely,

$$\begin{aligned}\dot{\hat{\theta}} &= 0, \\ \dot{r} &= \big[-\lambda_r \nabla_r \hat{J}(\hat{\theta}, r)^\top\big]_{T_{\mathcal{R}}(r)},\end{aligned} \tag{52}$$

which can also be modeled as a projected dynamical system (45) with the state $x := (\hat{\theta}, r)$ and the set $\mathcal{S} := \Theta \times \mathcal{R}$. Again, Lemma 5 ensures that there exists a solution $(\hat{\theta}_2, r_2)$ to (52) over $[t_1, \infty)$ with $(\hat{\theta}_2(t_1), r_2(t_1)) = (\hat{\theta}_1(t_1), r_1(t_1))$. Consider the corresponding observation error $e_{\text{obs},2}$ and switching signal $\lambda_{e,2}$ defined by (8) and (12), respectively, and let

$$t_2 := \inf\{t > t_1 : \|e_{\text{obs},2}(t)\| \geq \varepsilon_{\text{obs}}\}.$$

Then $(\hat{\theta}_2, r_2)$ is a solution to (11) and (15) over $[t_1, t_2)$ with $(\hat{\theta}_2(t_1), r_2(t_1)) = (\hat{\theta}_1(t_1), r_1(t_1))$. The proof is complete if $t_2 = \infty$. Otherwise $e_{\text{obs},2}(t_2) = \varepsilon_{\text{obs}}$; hence $\lambda_{e,2}(t_2) = \lambda_\theta$ due to (12), and we proceed with Step 1 above.

By switching between these two steps, we obtain an increasing time sequence $(t_k)_{k \in \mathbb{Z}_+}$ and a corresponding sequence $(\hat{\theta}_k, r_k)_{k \geq 1}$ of absolutely continuous functions $\hat{\theta}_k : [t_{k-1}, \infty) \to \Theta$ and $r_k : [t_{k-1}, \infty) \to \mathcal{R}$. Since all $\hat{\theta}_k$ and r_k evolve within the compact sets Θ and $\mathcal{R}$, respectively, the definition of e_{obs} in (8) and Assumption 1 imply that

$$\sup_{k \geq 1} \operatorname*{ess\,sup}_{t \geq t_{k-1}} \|\dot{e}_{\text{obs},k}(t)\| \leq M$$

for some finite constant $M \geq 0$. Then $t_k - t_{k-1} \geq (\varepsilon_{\text{obs}} - \varepsilon'_{\text{obs}})/M$ for all $k \geq 2$; hence $\lim_{k \to \infty} t_k = \infty$ (i.e., the so-called *Zeno behavior* [40, Sec. 1.2.2] cannot occur). Therefore, the absolutely continuous functions $\hat{\theta} : \mathbb{R}_+ \to \Theta$ and $r : \mathbb{R}_+ \to \mathcal{R}$ defined by

$$\hat{\theta}(t) := \hat{\theta}_k(t), \quad r(t) := r_k(t), \qquad k \geq 1, t \in [t_{k-1}, t_k)$$

form a solution to (11) and (15) over $\mathbb{R}_+$ with $(\hat{\theta}(0), r(0)) = (\hat{\theta}_0, r_0)$. Since θ and r evolve within the compact sets Θ and $\mathcal{R}$, respectively, the definition of e_{obs} in (8) and Assumption 1 imply that $\hat{\theta}$, r, e_{obs}, and their time derivatives are essentially bounded over $\mathbb{R}_+$.

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] J. P. Hespanha, *Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists*. Princeton University Press, 2017.

[4] T. Alpcan and T. Başar, *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press, 2010.

[5] D. Fudenberg and D. K. Levine, *The Theory of Learning in Games*. MIT Press, 1998.

[6] M. S. Kang, S. B. Lee, and V. D. Gligor, "The Crossfire attack," in *2013 IEEE Symposium on Security and Privacy*, 2013, pp. 127–141.

[7] H. von Stackelberg, *Market Structure and Equilibrium*. Springer, 2011, transl. from German.

[8] G. Yang, R. Poovendran, and J. P. Hespanha, "Adaptive learning in two-player Stackelberg games with continuous action sets," in *58th IEEE Conference on Decision and Control*, 2019, pp. 6905–6911.

[9] Y. A. Korilis, A. A. Lazar, and A. Orda, "Achieving network optima using Stackelberg routing strategies," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 161–173, Feb. 1997.

[10] T. Roughgarden, "Stackelberg scheduling strategies," *SIAM Journal on Computing*, vol. 33, no. 2, pp. 332–350, Jan. 2004.

[11] M. Bloem, T. Alpcan, and T. Başar, "A Stackelberg game for power control and channel allocation in cognitive radio networks," in *2nd International Conference on Performance Evaluation Methodologies and Tools*, 2007, 9 pages.

[12] J. Pita, M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus, "Deployed ARMOR protection: The application of a game-theoretic model for security at the Los Angeles International Airport," in *7th International Conference on Autonomous Agents and Multiagent Systems*, 2008, pp. 125–132.

[13] J. Tsai, S. Rathi, C. Kiekintveld, F. Ordóñez, and M. Tambe, "IRIS - A tool for strategic security allocation in transportation networks," in *8th International Conference on Autonomous Agents and Multiagent Systems*, 2009, pp. 37–44.

[14] G. G. Brown, W. M. Carlyle, J. Salmerón, and K. Wood, "Analyzing the vulnerability of critical infrastructure to attack and planning defenses," in *Emerging Theory, Methods, and Applications*, J. C. Smith, Ed. INFORMS, Sep. 2005, pp. 102–123.

[15] G. G. Brown, M. Carlyle, J. Salmerón, and K. Wood, "Defending critical infrastructure," *Interfaces*, vol. 36, no. 6, pp. 530–544, Dec. 2006.

[16] D. Sahabandu, J. Allen, S. Moothedath, L. Bushnell, W. Lee, and R. Poovendran, "Quickest detection of advanced persistent threats: A semi-Markov game approach," in *11th ACM/IEEE International Conference on Cyber-Physical Systems*, 2020, pp. 9–19.

[17] G. W. Brown, "Iterative solution of games by fictitious play," in *Activity Analysis of Production and Allocation*, T. C. Koopmans, Ed. John Wiley & Sons, 1951, pp. 374–376.

[18] J. Robinson, "An iterative method of solving a game," *The Annals of Mathematics*, vol. 54, no. 2, pp. 296–301, Sep. 1951.

[19] G. W. Brown and J. von Neumann, "Solutions of games by differential equations," in *Contributions to the Theory of Games*, H. W. Kuhn and A. W. Tucker, Eds. Princeton University Press, 1952, vol. I, ch. 6, pp. 73–80.

[20] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, Jul. 1965.

[21] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, Mar. 2005.

[22] S. Hart, "Adaptive heuristics," *Econometrica*, vol. 73, no. 5, pp. 1401–1430, Sep. 2005.

[23] L. Buşoniu, R. Babuška, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 2, pp. 156–172, Mar. 2008.

[24] J. R. Marden and J. S. Shamma, "Game theory and distributed control," in *Handbook of Game Theory with Economic Applications*, H. P. Young and S. Zamir, Eds. Elsevier, 2015, vol. 4, pp. 861–899.

[25] A. Gaunersdorfer and J. Hofbauer, "Fictitious play, Shapley polygons, and the replicator equation," *Games and Economic Behavior*, vol. 11, no. 2, pp. 279–303, Nov. 1995.

[26] J. Letchford, V. Conitzer, and K. Munagala, "Learning and approximating the optimal strategy to commit to," in *2nd International Symposium on Algorithmic Game Theory*, 2009, pp. 250–262.

[27] J. Marecki, G. Tesauro, and R. Segal, "Playing repeated Stackelberg games with unknown opponents," in *11th International Conference on Autonomous Agents and Multiagent Systems*, vol. 2, 2012, pp. 821–828.

[28] A. Blum, N. Haghtalab, and A. D. Procaccia, "Learning optimal commitment to overcome insecurity," in *28th Conference on Neural Information Processing Systems*, 2014, pp. 1826–1834.

[29] G. Leitmann, "On generalized Stackelberg strategies," *Journal of Optimization Theory and Applications*, vol. 26, no. 4, pp. 637–643, Dec. 1978.

[30] M. Breton, A. Alj, and A. Haurie, "Sequential Stackelberg equilibria in two-person games," *Journal of Optimization Theory and Applications*, vol. 59, no. 1, pp. 71–97, Oct. 1988.

[31] G. Yang and J. P. Hespanha, "Modeling and mitigating link-flooding distributed denial-of-service attacks via learning in Stackelberg games," in *Handbook of Reinforcement Learning and Control*, K. G. Vamvoudakis,

Y. Wan, F. L. Lewis, and D. Cansever, Eds. Springer, 2021, pp. 433–463.

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes The Art of Scientific Computing*, 3rd ed. Cambridge University Press, 2007.

[33] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. Springer, 1998.

[34] J.-P. Aubin, *Viability Theory*. Birkhäuser, 1991.

[35] A. S. Morse, D. Q. Mayne, and G. C. Goodwin, “Applications of hysteresis switching in parameter adaptive control,” *IEEE Transactions on Automatic Control*, vol. 37, no. 9, pp. 1343–1354, Sep. 1992.

[36] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Prentice Hall, 1996.

[37] A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*. Springer, 1988.

[38] C. Henry, “An existence theorem for a class of differential equations with multivalued right-hand side,” *Journal of Mathematical Analysis and Applications*, vol. 41, no. 1, pp. 179–186, Jan. 1973.

[39] E. P. Ryan, “An integral invariance principle for differential inclusions with applications in adaptive control,” *SIAM Journal on Control and Optimization*, vol. 36, no. 3, pp. 960–980, May 1998.

[40] D. Liberzon, *Switching in Systems and Control*. Birkhäuser, 2003.