Online Optimization of LTI Systems Under Persistent Attacks: Stability, Tracking, and Robustness

Felipe Galarza-Jimenez, Gianluca Bianchin, Jorge I. Poveda, Emiliano Dall'Anese*

Abstract

We study the stability properties of the interconnection of an LTI dynamical plant and a controller that generates control signals that are maliciously compromised by an attacker. We consider two classes of controllers: a static output-feedback controller, and a dynamical gradient-flow controller that seeks to steer the output of the plant towards the solution of a convex optimization problem. We analyze the stability of the system under a class of switching attacks that persistently modify the control inputs generated by the controllers. The stability analysis leverages the framework of Hybrid Dynamical Systems, Lyapunov-based arguments for switching systems with unstable modes, and singular perturbation theory. Our results show that the stability of the interconnected system can be preserved when an attack defense mechanism can mitigate "sufficiently often" the activation time of the attack action in any bounded time interval. We present simulation results that corroborate the technical findings.

Keywords: Switched Systems, Cyber-Physical Security, Online Optimization, Stability and Tracking

1 Introduction

This paper studies the stability properties of the interconnection of a Linear Time Invariant (LTI) dynamical system and an output feedback controller, with control inputs maliciously compromised by an external attacker. In particular, the output feedback controller is designed based on a gradient flow, with the objective of steering the output of the LTI plant towards the solution of an optimization problem [1, 2, 3, 4, 5]. The theoretical and algorithmic endeavors are motivated by a number of applications within the realm of cyber-physical systems (CPSs) – that is, physical stystems integrated with computational resources by means of a communication infrastructure; in particular, the modeling adopted in this paper is well suited for a number of applications in power systems [6, 5], transportation systems [7], communication networks [8], and robotics [9], to mention just a few. While advances in communication and cyber technologies provide enhanced functionality, efficiency, and autonomy of a CPS, the presence of communication channels as well as the tight integration between cyber and physical components unavoidibly introduces security vulnerabilities.

As illustrated in Figure 1, we consider an LTI dynamical system with two types of feedback controllers: (i) a static output-feedback controller that is used as an inner loop to stabilize the LTI system; and, (ii) an output feedback controller designed based on an appropriate modification of a gradient flow. For the synthesis of the gradient-flow controller, we start by formulating an unconstrained optimization problem with a composite cost function that captures performance indexes associated with the input and the output of the plant. The cost function is assumed to be smooth and to satisfy the Polyak-Łojasiewicz (PL) inequality condition (where we recall that the PL inequality is a weaker assumption than convexity, and it implies invexity) [10]. We consider attack actions of the form of *switching multiplicative attacks against the interconnection between the LTI plant and the controller*, whereby an attacker can persistently modify the inputs of the dynamical system with the objective of destabilizing the equilibrium points. This attack model is rather general and captures different classes of multiplicative attacks that can *persistently* modify the sign and/or the magnitude of the control signals, and also jam the communication channels.

With this setting in place, the problem addressed in this paper is that of finding sufficient conditions under which asymptotic and exponential stability is preserved for the closed-loop system. To this end, we adopt a

*The authors are with the Department of Electrical, Computer, and Energy Engineering at the University of Colorado Boulder, Boulder, CO, USA. email: {firstname.lastname}@colorado.edu. This work is supported in part by the National Science Foundation awards 1941896 and 1947613, and the CU Boulder Autonomous Systems IRT.

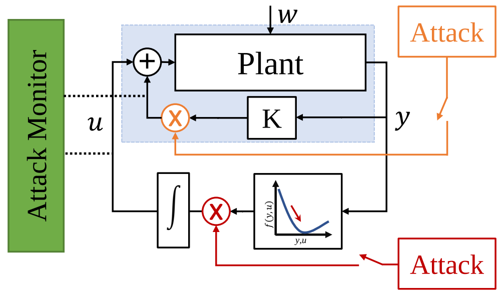

Figure 1: An inner control loop is used to stabilize the plant, while an outer control loop regulates the plant to an optimizer of (6).

hybrid systems framework, and we leverage arguments from singular perturbation analysis and input-to-state stability (ISS) for hybrid systems. This framework allows us to establish sufficient conditions in terms of the total activation time of the attacks acting on the system, and the time-scale separation between the plant and the gradient flow controller that preserves the stability properties of the interconnection.

Related Works. The design of feedback-based optimization controllers has recently received significant attention; see, for example [1, 2, 3, 4, 5, 9, 7]. In particular, in [3] sufficient conditions on the time-scale separation between plant and controller were derived to induce asymptotic stability properties. Similarly, the joint stabilization and optimal steady-state regulation of LTI dynamical systems was considered in [4]. LTI dynamical systems with time-varying exogenous inputs were considered in [5], along with the problem of tracking an optimal solution trajectory of a time-varying problem with a strongly convex cost; prediction-correction-type controllers were utilized to track the trajectory of a time-varying problem in [9]. Recently, [7] established exponential stability results for the interconnection of a switched LTI system and a hybrid feedback controller based on accelerated gradient dynamics with resets. Finally, [11] studied extremum-seeking algorithms for *static* optimization problems under deception attacks.

In recent years, and motivated by the increasing vulnerability of cyber-physical systems operating in adversarial environments, several works have investigated the stability properties of systems under denial of service attacks, whereby an attacker compromises system resources such as sensors or actuators, as well as infrastructure such as communication channels. We refer to [12, 13] (see also references therein) for a comprehensive list of references, while we present a list of representative references below. The authors in [14] designed a stabilizing controller for communication channels that face malicious random packet losses, while [15] presented a class of stabilizing event-triggered controllers. The work [16] designed scheduling policies that preserve the stability of an interconnected system when an attacker jams the control inputs sent to the plants. The works [17] and [18] investigated system stability in the presence of deception attacks, namely attacks where the integrity of control packets or measurements is compromised. A self-triggered consensus networks in the presence of communication failures caused by denial-of-service attacks was considered in [13]. Consensus in the presence of the denial-of-service was also considered in [19]. Jamming attacks were also studied in [20]. Event-triggered communication and decentralized control of switched systems under cyber attacks are analyzed in [21]; in particular, conditions on the dwell-time and the gain for the controller are derived to ensure stability. Other lines of work have studied the robustness properties with respect to attacks of different (discrete-time) optimization algorithms; e.g., [22, 23, 24]. In these works, Byzantine attacks in distributed algorithms are modeled by a group of malicious nodes that modify the data transmitted to their neighbors. A similar model was considered in [25] for sub-gradient methods. Furthermore, a model-free moving target defense framework for the detection and mitigation of sensor and/or actuator attacks with discrete-time dynamics was considered in [26].

Contributions. The contribution of this work is threefold. i) We present the first stability analysis of the interconnection between an LTI dynamical system and a gradient-flow controller operating under switching multiplicative attacks. We present a general framework that leverages analytical tools from set-valued hybrid dynamical systems theory, optimization, and feedback control theory; ii) We formulate a new class of switching-mode multiplicative attacks against the feedback signals, whereby the attacker transforms the control inputs produced by the controller according to a linear map with the objective of destabilizing the closed-loop system. This class of attacks is novel in the literature and includes as a special case denial of service and deception attacks. We characterize the level of defense (i.e., attack rejection) needed by the defense mechanism to guarantee exponential stability of the closed-loop system under a suitable time scale separation between the dynamics of the plant and the dynamics of the controller. iii) For systems with external exogenous inputs or disturbances, we establish sufficient conditions to guarantee exponential input-to-state stability with respect to the essential supremum of the time derivative of the disturbance. To establish this result, we derive an auxiliary lemma for switched systems with inputs and unstable modes, where the switching signals are generated by a hybrid dynamical system.

Organization. The rest of this paper is organized as follows. In Section 2 we present the notation used throughout the paper, and some preliminaries on hybrid dynamical systems. In Section 3, we formalize the model of the system and the problem under study. Section 4 presents the main results, followed by the analysis presented in Section 5. Section 6 presents some numerical experiments, and Section 7 ends with the conclusions.

2 Preliminaries and Notation

We begin by introducing the notation that will be used throughout the paper. Given a compact set $\mathcal{A} \subset \mathbb{R}^n$, $n \in \mathbb{Z}_{\geq 0}$, and a vector $z \in \mathbb{R}^n$, we let $|z|_{\mathcal{A}} := \min_{s\in\mathcal{A}} \|z - s\|_2$ denote the minimum distance between z and $\mathcal{A}$. For a function $w : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, we denote $\|w\|_t$ as the norm supremum on the interval $[0, t]$. For a matrix $A \in \mathbb{R}^{n\times d}$, with $d \in \mathbb{Z}_{\geq 0}$, we use $\|A\|$ to denote the induced Euclidean norm of A. When $n = d$, we use $\overline{\lambda}(A)$ and $\underline{\lambda}(A)$ to denote the largest and smallest eigenvalue of A, respectively. Given a set O, we use $I_O(x)$ to denote the indicator function for the set O, namely $I_O(x) = 1$ if $x \in O$ and $I_O(x) = 0$ if $x \notin O$. Given vectors $p_1, p_2 \in \mathbb{R}^n$, we denote by (p_1, p_2) their concatenation. A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is class-$\mathcal{K}$ if it is continuous, zero at zero, and strictly increasing. If, in addition, $\lim_{r\to\infty} \alpha(r) = \infty$, it is said to be of class-$\mathcal{K}_\infty$. A function $\sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is called of class-$\mathcal{L}$ if it is continuous, decreasing and $\lim_{r\to\infty} \sigma(r) = 0$. And, a function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be of class-$\mathcal{KL}$ if it is class-$\mathcal{K}$ in its first argument and class-$\mathcal{L}$ in its second argument.

In this paper, we will consider mathematical models corresponding to set-valued dynamical systems that combine continuous-time dynamics (with inputs) and discrete-time dynamics. These systems are referred to as Hybrid Dynamical Systems (HDS) [27], an they are characterized by the following hybrid inclusions:

$$p \in C, \quad \dot{p} \in F(p, u), \qquad\qquad p \in D, \quad p^+ \in G(p), \tag{1}$$

where $p \in \mathbb{R}^n$ is the state system, $u \in \mathbb{R}^m$ is the input, $F : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is a set-valued map (called the *flow map*) that governs the continuous-time dynamics of the system when the state belongs to the flow set $C \subset \mathbb{R}^n$. Similarly, $G : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is a set-valued map (called the *jump map*) that describes the discrete-time dynamics of the system when the state belongs to the jump set $D \subset \mathbb{R}^n$. The tuple $\mathcal{H} = \{C, F, D, G\}$ completely characterizes the HDS, and it is called the *data* of the HDS. Solutions to (1) are defined on *hybrid time domains*, namely they are indexed by a parameter $t \in \mathbb{R}_{\geq 0}$ that increases continuously during flows, and by a parameter $j \in \mathbb{Z}_{\geq 0}$ that increases by one unit during the jumps. A set $E \subset \mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ is called a *compact* hybrid time domain if $E = \cup_{j=0}^{J-1}([t_j, t_{j+1}], j)$ for some $0 = t_0 \leq t_1 \ldots \leq t_J$. The set E is a hybrid time domain if for all $(T, J) \in E$, the set $E \cap ([0, T] \times \{0, \ldots, J\})$ is a compact hybrid time domain. Using the notion of hybrid time domains we can formally introduce the concept of solution to systems of the form (1).

Definition 1 *A function $p : dom(p) \to \mathbb{R}^n$ is a hybrid arc if $dom(p)$ is a hybrid time domain and $t \mapsto p(t, j)$ is locally absolutely continuous for each j such that the interval $I_j := \{t : (t, j) \in dom(p)\}$ has nonempty interior. A hybrid arc p is a* solution *to* (1) *with $dom(u) = dom(p)$ if $p(0, 0) \in \overline{C} \cup D$, and the following two conditions hold:*

1. *For each $j \in \mathbb{Z}_{\geq 0}$ such that I_j has nonempty interior: $p(t, j) \in C$ for all $t \in int(I_j)$, and $\dot{p}(t, j) \in F(p(t, j), u(t, j))$ for almost all $t \in I_j$.*
2. *For each $(t, j) \in dom(p)$ such that $(t, j + 1) \in dom(p)$: $p(t, j) \in D$, and $p(t, j + 1) \in G(p(t, j))$.*

By working with hybrid time domains we can exploit suitable graphical convergence notions to establish sequential compactness results for the solutions of (1), e.g., the graphical limit of a sequence of solutions is also a solution. Such types of results will be instrumental for the robustness analysis of the systems studied in this paper.

Definition 2 *A hybrid solution p is maximal if there does not exist another solution ψ to $\mathcal{H}$ such that dom(p) is a proper subset of dom(ψ), and $p(t,j) = \psi(t,j)$ for all $(t,j) \in$ dom(p). A maximal hybrid solution is said to be complete if its domain is unbounded.*

In this paper, we are interested in establishing suitable *convergence and stability* properties for a class of dynamical systems under attacks. To do this, the following definitions will be instrumental.

Definition 3 *Given a compact set $\mathcal{A} \subset C \cup D$, system (1) with $u = 0$ is said to render the set $\mathcal{A}$ uniformly globally asymptotically stable (UGAS) if there exists a class $\mathcal{KL}$ function β such that every solution of (1) satisfies $|p(t,j)|_{\mathcal{A}} \leq \beta(|p(0,0)|_{\mathcal{A}}, t+j)$ for all $(t,j) \in$ dom(p). If $\beta(r,s) = c_1 r e^{-c_2 s}$ for $c_1, c_2 > 0$, the system is said to render the set $\mathcal{A}$ uniformly globally exponentially stable (UGES).*

It is important to note that the stability notion of Definition 3 subsumes the standard asymptotic and stability notions considered in the literature of continuous-time systems ($D = \emptyset$) and discrete-time systems ($C = \emptyset$). In particular, recall that an LTI dynamical system given by $\dot{x} = Ax$ renders the origin $\mathcal{A} = \{0\}$ UGES if A is Hurwitz, i.e., $\text{Re}\{\lambda(A)\} < 0$.

When the input u in (1) is not identically zero, the notion of *input-to-state stability* can be used to qualitatively characterize the effect of the input on the stability properties of the system

Definition 4 *For every measurable function $u : \text{dom}(u) \to \mathbb{R}^m$ with dom$(u) =$ dom(x), system (1) is said to render the compact set $\mathcal{A}$ input-to-state stable (ISS) with respect to u if there exists a class $\mathcal{KL}$ function β, and a class $\mathcal{K}$ function γ, such that every solution of the system satisfies the bound*

$$|p(t,j)|_{\mathcal{A}} \leq \beta(|p(0,0)|_{\mathcal{A}}, t+j) + \gamma\left(\|u\|_{(t,j)}\right), \tag{2}$$

for all $(t,j) \in$ dom(p). When β has an exponential form, we say that system (1) renders the set $\mathcal{A}$ exponentially input-to-state stable (E-ISS).

The properties of UGAS, UGES, and ISS for hybrid systems can be readily established via suitable Lyapunov-based conditions; see [27, Ch.3] for sufficient conditions that certify UGAS, [28, Thm. 1] for UGES, and [29, Thm. 3.1] for ISS. We will leverage some of these tools in our analysis.

3 Problem Formulation

In this paper, we study the stability properties of a class of online optimization controllers designed to regulate the output of an LTI plant towards the solution of a general convex optimization problem. However, we are interested in settings where the plant and the controller operate in adversarial environments subject to persistent attacks that aim to destabilize the closed-loop system. To formally describe this behavior, we first proceed to characterize the nominal model of the plant and the structure of the proposed controller. After this, we present our attack model, and we formalize the joint control-optimization problem that is addressed in this work.

3.1 Nominal Model of the Plant and the Controller

We consider plants modeled as LTI dynamical systems of the form:

$$\dot{x} = Fx + Nv + Bu + Ew, \quad y = Cx, \tag{3}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^{m_v}$ are control inputs, $w \in \mathbb{R}^q$ is an *unknown* exogenous signal, $y \in \mathbb{R}^p$ is the measurable output, and F, N, B, E, C are matrices of suitable dimensions. In general, the matrix F might be not Hurwitz.

To stabilize and optimize system (3), as shown in Figure 1, two control loops are applied to the plant. First, an inner control loop, describing a static feedback law, is applied via the control input v:

$$v = Ky, \quad K \in \mathbb{R}^{m_v \times p}. \tag{4}$$

Using (3) and (4), we denote by $A := F + NKC$ the closed-loop matrix of the plant with inner loop. We assume that the feedback law has the *stabilizing property*, which is formalized by the following assumption.

Assumption 1 *For every positive definite matrix $R \in \mathbb{R}^{n\times n}$, there is a unique positive definite matrix $P \in \mathbb{R}^{n\times n}$ such that $A^\top P + PA = -R$. Moreover, the function $t \mapsto w(t)$ is continuously differentiable, and there exists $\lambda > 0$ such that $||w(t)|| \leq \lambda$ for all $t \geq 0$.*

Functions w satisfying the conditions of Assumption 1 are said to be *admissible.*

Next, an outer control loop of the form of a low-gain dynamical controller acts on the input u:

$$\dot{u} = -\varepsilon \left(\nabla f_u(u) + G^\top \nabla f_y(y)\right), \quad G := -CA^{-1}B, \tag{5}$$

where $f_u : \mathbb{R}^m \to \mathbb{R}$ and $f_y : \mathbb{R}^p \to \mathbb{R}$ are suitable cost functions (formally defined below), and $\varepsilon > 0$ is a tunable gain.

Under a suitable choice of the controller gain ε, the controller (5) steers the dynamical system (3) towards the solutions of the following optimization problem:

$$\min_{u,y} \; f_u(u) + f_y(y), \qquad \text{s.t.} \quad 0 = Ax + Bu + Ew, \qquad y = Cx. \tag{6}$$

Thus, the controller (5) regulates system (3) towards an equilibrium point that minimizes the cost $f_u(u) + f_y(y_{ss})$, where $f_u(\cdot)$ is interpreted as a cost associated with the steady-state control input, and $f_y(\cdot)$ is interpreted as a cost associated with the steady-state system output y_{ss}.

Remark 1 *Feedback controllers inspired from optimization algorithms of the form* (5) *have received significant research attention during the last decade thanks to their flexibility and performance guarantees [3, 4, 5]. In particular, they have been applied to control power systems [5], transportation networks [30], and communication systems.*

In what follows, we denote by $f(u, w) := f_u(u) + f_y(Gu + Hw)$, where we recall that $G := -CA^{-1}B$, and we let $H := -CA^{-1}E$. Then, the optimization problem (6) can be rewritten as:

$$\min_{u} f(u, w) := f_u(u) + f_y(Gu + Hw), \tag{7}$$

where w acts as an exogenous signal that parametrizes the solution of problem (7). To guarantee that this problem is well-defined, we make the following standard regularity assumptions on the cost functions $f_u(\cdot)$ and $f_y(\cdot)$.

Assumption 2 *The functions $u \mapsto f_u(u)$ and $y \mapsto f_y(y)$ are continuously differentiable, and there is ℓ_u, $\ell_y > 0$ such that for every $u, u' \in \mathbb{R}^m$ and $y, y' \in \mathbb{R}^p$, we have $\|\nabla f_u(u) - \nabla f_u(u')\| \leq \ell_u \|u - u'\|$, and $\|\nabla f_y(y) - \nabla f_y(y')\| \leq \ell_y \|y - y'\|$.*

Assumption 3 *The function $u \mapsto f(u, w)$ satisfies the PL inequality, uniformly in w; that is, $\exists \mu > 0$ such that $\frac{1}{2}\|\nabla f(u, w)\|^2 \geq \mu(f(u, w) - f(u^*, w))$, $\forall u \in \mathbb{R}^m$, $\forall\ ||w|| \leq \lambda$. Furthermore, for each w the minimizer u^*_w is unique.*

Remark 2 *Assumption 2 guarantees that the cost functions are sufficiently smooth, which is standard in the literature of online optimization. Indeed, by Assumption 2, the mapping $u \mapsto f(u, w)$ has a globally Lipschitz gradient, uniformly in w, with Lipschitz constant $\ell = \ell_v + \ell_y ||G||^2$. Note that Assumption 3 implies that the function f is invex; it is one of the weakest assumptions in the optimization literature that ensures optimization algorithms to exhibit linear convergence [10]. We also note that the PL inequality implies the quadratic growth condition $f(u, w) - f(u^*_w, w) \geq \frac{\mu}{2}\|u - u^*_w\|^2$, $\forall u \in \mathbb{R}^m$ and $\forall\ ||w|| \leq \lambda$, which implies radial unboundedness of f.*

By combining the plant model (3) with the controllers (4)-(5) we obtain the following nominal closed-loop system with states $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{aligned} \dot{x} &= (F + NKC)x + Bu + Ew, \quad y = Cx, \\ \dot{u} &= -\varepsilon\Big(\nabla f_u(u) + G^\top \nabla f_y(y)\Big). \end{aligned} \tag{8}$$

The stability and convergence properties of the nominal closed-loop system can be studied using suitable Lyapunov functions with a sufficiently small ε, which introduces a time-scale separation between the dynamics of the controller and the dynamics of the plant [31, 32, 3, 7]. However, unlike this type of standard model, we are interested in studying the stability properties of (8) under the classes of attacks described in the following section.

3.2 Attack Model

Systems of the form (8) often operate in adversarial environments, where an attacker can target the system, controller, or communication channels in order destabilize the interconnection. In this work, we focus our attention on attacks that are able to modify the control signals, and that are persistent. In particular, the attacks can be intermittent and not necessarily periodic, see Figure 1. Such types of non-constant aperiodic attacks are difficult to completely repeal or mitigate using defense mechanisms and monitoring technologies. We consider two types of persistent attacks against the control signals. First, we consider attacks acting on the gradient-flow controller (5), which we model as follows:

$$\text{Nominal System: } \dot{u} = -\varepsilon\left(\nabla f_u(u) + G^{\mathsf{T}}\nabla f_y(y)\right), \tag{9a}$$

$$\text{System Under Attack: } \dot{u} = -\varepsilon M_{\sigma_u}\left(\nabla f_u(u) + G^{\mathsf{T}}\nabla f_y(y)\right), \tag{9b}$$

where $\sigma_u : \mathbb{R}_{\geq 0} \to \Sigma_{u,a}$ is a switching signal taking values in the finite set of indices $\Sigma_{u,a} \subset \mathbb{Z}_{>0}$, and $M_{\sigma_u} \in \mathbb{R}^{m\times m}$ is a matrix that describes the attack transformation map. The model (9) is general enough to capture different types of attacks, including those that are able modify the sign and/or the magnitude of the gradient (or some partial derivatives); see [22, 23, 24, 25]. The model (9) can also be used to capture jamming attacks and denial-of-service attacks; see [13, 20].

Similarly, we consider attacks acting on the plant. These attacks are modeled as follows:

$$\text{Nominal System:} \quad v = Ky, \qquad \dot{x} = Fx + Nv + Bu + Ew, \tag{10a}$$

$$\text{System Under attack:} \quad v = L_{\sigma_v}Ky, \quad \dot{x} = Fx + Nv + L^b_{\sigma_v}Bu + L^e_{\sigma_v}Ew, \tag{10b}$$

where $\sigma_v : \mathbb{R}_{\geq 0} \to \Sigma_{v,a}$ is a switching signal taking values in a finite set of indices $\Sigma_{v,a} \subset \mathbb{Z}_{>0}$. For each $\sigma_v \in \Sigma_{v,a}$, $L_{\sigma_v} \in \mathbb{R}^{m_v\times m_v}$ is a matrix that describes the attack transformation map applied to the internal feedback control loop. On the other hand, $L^e_{\sigma_v}, L^b_{\sigma_v}$ are real matrices of appropriate dimensions that induce a transformation of the inputs Bu and Ew, respectively.

To state our technical assumptions on the attacks, we next formulate the *Nominal System* and the *System Under Attack* using a unified modeling framework. We let $\Sigma_{v,s} = \{s\}$ and $L_s = I \in \mathbb{R}^{m_v\times m_v}$ denote the index and the transformation map corresponding to the *Nominal System* in (10). Similarly, we let $\Sigma_{u,s} = \{s\}$ and $M_s = I \in \mathbb{R}^{m\times m}$ denote the index and the transformation map corresponding to the *Nominal System* in (9). Using these definitions, we can rewrite systems (9)-(10) as switching systems of the form

$$\dot{u} = -\varepsilon M_{\sigma_u}\left(\nabla f_u(u) + G^{\mathsf{T}}\nabla f_y(y)\right), \qquad v = L_{\sigma_v}Ky, \qquad B_{\sigma_u} := L^b_{\sigma_v}B, \qquad E_{\sigma_u} := L^e_{\sigma_v}E, \tag{11}$$

where now the switching signal σ_v takes values in the extended set $\Sigma_v := \Sigma_{v,s} \cup \Sigma_{v,a}$, and the switching signal σ_u takes values in the extended set $\Sigma_u := \Sigma_{u,s} \cup \Sigma_{u,a}$.

Assumption 4 *The sets Σ_v and Σ_u are finite. Additionally, any attack is destabilizing, namely:*

(i) For all $\sigma_v \in \Sigma_{v,a}$, the matrix $F+NL_{\sigma_v}KC$ is invertible and admits at least one eigenvalue with positive real part.

(ii) For all $\sigma_u \in \Sigma_{u,a}$, the matrix M_{σ_u} has at least one eigenvalue with negative real part. Moreover, there exists $\bar{M} \in \mathbb{R}_{>0}$ such that $\|M_{\sigma_u}\| \leq \bar{M}$ holds for all $\sigma_u \in \Sigma_{u,a}$.

It follows from Assumption 4 that all modes described by the sets $\Sigma_{v,a}$ and $\Sigma_{u,a}$ are unstable for the closed-loop system (8). Moreover, since the sets $\Sigma_{v,a}$ and $\Sigma_{u,a}$ are countable, in the remainder we will enumerate the unstable modes as follows: $\Sigma_{v,a} := \{a_1, \ldots, a_{|\Sigma_{v,a}|}\}$ and $\Sigma_{u,a} := \{a_1, \ldots, a_{|\Sigma_{u,a}|}\}$, where the distinction between the two sets will be made clear by the context. By combining Assumptions 1 and 4(i), it follows that for every $a_i \in \Sigma_{v,a}$, the matrix

$$(F + NL_{a_i}KC)^\top P + P(F + NL_{a_i}KC) := \hat{R}_{a_i},$$

has at least one positive real eigenvalue. Throughout the rest of this paper, we denote as $\bar{\lambda}(\hat{R}_a) := \max_{a_i \in \Sigma_{v,a}} \bar{\lambda}(\hat{R}_{a_i})$ the largest eigenvalue of all the matrices $\hat{R}_{a_i}$.

3.3 Detection and Mitigation Mechanism

The models (9) and (10) describe a family of persistent multiplicative attacks that intermittently modify the control signal generated by the controller via the switching maps L_{σ_v} and M_{σ_u}. In gradient-based controllers, such types of attacks can easily induce instability in the closed-loop system (8) by persistently altering the direction of the vector field $\dot{u}$, or the gains of the feedback law (4). Since the attacker can switch between multiple adversarial matrices L_{σ_v} and M_{σ_u}, a defensive mechanism able to effectively anticipate and block the attacks *at all times* is practically and economically unfeasible. Indeed, since the attack is multiplicative rather than additive, standard defensive mechanisms for constant adversarial signals, including dynamic filters [17], identification mechanisms [18], dynamic compensators [15], etc, cannot be used to reject the attacks at all times. Therefore, instead of assuming the existence of an ideal defense mechanism able to completely reject an attack, in this paper we consider a more realistic scenario where the control system is equipped with a security mechanism that is able to mitigate the attacks only *sufficiently often*. To model this scenario, for every time-interval $0 \leq s < t$, we define the activation time of the attacks (9) and (10) as

$$T_u(s,t) := \int_s^t I_{\Sigma_{u,a}}(\sigma_u(\tau))d\tau, \qquad \text{and} \qquad T_v(s,t) := \int_s^t I_{\Sigma_{v,a}}(\sigma_v(\tau))d\tau,$$

where we recall that $I_\Sigma(\sigma)$ denotes the indicator function for the set Σ (see Section 2). Notice that, if $\sigma_v = \sigma_u = s$ at all times, then $T_v(s,t) = T_u(s,t) = 0$ for any choice of $0 \leq s < t$ since no attack is active in the closed-loop.

In what follows, we consider defensive mechanisms that guarantee the following *persistent rejection* property.

Definition 5 ***(Persistent Rejection)*** *For the attacks* (10) *and* (9)*, a defensive mechanism is said to satisfy the* uniform persistent rejection *property with time-ratio parameters* $\kappa_{v,2}, \kappa_{u,2} \in (0,1)$*, if the following conditions hold for any time interval* $[s,t]$*:*

$$T_v(s,t) \leq \kappa_{v,2}(t-s) + T_{0,v}, \quad T_u(s,t) \leq \kappa_{u,2}(t-s) + T_{0,u}, \tag{12}$$

where $T_{0,v}, T_{0,u} \in \mathbb{R}_{\geq 0}$*; and, for* $\kappa_{v,1}, \kappa_{u,1} > 0$ *the number of switches of the signal* $\sigma(t)$ *in the interval* $[s,t]$ *satisfies the bounds:*

$$N_v(s,t) \leq \kappa_{v,1}(t-s) + N_{0,v}, \quad N_u(s,t) \leq \kappa_{u,1}(t-s) + N_{0,u}, \tag{13}$$

where $N_{0,v}, N_{0,u} \in \mathbb{Z}_{\geq 1}$.

According to (12), defensive mechanisms that satisfy the uniform persistent rejection property guarantee that the total activation time of any attack in a time interval grows at most linearly with the length of the time interval. Moreover, by condition (13), the defense mechanism prevents the emergence of Zeno behavior in the system, permitting only switches of σ that satisfy an average dwell-time constraint. Similar attack models have been considered in the literature; see [13, 20, 11]. Indeed, conditions (12)-(13) are quite general: they capture periodic or aperiodic attacks, and they allow a finite number of consecutive switches in every interval of time of sufficiently large length. Note that any cyber-security system endowed with (12) might not be able to guarantee closed loop stability when the time-ratio parameters are close to 1. Indeed, lower values of $\kappa_{v,2}$ and $\kappa_{u,2}$ describe more effective mitigation mechanism able to reduce the activation time of the attacks in the controllers. This reduction usually comes at the price of requiring more effective sensors, actuators, and computational power in the defense system, thus inducing a trade-off between the economic/technological feasibility of the controller and the effectiveness of the defensive mechanism.

Remark 3 *The quantities* $\{T_{0,v}, \kappa_{v,2}, T_{0,u}, \kappa_{u,2}\}$ *can be interpreted as design parameters that can be selected by a system planner by properly designing the attack monitor and the defense mechanism to limit the effects of possible attack actions against the system. Alternatively, the quantities* $\{T_{0,v}, \kappa_{v,2}, T_{0,u}, \kappa_{u,2}\}$ *can also be interpreted as attack design parameters, namely given a pre-designed attack rejection mechanism, the above quantities can be tuned by an attacker by modifying the frequency at which the attack is activated. Both engineering interpretations will be consistent with our theoretical framework.*

3.4 Control Objectives: Approximate Optimal Tracking Under Attacks

Based on the model introduced in the previous sections, we can now formalize the problem under study in this paper. In particular, for the family of closed-loop systems under attacks described by system (8), satisfying Assumptions 1-4, we are interested in characterizing the amount of persistent rejection of attacks, and the time-scale separation between the plant and the controller, needed to preserve the stability properties of the closed-loop system, and to guarantee exponential convergence of the trajectories to the set of solutions of the online optimization problem described by (6).

The following problem formalizes the stated objectives

Problem 1 *Determine, when possible, a set of parameters* $(\varepsilon, \kappa_{v,2})$*, or* $(\varepsilon, \kappa_{u,2})$ *such that the equilibrium and optimal point* $z^* = (x^*, u_w^*)$ *of system* (8) *under attack is exponentially input-to-state stable (E-ISS) with respect to the time-variation of the exogenous signal* $t \mapsto w(t)$*.*

In Problem 1, the motivation for the study of ISS instead of just UGAS or UGES comes from the potentially time-varying nature of the disturbance w. In this case, the performance of the controller will be quantified by the ISS gain γ that will characterize the effect of $\|\dot{w}\|$ in the size of the residual set where the trajectories converge.

Remark 4 *To simplify our presentation, and also due to space limitations, we will focus on attacks that are exclusive either to the plant or to the controller, i.e., we do not address simultaneous attacks to the static and dynamic feedback controller. This allow us to simplify our notation since in this case if* $\sigma_v = a_i$ *then* $\sigma_u = s$*, and if* $\sigma_u = a_i$ *then* $\sigma_v = s$ *as defined before. Note, however, that considering simultaneous attacks in the plant and the controller is a natural extension of this work; the stability and tracking properties can be established by following similar steps as those presented in the paper.*

Before presenting our main results, we list three representative examples of applications in cyber-physical systems for which the modelling and control framework considered in this paper is particularly well-suited.

Example 1 *The dynamical model* (3) *and the interconnection* (8) *fit within the context of real-time frequency control and economic optimization of power transmission systems. In particular, the matrix* K *can be designed based on the so-called automatic generation control (AGC) as well as pertinent droop controllers implemented in the generation units [33]; on the other hand, the gradient-flow controller in* (8) *can be utilized to produce setpoints for the generators (both conventional fossil-fuel and renewable-based units) in order to solve an economic dispatch (ED) or optimal power flow (OPF) problem in real time [34]; see also the example provided in [5]. The models of attacks considered in this paper capture in this case malicious attacks to the AGC signals, droop control loops, and ED/OPF commands.*

Example 2 *The model* (3) *and* (8) *finds pertinent applications in the context of intelligent transportation systems, where the properties of safety and resilience are critical for the feedback controllers that operate the system. For example, for a ramp metering problem in a highway system, the inner control loop models local controllers such as the ALINEA [35]. On the other hand, gradient-flow controllers have been used to drive the equilibrium point of the traffic flows on the highway system towards the solution of a network-level problem as explained in, e.g., [30] and [36]. It then follows that the attack models consider in* (11) *can be used to capture adversarial jamming of the feedback gains of the ALINEA controller, or attacks acting on the control directions obtained from the gradient flow algorithm.*

Example 3 *The model* (11) *is also applicable to several applications in robotics, such as the control of autonomous vehicles using navigation laws based on potential and anti-potential functions; see [37, 38, 39].*

Indeed, in this scenario, it is common to approximate the vehicle dynamics as a linear system (3) *for which an internal feedback loop is designed to stabilize an external reference u; see [40]. In general, the internal controller is designed to guarantee a unitary DC gain, and to operate in a faster time scale (limited by the actuator dynamics of the vehicle), such that the "steady state" of the vehicle satisfies $x^* = u$, where x can model positions or velocities. The input u is then regulated via a gradient flow of the form* (11) *aiming to converge to the minimizer (e.g., the target point) of an artificial potential field. In this case, to prevent the vehicle from converging to the target, an external attacker can persistently alter the sign of the gradient dynamics, i.e., $M_{\sigma_u} \in \{I, -I\}$, effectively moving the vehicle* away *from the target. Note that this type of attacks will be particularly detrimental for navigation laws that implement anti-potential fields to avoid obstacles, given that in this case the attacks will effectively push the vehicles* towards *the obstacles.*

4 Main Results

To address Problem 1, we will use the framework of HDS (1), which will allows us to model the switching systems (9)-(10) as time-invariant (hybrid) systems, and also to consider suitable (set-valued) dynamic models of the uniform persistent rejection property induced by the defensive mechanism of the controllers. In order to do this, and for simplicity of exposition, we drop the subscripts from (12)-(13), and we consider a dynamical system with state $\varrho := (\tau_1, \tau_2, \sigma) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \Sigma$, and the following hybrid dynamics:

$$\varrho \in C_\varrho := [0, N_0] \times [0, T_0] \times \Sigma, \tag{14a}$$

$$\begin{pmatrix} \dot{\tau}_1 \\ \dot{\tau}_2 \\ \dot{\sigma} \end{pmatrix} \in F(\varrho) := \begin{pmatrix} [0, \kappa_1] \\ [0, \kappa_2] - I_{\Sigma_a}(\sigma) \\ 0 \end{pmatrix}, \tag{14b}$$

$$\varrho \in D_\varrho := [1, N_0] \times [0, T_0] \times \Sigma, \tag{14c}$$

$$\begin{pmatrix} \tau_1^+ \\ \tau_2^+ \\ \sigma^+ \end{pmatrix} \in G(\varrho) := \begin{pmatrix} \tau_1 - 1 \\ \tau_2 \\ \Sigma \backslash \sigma \end{pmatrix}, \tag{14d}$$

where $T_0 \geq 0$, $N_0 \in \mathbb{Z}_{\geq 1}$, $\kappa_1 > 0$, $\kappa_2 \in (0, 1)$. The following Lemma, originally established in [41], will be instrumental for our results.

Lemma 1 *Every complete solution ϱ of* (14) *has a hybrid time-domain dom(ϱ) that satisfies condition* (12)*. Moreover, every signal τ satisfying the persistent rejection property can be generated by the dynamical system* (14) *with a suitable initial condition.*

Lemma 1 allows us to study the closed-loop system under persistent switching attacks by studying the interconnection between the dynamics (9), (10) and (14). Indeed, the hybrid dynamics (14) will effectively act as a model of the switching behavior of the persistent attacks acting in the system. Similar models have been studied in the literature for the analysis of *static* model-based and model-free optimization problems; see [11, 41]. However, unlike these works, Problem 1 describes an *online optimization problem* with a *dynamic* plant in the loop, which has not been studied before in the literature.

In the ensuing section, we derive sufficient conditions for the stability of the interconnected system when the gradient-flow controller is affected by attacks as in (9). After this, we establish similar results for the case when the static feedback controller is under attack, as in (10).

4.1 Attacks Against Dynamic Feedback

In this section, we analyze the stability properties of the closed-loop system with attacks acting on the dynamic gradient-based controller. To model this scenario, we consider the following HDS with state $\vartheta :=$

$(x, u, \tau_{v,1}, \tau_{v,2}, \sigma_u)$, and dynamics:

$$\vartheta \in C := \mathbb{R}^n \times \mathbb{R}^m \times [0, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u, \tag{15a}$$

$$\begin{pmatrix} \dot{x} \\ \dot{u} \\ \dot{\tau}_{u,1} \\ \dot{\tau}_{u,2} \\ \dot{\sigma}_u \end{pmatrix} \in F_{\sigma_u}(\vartheta, w) := \begin{pmatrix} Ax + Bu + Ew \\ -\varepsilon M_{\sigma_u} \left(\nabla f_u(u) + G^\top \nabla f_y(y) \right) \\ [0, \varepsilon \kappa_{u,1}] \\ [0, \varepsilon \kappa_{u,2}] - \varepsilon I_{\Sigma_{u,a}}(\sigma_u) \\ 0 \end{pmatrix}, \quad y = Cx, \tag{15b}$$

$$\vartheta \in D := \mathbb{R}^n \times \mathbb{R}^m \times [1, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u, \tag{15c}$$

$$\begin{pmatrix} x^+ \\ u^+ \\ \tau_{u,1}^+ \\ \tau_{u,2}^+ \\ \sigma_u^+ \end{pmatrix} \in G_{\sigma_u}(\vartheta) := \begin{pmatrix} x \\ u \\ \tau_{u,1} - 1 \\ \tau_{u,2} \\ \Sigma_u \backslash \{\sigma_u\} \end{pmatrix}. \tag{15d}$$

This HDS models the interconnection between the signal generator (14) and the closed-loop system (8) with the dynamic controller under attack (9).

Remark 5 *In* (15b)*, the dynamics of the states $\tau_{u,1}$ and $\tau_{u,2}$ are multiplied by ε. This is done without loss of generality, and only to simplify the modeling framework such that the attacks operate in the same time scale as the gradient controller. In fact, since in our case $\varepsilon \in (0,1)$, the dynamics of* (15b) *could also be written in terms of the scaled time ratios $\tilde{\kappa}_{u,1} = \varepsilon \kappa_{u,1}$ and $\tilde{\kappa}_{u,2} = \varepsilon \kappa_{u,2}$, and a ε-scaled indicator function.*

The following Lemma, which follows directly by construction, will play an important role in the robustness analysis of system (15).

Lemma 2 *For the HDS* (15) *the following holds: (a) The sets C and D are closed; (b) For each w, the set-valued mapping $F_{\sigma_u}(\cdot, w)$ is outer-semicontinuous, locally bounded, and convex-valued in C; (c) The set-valued mapping G_{σ_u} is outer-semicontinuous and locally bounded in D.*

For system (15), we are interested in characterizing sufficient conditions on the gain ε of the controller, and the time-ratio parameter $\kappa_{u,2}$, to guarantee the solution of Problem 1. To do this, we first neglect the plant dynamics by assuming they are infinitely fast compared to the dynamics of the controller. Namely, we disregard the dynamics $\dot{x}$, and we substitute x by the steady-state mapping $x(u, w) = A^{-1}Bu + Ew$, where w is assumed to be a constant disturbance, i.e., $\dot{w}$=0. For this simplified system we establish the following proposition, which is the first result of this paper. It establishes a novel characterization of the level of persistent rejection of attacks needed by the defense mechanism to guarantee uniform exponential stability of the gradient-based controller under attacks.

Proposition 1 *Suppose that Assumptions 1-4 hold, and that the time-ratio parameter $\kappa_{u,2}$ satisfies*

$$0 < \kappa_{u,2} < \frac{1}{1+\alpha}, \qquad \textit{with} \qquad \alpha := -\min_{\sigma_u \in \Sigma_{u,a}} \underline{\lambda}(M_{\sigma_u}) \geq 0. \tag{16}$$

Then, for each admissible and constant exogenous signal w, and each compact set $K_x \subset \mathbb{R}^n$, the set

$$\mathcal{A}^*_{K_x} = \left\{ (x, u, \tau_{u,1}, \tau_{u,2}, \sigma_u) : x \in K_x, \ u = u^*_w, \ \tau_{u,1} \in [0, N_{0,u}], \tau_{u,2} \in [0, T_{0,u}], \sigma_u \in \Sigma_u \right\},$$

is UGES for the HDS (15) *with $\dot{x} = 0$ and $\varepsilon = 1$, and with restricted flow and jump sets*

$$C_{K_x} = (\mathbb{R}^n \cap K_x) \times \mathbb{R}^m \times [0, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u, \qquad D_{K_x} = (\mathbb{R}^n \cap K_u) \times \mathbb{R}^m \times [1, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u. \tag{17}$$

The result of Proposition 1 establishes an explicit characterization of a time-ratio parameter $\kappa_{u,2}$ that guarantees exponential stability of the plant dynamics under persistent attacks. In particular, by (12), we obtain that the persistent rejection property is dependent on the smallest eigenvalue of all the unstable matrices M_{σ_u} induced by the attacks in the controller. By construction, it follows that $\alpha > 0$. Moreover, the time-ratio parameter $\kappa_{u,2}$ is uniform on the initial conditions of the plant and also on the controller. This uniformity property, established via hybrid Lyapunov functions in the next section, will be instrumental for the analysis of the closed-loop system for the case when $\dot{x} \neq 0$.

Remark 6 *The restriction to a compact set K_x of the x-component of the flow and jump set is imposed only for the purpose of regularity –namely, to guarantee that $\mathcal{A}^*_{K_x}$ is compact. Similarly, the stability properties of the signal generator* (14) *are asserted with respect to the set $[0, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u$, which is forward pre-invariant and trivially attractive since, by construction of the flow and jump set, no solution can start outside $[0, N_{0,u}] \times [0, T_{0,u}] \times \Sigma_u$.*

The next result incorporates the dynamics of the plant into the stability analysis of the closed-loop system. In this case, we provide a novel upper bound for the gain of the controller, which guarantees exponential stability of the optimal point $z^* = (x^*, u^*_w)$ for the case when $\dot{w} = 0$.

Proposition 2 *Suppose that Assumptions 1-4 hold, and let $\kappa_{u,2}$ satisfy the bound* (16). *If the gain ε satisfies:*

$$0 < \varepsilon < \varepsilon^* := \frac{\rho \underline{\lambda}(R)}{2\ell_y \max\{\bar{M}, 1\} \|C\| \|G\| \|PA^{-1}B\| (\rho + 2\max\{\bar{M}, 1\}\mu)}, \tag{18}$$

where $\rho, \tau_0 > 0$. Then, for each admissible and constant w, the set

$$\mathcal{A}^* = \{(x, u, \tau_{u,1}, \tau_{u,2}, \sigma_u) : x = x^*, u = u^*_w,\ \tau_{u,1} \in [0, N_{0,u}], \tau_{u,2} \in [0, T_{0,u}], \sigma_u \in \Sigma_u\}, \tag{19}$$

is UGES for the HDS (15).

The result of Proposition 2 establishes a solution to Problem 1 under constant exogenous signals w. Note that the result is actually uniform in w, i.e., the time ratio $\kappa_{u,2}$ and the gain ε_2 are independent of w. Indeed, as we will show below, they are also independent of $|\dot{w}|$, which is a critical property needed to avoid vanishing gains and safety margins under highly oscillating exogenous signals.

Given that by Lemma 2 the HDS (15) is well-posed, the following additional robustness result can be asserted for the closed-loop system under attacks *and* small disturbances e acting on the states and dynamics. In general, such type of disturbances are unavoidable in practical applications due to measurements noise, numerical approximations, etc.

Lemma 3 *Consider the HDS* (15) *and suppose that the conditions of Propositions 1 and 2 hold for the time-ratio parameter $\kappa_{u,2}$ and the gain $\varepsilon > 0$. Then, for each constant w there exists $c_1, c_2 > 0$ such that for each compact set of initial conditions $K \subset \mathbb{R}^n \times \mathbb{R}^m$, and for each $\delta > 0$, there exists $\bar{e} > 0$ such that for any measurable disturbance e with $\sup_{t+j \geq 0} |e(t, j)| \leq \bar{e}$ and every initial condition $\vartheta(0, 0) \in K \times [0, N_{0,u}] \times [0, T_{0,u}] \times \Sigma$, the trajectories of the perturbed HDS*

$$\vartheta + e \in C, \quad \dot{\vartheta} \in F_{\sigma_u}(\vartheta + e) + e, \tag{20a}$$

$$\vartheta + e \in D, \quad \vartheta^+ \in G_{\sigma_u}(\vartheta + e) + e, \tag{20b}$$

satisfy the bound

$$|z(t, j) - z^*| \leq c_1 |z(0, 0) - z^*| e^{-c_2(t+j)} + \delta, \quad \forall\ (t, j) \in dom(\vartheta), \tag{21}$$

The robustness result of Lemma 3 is not trivial. Indeed, for general hybrid dynamical systems it is difficult to guarantee closeness of solutions between nominal and perturbed dynamics, even when the perturbations are arbitrarily small. On the other hand, when the HDS satisfies the regularity conditions of Lemma 2, stability properties turn out to be robust to small disturbances [27, Ch.7]. This idea is at the core of our modeling framework (15), which subsumes the closed-loop system under attack as a well-posed hybrid dynamical system with sufficiently slow control dynamics.

By leveraging the results of Propositions 1 and 2, we can now state the first *main* result of this paper, which establishes uniform global exponential input-to-state stability (E-ISS) for the closed-loop system under attacks and time-varying exogenous inputs or disturbances w. In particular, under the time ratio established in Proposition 1, and the maximal gain established in Proposition 2, we provide an explicit characterization of the ISS gain γ that maps $\|\dot{w}(t)\|_t$ to the radius of the residual set where the trajectories converge.

Theorem 1 *Let $\mathcal{A}^*$ be given by (19). Suppose that Assumptions 1-4 hold, and that the conditions (16) and (18) are also satisfied for the pair $(\kappa_{u,2}, \varepsilon)$. Then, the HDS (15) renders E-ISS the set $\mathcal{A}^*$ with respect to $|\dot{w}|$, with linear asymptotic gain γ given by*

$$\gamma\left(\|\dot{w}(t)\|_t\right) := \frac{1}{\varepsilon}\left[\left(\frac{\max\{\theta\overline{\lambda}(P), (1-\theta)e^{\tau_0}\ell/2\}}{\min\{\theta\underline{\lambda}(P), (1-\theta)\mu/2\}}\right)^{1/2}\frac{\|r\|}{k\min\{\mu^2, 1\}}\right]\|\dot{w}(t)\|_t, \tag{22}$$

where k and τ_0 are positive constants, and $r := [2\theta\|PA^{-1}B\|, \ell_y(1-\theta)e^{\tau_0}\|H\|\|G\|]^\top$ with $\theta := \frac{\ell_y\|G\|\|C\|}{\ell_y\|G\|\|C\|+2e^{-\tau_0}\|PA^{-1}B\|}$.

Remark 7 *The gain of the function γ in (22) reflects explicitly the stability of the plant through the eigenvalues of the matrix P, and the qualitative behaviour of the cost function given by the Lipschitz, and PL constant ℓ, and μ, respectively. Additionally, this gain is weighted by $1/\varepsilon$, which implies that a larger time scale separation between the dynamics of the plant and the controller leads to larger tracking errors. This behavior is expected since decreasing ε effectively "slows down" the controller, leading to a detrimental effect on its tracking performance.*

Remark 8 *When $\bar{\lambda}(P) >> 1$ and $\underline{\lambda}(P) << 1$, the asymptotic gain simplifies to*

$$\gamma\left(\|\dot{w}(t)\|_t\right) := \frac{1}{\varepsilon}\left[\sqrt{cond(P)}\frac{\|r\|}{k\min\{\mu^2, 1\}}\right]\|\dot{w}(t)\|_t, \qquad cond(P) := \frac{\bar{\lambda}(P)}{\underline{\lambda}(P)}, \tag{23}$$

which shows the effect of the condition number of the Lyapunov matrix P on the residual set where the trajectories converge.

Remark 9 *In (22), the constant τ_0 comes from the Lyapunov-based analysis used to study the HDS (15). This is an upper bound for the weighted time that the closed-loop system (8) can flow under attacks. On the other hand, constant k is upper bounded by the smallest eigenvalue of a matrix that relates the interconnected system. A detailed characterization of these constants is presented in Section 5.*

4.2 Attacks to the Static Feedback

In the previous section, we focused on attacks acting on the dynamic controller. In this section, we now turn our attention to study the effect of persistent attacks acting on the static feedback controller (10). To study this scenario, we consider the following set-valued HDS with state $\vartheta := (x, u, \tau_{v,1}, \tau_{v,2}, \sigma_v)$, which corresponds to the interconnection of the plant dynamics under attacks (10), the nominal dynamic gradient-based controller (9), and the signal generator (14):

$$\vartheta \in C := \mathbb{R}^n \times \mathbb{R}^m \times [0, N_{0,v}] \times [0, T_{0,v}] \times \Sigma_v, \tag{24a}$$

$$\begin{pmatrix} \dot{x} \\ \dot{u} \\ \dot{\tau}_{v,1} \\ \dot{\tau}_{v,2} \\ \dot{\sigma}_v \end{pmatrix} \in F_{\sigma_v}(\vartheta, w) := \begin{pmatrix} A_{\sigma_v}x + B_{\sigma_v}u + E_{\sigma_v}w \\ -\varepsilon(\nabla f_u(u) + G^\top \nabla f_y(y)) \\ [0, \kappa_{v,1}] \\ [0, \kappa_{v,2}] - I_{\sigma_{v,a}}(\sigma_v) \\ 0 \end{pmatrix}, \quad y = Cx, \tag{24b}$$

$$\vartheta \in D := \mathbb{R}^n \times \mathbb{R}^m \times [1, N_{0,v}] \times [0, T_{0,v}] \times \Sigma_v, \tag{24c}$$

$$\begin{pmatrix} x^+ \\ u^+ \\ \tau_{v,1}^+ \\ \tau_{v,2}^+ \\ {\sigma_v}^+ \end{pmatrix} \in G_{\sigma_v}(\vartheta) := \begin{pmatrix} x \\ u \\ \tau_{v,1} - 1 \\ \tau_{v,2} \\ \Sigma_v \backslash \{\sigma_v\} \end{pmatrix}, \tag{24d}$$

where $T_{0,v} \geq 0$, $\kappa_{v,1} > 0$, and $\kappa_{v,2} \in (0, 1)$ is a time-ratio parameter that describes the persistent rejection property of the defensive mechanism. Note that when $\sigma_v = s$, we have $A_s = A$, hence, Assumption 1 holds for A_s. Moreover, to simplify notation we denote $A_{a_i} := F + NL_{a_i}KC$ for any $a_i \in \Sigma_{v,a}$. By construction, system (24) also satisfies the properties of Lemma 2.

For attacks acting on the plant, we introduce an additional technical assumption to guarantee that the switching system has a well-defined unique equilibrium point.

Assumption 5 *For any* $w \in \mathbb{R}^q$ *there exists* $\bar{x} \in \mathbb{R}^n$ *and* $\bar{u} \in \mathbb{R}^m$ *such that* $0 = A_\sigma \bar{x} + B_\sigma \bar{u} + E_\sigma w,\ \forall \sigma \in \Sigma$.

Remark 10 *The conditions of Assumption 5 assert that the attacks to the plant modify the stability of the equilibria, but they do not alter the set of equilibria [13]. We note that this assumption is aligned with existing works in context [13, 20], and it enables the use of analytical tools based on Lyapunov theory for HDS. Without this assumption in place, each attack could generate a different equilibrium point (if any), and to assert stability properties we will need to introduce additional restrictive assumptions on the frequency of the attacks, e.g., condition* (13) *with a "sufficiently small" constant* $\kappa_{v,1}$. *The study of online optimization problems with multiple equilibria (even without attacks) is an open problem that is out of the scope of this paper.*

In contrast to the analysis of the previous section, we now first consider the plant under attacks, operating with a constant controller command u (i.e., $\dot{u} = 0$) and a fixed reference w (i.e., $\dot{w} = 0$). Thus, we study the stability properties of the equilibrium point $x^*(u, w) := A_{\sigma_v}^{-1} B_{\sigma_v} u + A_{\sigma_v}^{-1} E_{\sigma_v} w$. For this reduced system under attacks, we establish the following characterization of the time-ratio parameter $\kappa_{v,2}$ to guarantee uniform global exponential stability.

Proposition 3 *Suppose that Assumptions 1-5 hold, and that the time-ratio parameter* $\kappa_{v,2}$ *satisfies*

$$0 < \kappa_{v,2} < \frac{1}{1+\alpha}, \qquad \text{with} \qquad \alpha := \frac{\overline{\lambda}(\hat{R}_a)\overline{\lambda}(P)}{\underline{\lambda}(R)\underline{\lambda}(P)} > 0, \tag{25}$$

where P, R *satisfy Assumption 1, and* $\overline{\lambda}(\hat{R}_a)$ *is a constant derived from Assumption 4. Then, if* $\varepsilon = 0$, *for each admissible and constant* w, *and each compact set* $K_u \subset \mathbb{R}^m$, *the set*

$$\mathcal{A}^*_{K_u} = \{(x, u, \tau_{v,1}, \tau_{v,2}, \sigma_v) : x = x^*(u, w),\ u \in K_u,\ \tau_{v,1} \in [0, N_{0,v}], \tau_{v,2} \in [0, T_{0,v}], \sigma_v \in \Sigma_v\},$$

is UGES for the HDS (24) *with restricted flow and jump sets*

$$C_{K_u} = \mathbb{R}^n \times (\mathbb{R}^m \cap K_u) \times [0, N_{0,v}] \times [0, T_{0,v}] \times \Sigma_v, \qquad D_{K_u} = \mathbb{R}^n \times (\mathbb{R}^m \cap K_u) \times [1, N_{0,v}] \times [0, T_{0,v}] \times \Sigma_v. \tag{26}$$

The result of Proposition 3 establishes an explicit characterization of the level of persistent rejection (c.f. Definition 5) needed to guarantee exponential stability in the system. For the internal feedback loop controller, this rejection capability is related to the largest and smallest eigenvalues of the matrices P and $\hat{R}_a$.

By leveraging the result of Proposition 3, we can now state a stability and convergence result for the case where the controller is activated, and the disturbance w remains static.

Proposition 4 *Suppose that Assumptions 1-5 hold, and let* $\kappa_{v,2}$ *satisfy the bounds in Proposition 3. If* ε *satisfies:*

$$0 < \varepsilon < \varepsilon^* := \frac{\rho \underline{\lambda}(P)}{4\ell_y e^{\tau_0} \|C\| \|G\| \|PA^{-1}B\|}, \tag{27}$$

where ρ, *and* τ_0 *are positive constants, and* P *is defined in Assumption 1 then, for any admissible and constant* w *the set*

$$\mathcal{A}^* = \{(x, u, \tau_{v,1}, \tau_{v,2}, \sigma_v) : x = x^*(u^*, w), u = u^*,\ \tau_{v,1} \in [0, N_{0,v}], \tau_{v,2} \in [0, T_{0,v}], \sigma_v \in \Sigma_v\},$$

is UGES for the HDS (24).

Remark 11 *The result of Proposition 4 establishes a sufficient condition on the bounds of the gain of the controller and the time-ratio parameter* $\kappa_{v,2}$ *to obtain exponential convergence to the solution of Problem* (1) *using a gradient-based controller under persistent attacks. Since the HDS is well-posed the robustness results of Lemma 3 also hold for system* (24).

Attack action on:	Time-Ratio Factor: α	Controller Gain: ε^*	ISS Gain: $\gamma(\lvert\lvert\dot{w}\rvert\rvert_t)$
Dynamic Feedback	$-\min_{\sigma_u\in\Sigma_{u,a}}\underline{\lambda}(M_{\sigma_u})$	$\frac{\rho\underline{\lambda}(R)}{2\ell_y\max\{\bar{M},1\}\lvert\lvert C\rvert\rvert\lvert\lvert G\rvert\rvert\lvert\lvert PA^{-1}B\rvert\rvert(\rho+2\max\{\bar{M},1\}\mu)}$	$\frac{1}{\varepsilon}\left[\left(\frac{\max\{\theta\bar{\lambda}(P),(1-\theta)e^{\tau_0}\ell/2\}}{\min\{\theta\underline{\lambda}(P),(1-\theta)\mu/2\}}\right)^{1/2}\frac{\lvert\lvert r\rvert\rvert}{k\min\{\mu^2,1\}}\right]\lvert\lvert\dot{w}(t)\rvert\rvert_t$
Static Feedback	$\frac{\bar{\lambda}(\hat{R}_a)\bar{\lambda}(P)}{\underline{\lambda}(R)\underline{\lambda}(P)}$	$\frac{\rho\underline{\lambda}(P)}{4\ell_y e^{\tau_0}\lvert\lvert C\rvert\rvert\lvert\lvert G\rvert\rvert\lvert\lvert PA^{-1}B\rvert\rvert}$	$\frac{1}{\varepsilon}\left[\left(\frac{\max\{\theta\bar{\lambda}(P)e^{\tau_0},(1-\theta)\ell/2\}}{\min\{\theta\underline{\lambda}(P),(1-\theta)\mu/2\}}\right)^{1/2}\frac{\lvert\lvert r\rvert\rvert}{k\min\{\mu^2,1\}}\right]\lvert\lvert\dot{w}(t)\rvert\rvert_t$

Figure 2: Summary of the results of Propositions 1-4, and Theorems 1-2. The time ratio parameters are given by $1/1+\alpha$.

By leveraging the results of Proposition 3 and Proposition 4, we now establish the second main result of this paper, which asserts exponential input-to-state stability of the closed-loop system under attacks, with respect to the time-variation of w. In conjunction with Theorem 1, this result provides a complete answer to Problem 1.

Theorem 2 *Suppose that Assumptions 1-5 hold, and that conditions* (25) *and* (27) *also hold. Then, the HDS* (24) *renders the set $\mathcal{A}^*$ E-ISS with respect to $|\dot{w}|$, with linear asymptotic gain γ given by*

$$\gamma\left(\|\dot{w}(t)\|_t\right) := \frac{1}{\varepsilon}\left[\left(\frac{\max\{\theta\bar{\lambda}(P)e^{\tau_0},(1-\theta)\ell/2\}}{\min\{\theta\underline{\lambda}(P),(1-\theta)\mu/2\}}\right)^{1/2}\frac{\|r\|}{k\min\{\mu^2,1\}}\right]\|\dot{w}(t)\|_t, \tag{28}$$

where k and τ_0 are positive constants, $r := [2e^{\tau_0}\theta\|PA^{-1}B\|, \ell_y(1-\theta)\|H\|\|G\|]^\top$ with $\theta := \frac{\ell_y\|G\|\|C\|}{\ell_y\|G\|\|C\|+2e^{\tau_0}\|PA^{-1}B\|}$.

Remark 12 *As in Theorem 1, the asymptotic gain* (28) *relates the magnitude of $\dot{w}$ to the size of the residual set where the trajectories converge. Note that in this case the coefficient e^{τ_0} appears next to the the values and gains related to the plant. Therefore, when $\ell >> 1$ and $\mu << 1$ (i.e., the cost function f is ill-conditioned) the asymptotic gain simplifies to*

$$\gamma\left(\|\dot{w}(t)\|_t\right) := \frac{1}{\varepsilon}\left[\sqrt{cond(f)}\frac{\|r\|}{k\min\{\mu^2,1\}}\right]\|\dot{w}(t)\|_t, \qquad cond(f) = \frac{\ell}{\mu}, \tag{29}$$

which shows the effect of the condition number of $u \mapsto f(u)$ on the residual set where the trajectories converge. Note that a cost function f that is ill-conditioned can be identified by examining the eccentricity of the sub-level sets.

We summarize in Table 2 the main results of this section. Each row indicates which subsystem is under attack, and the columns summarize the upper bounds on the time ratios and gains, and also the form of the ISS gain.

5 Analysis

In this section, we present the analysis and the proofs of Theorems 1 and 2. In both cases, our analysis leverages the results of Lemma 4 in the Appendix. The structure of the proofs will follow four main steps:

1. For each of the feedback loops under attack, we will introduce an auxiliary Lyapunov-like function for which a strong exponential decrease can be asserted whenever the feedback loop operates in the stable mode, and the dynamics of the other loop are neglected.
2. When the feedback loop is under attacks, we incorporate the switching behavior into the system by extending the auxiliary Lyapunov-like function with a state that depends on the hybrid signal generator. To establish suitable stability properties for the switched system under attacks we make use of Lemma 4 in the Appendix.

3. We use the extended auxiliary functions constructed in the previous step, to construct a new Lyapunov function for the complete hybrid system, and we characterize the upper bound ε^* on the gain of the controller needed to guarantee exponential decrease of the function during flows outside of the compact set of interest. We also show that the Lyapunov function does not increase during jumps.

4. The previous argument, in addition to the average dwell-time constraint, allows us to establish uniform global exponential stability and/or uniform global exponential ISS.

To model the exogenous input w acting on the system, we will assume that w is generated by an exosystem of the form $\dot{w} = F_w(w)$, which evolves in the compact set $\Lambda = \lambda\mathbb{B}$. For simplicity we assume that F_w is continuous, and that every trajectory w generated by the exosystem is complete. We assume that during a jump of the system the signal w does not change, i.e., $w^+ = w$. This model is quite standard in the literature of nonlinear systems with exogenous inputs.

5.1 Analysis of System with Dynamic Feedback Controller under Attacks

Consider the closed-loop system with attacks on the gradient-flow controller (15), modeled by the HDS (15). Define the affine map $(u,w)\mapsto \bar{x}(u,w)$ as $\bar{x}(u,w) := -A^{-1}Bu - A^{-1}Ew$, and consider the change of variables $x_e := x-\bar{x}$. In the new variables, the dynamics can be expressed as:

$$\dot{x}_e = Ax_e + A^{-1}B\dot{u} + A^{-1}E\dot{w}, \qquad \dot{u} = -\varepsilon M_{\sigma_u}\varphi(u,w,x_e) = -\varepsilon M_{\sigma_u}\left(\nabla f_u(u) + G^{\mathsf{T}}\nabla f_y(Cx_e+Gu+Hw)\right),$$
$$G := -CA^{-1}B, \qquad H := -CA^{-1}E,$$

where we note that $-\varphi(u,w,0) = -\nabla f(u,w) = -\nabla f_u(u) - G^{\mathsf{T}}\nabla f_y(Gu+Hw)$. First, we investigate the stability properties of the "stable mode" of the system by leveraging the stability properties of the gradient flow. Consider the following auxiliary function:

$$V(u,w) = f(u,w) - f(u^*),$$

where we recall that $f(u,w) = f_u(u) + f_y(Gu+Hw)$. For $\vartheta\in D$, during the jumps it holds that $V(u^+,w^+) = V(u,w)$. Moreover, for $\vartheta\in C$, $\dot{V}(u,w)$ can be bounded as follows during the flows:

$$\begin{aligned}
\dot{V}(u,w) &= -\varepsilon\nabla f(u,w)^\top M_{\sigma_u}\varphi(u,w,x_e) + H^\top(\nabla f_y(Gu+Hw) - \nabla f_y(Gu^*+Hw))\dot{w}\\
&= -\varepsilon\varphi(u,w,0)^\top M_{\sigma_u}(\varphi(u,w,0) + \varphi(u,w,x_e) - \varphi(u,w,0)) + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|\\
&\le -\varepsilon\underline{\lambda}(M_{\sigma_u})\|\nabla f(u,w)\|^2 + \varepsilon\ell_y\|C\|\|G\|\|M_{\sigma_u}\|\|\nabla f(u,w)\|\|x_e\| + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|\\
&\le -2\mu\varepsilon\underline{\lambda}(M_{\sigma_u})V(u,w) + \varepsilon\ell_y\|C\|\|G\|\|M_{\sigma_u}\|\|\nabla f(u,w)\|\|x_e\| + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|\\
&\le -2\mu\varepsilon\underline{\lambda}(M_{\sigma_u})V(u,w) + \varepsilon\ell_y\max\{\bar{M},1\}\|C\|\|G\|\|\nabla f(u,w)\|\|x_e\| + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|.
\end{aligned}$$

Next, to study the stability properties of the controller under switching signals with unstable modes, we consider the following auxiliary function:

$$\bar{V}(u,w,\tau_{u,1},\tau_{u,2}) := V(u,w)e^{\tau},\ \tau := \ln(\omega)\tau_{u,1} + \tau_{u,2}(\rho_s+\rho_a),\ \tau_0 := T_{0,u}(\rho_s+\rho_a),\ \rho_s = 2\mu,\ \rho_a = -2\mu\min_{\sigma_u\in\Sigma_{u,a}}\underline{\lambda}(M_{\sigma_u}),$$

where we recall that μ is the coefficient in the PL inequality. Since we will analyze all the modes of the system using the same Lyapunov function, we can set $\omega=1$; see Lemma 4 in the Appendix. Now, during jumps the function $\bar{V}(u,w,\tau_{u,1},\tau_{u,2})$ satisfies $\bar{V}(u^+,w^+,\tau_{u,1}^+,\tau_{u,2}^+) = \bar{V}(u,w,\tau_{u,1},\tau_{u,2})$, and during flows it satisfies:

$$\begin{aligned}
\dot{\bar{V}}(u,w,\tau_{u,1},\tau_{u,2}) &= \dot{V}(u,w)e^{\tau} + V(u,w)e^{\tau}\dot{\tau}\\
&= \dot{V}(u,w)e^{\tau} + V(u,w)e^{\tau}(\rho_s+\rho_u)\dot{\tau}_{u,2}\\
&\le \dot{V}(u,w)e^{\tau} + \varepsilon V(u,w)e^{\tau}(\rho_s+\rho_u)\kappa_{u,2}\\
&\le -2\mu\varepsilon V(u,w)e^{\tau} + \varepsilon\ell_y\max\{\bar{M},1\}\|C\|\|G\|\|\nabla f(u,w)\|\|x_e\|e^{\tau} + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|e^{\tau}\\
&\quad + \varepsilon V(u)e^{\tau}(\rho_s+\rho_u)\kappa_{u,2}\\
&= -\varepsilon\rho\bar{V}(u) + \varepsilon\ell_y\max\{\bar{M},1\}\|C\|\|G\|\|\nabla f(u,w)\|\|x_e\|e^{\tau} + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|e^{\tau}\\
&\le -\frac{\varepsilon\rho}{2\mu}\|\nabla f(u,w)\|^2e^{\tau_0} + \varepsilon\ell_y\max\{\bar{M},1\}\|C\|\|G\|\|\nabla f(u,w)\|\|x_e\|e^{\tau_0} + \ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|e^{\tau_0},
\end{aligned}$$

where $\rho = \rho_s - \kappa_{u,2}(\rho_s + \rho_a) > 0$, hence $\kappa_{u,2} < \frac{\rho_s}{\rho_s+\rho_a} = \frac{1}{1+\alpha}$, and $\alpha := -\min_{\sigma_u \in \Sigma_{u,a}} \underline{\lambda}(M_{\sigma_u})$. Notice that $\alpha \geq 0$ due to the definition of M_{a_i}.

We now analyze the interconnection of the plant dynamics and the switched controller. To do this, we first consider the auxiliary function $W(x_e) = x_e^\top P x_e$, where $P \succ 0$ and satisfies $A^\top P + PA \preceq -R$ for $R \succ 0$ due to Assumption 1. Along the trajectories of the plant dynamics the time derivative of W satisfies the following inequalities:

$$
\begin{aligned}
\dot{W}(x_e) &= \dot{x}_e^\top P x_e + x_e^\top P \dot{x}_e \\
&= 2x_e^\top P(Ax_e + A^{-1}B\dot{u} + A^{-1}E\dot{w}) \\
&= x_e^\top (A^\top P + PA)x_e + 2x_e^\top \left(PA^{-1}B\dot{u}\right) + 2x_e^\top \left(PA^{-1}B\dot{w}\right) \\
&\leq -x_e^\top R x_e - 2\varepsilon x_e^\top \left(PA^{-1}BM_{\sigma_u}\varphi(u,w,x_e)\right) + 2x_e^\top \left(PA^{-1}B\dot{w}\right) \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\|PA^{-1}B\|\|M_{\sigma_u}\|\|\varphi(u,w,x_e)\|\|x_e\| + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\|PA^{-1}B\|\|M_{\sigma_u}\|\|x_e\|\|\varphi(u,w,0) + \varphi(u,w,x_e) - \varphi(u,w,0)\| + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\|PA^{-1}B\|\|M_{\sigma_u}\|\|x_e\|(\|\varphi(u,w,0)\| + \|\varphi(u,w,x_e) - \varphi(u,w,0)\|) + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\|PA^{-1}B\|\|M_{\sigma_u}\|\|x_e\|(\|\nabla f(u,w)\| + \ell_y\|C\|\|G\|\|x_e\|) + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \\
&= -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\ell_y\|PA^{-1}B\|\|M_{\sigma_u}\|\|C\|\|G\|\|x_e\|^2 + 2\varepsilon\|PA^{-1}B\|\|M_{\sigma_u}\|\|x_e\|\|\nabla f(u,w)\| \\
&\quad + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\ell_y\max\{\bar{M},1\}\|PA^{-1}B\|\|C\|\|G\|\|x_e\|^2 + 2\varepsilon\max\{\bar{M},1\}\|PA^{-1}B\|\|x_e\|\|\nabla f(u,w)\| \\
&\quad + 2\|PA^{-1}B\|\|x_e\|\|\dot{w}\|.
\end{aligned}
$$

Next, we consider the following Lyapunov function for the complete HDS (15):

$$U(\vartheta) = (1-\theta)\bar{V}(u,w,\tau_{u,1},\tau_{u,2}) + \theta W(x_e), \qquad \theta \in (0,1). \tag{30}$$

This function is bounded from below and above as follows:

$$\min\left\{\theta\underline{\lambda}(P), \frac{\mu}{2}(1-\theta)\right\}|z|^2_{\mathcal{A}^*} \leq U(\vartheta) \leq \max\left\{\theta\bar{\lambda}(P), \frac{\ell}{2}(1-\theta)e^{\tau_0}\right\}|z|^2_{\mathcal{A}^*}, \quad \forall\, \vartheta \in C \cup D, \tag{31}$$

where we used the fact that $\tau_{u,1} \in [0, N_{0,u}]$, $\tau_{u,2} \in [0, T_{0,u}]$, and $\sigma_u \in \Sigma_u$ at all times. Moreover, for any $\vartheta \in D$, it follows that $U(\vartheta^+) = U(\vartheta)$. Also, for any $\vartheta \in C$, the time derivative of $\dot{U}$ can be upper bounded as follows:

$$
\begin{aligned}
\dot{U}(\vartheta) &= (1-\theta)\dot{\bar{V}}(u,w,\tau_1,\tau_2) + \theta\dot{W}(x_e) \\
&\leq -(1-\theta)\frac{\varepsilon\rho}{2\mu}\|\nabla f(u,w)\|^2 e^{\tau_0} + \varepsilon\ell_y(1-\theta)\max\{\bar{M},1\}\|C\|\|G\|\|\nabla f(u,w)\|\|x_e\|e^{\tau_0} \\
&\quad + \ell_y(1-\theta)\|H\|\|G\|\|u-u^*\|\|\dot{w}\|e^{\tau_0} - \theta\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\ell_y\theta\max\{\bar{M},1\}\|PA^{-1}B\|\|C\|\|G\|\|x_e\|^2 \\
&\quad + 2\varepsilon\theta\max\{\bar{M},1\}\|PA^{-1}B\|\|x_e\|\|\nabla f(u,w)\| + 2\theta\|PA^{-1}B\|\|x_e\|\|\dot{w}\| \;.
\end{aligned}
$$

Let $\xi := [\|x_e\|, \|\nabla f(u,w)\|]^\top$ and $\zeta := [\|x_e\|, \|u-u^*\|]^\top$; using these definitions, we obtain:

$$\dot{U}(\vartheta) \leq -\varepsilon\xi^\top\Xi\xi + r^\top\zeta\|\dot{w}\|, \tag{32}$$

where $r := [2\theta\|PA^{-1}B\|, \ell_y(1-\theta)e^{\tau_0}\|H\|\|G\|]^\top$, and Ξ is a symmetric matrix of the form

$$\Xi = \begin{bmatrix} \theta\left(\frac{\alpha}{\varepsilon} - \beta\right) & -\frac{1}{2}\left((1-\theta)\delta + \theta\chi\right) \\ -\frac{1}{2}\left((1-\theta)\delta + \theta\chi\right) & (1-\theta)\gamma \end{bmatrix}. \tag{33}$$

with $\alpha := \underline{\lambda}(R)$, $\beta := 2\ell_y\max\{\bar{M},1\}\|PA^{-1}B\|\|C\|\|G\|$, $\delta := \ell_y e^{\tau_0}\max\{\bar{M},1\}\|G\|\|C\|$, $\chi := 2\max\{\bar{M},1\}\|PA^{-1}B\|$, $\gamma := \frac{\rho e^{\tau_0}}{2\mu}$, $\theta := \frac{\delta}{\delta+\chi}$, and with ε satisfying (18).

The last step is to show that the quadratic term on the right-hand-side of (32) dominates the linear term. Using the PL inequality, it holds that $\|\xi\|^2 \geq \mu^2\|u-u^*\|^2 + \|x_e\|^2 \geq \min\{\mu^2,1\}\|\zeta\|^2$. It follows that for $0 < k \leq \underline{\lambda}(\Xi)$, (32) can be further upper bounded as follows:

$$\begin{aligned}
\dot{U}(\vartheta) &\leq -\varepsilon\xi^\top(\Xi - kI)\xi - \varepsilon k\|\xi\|^2 + r^\top\zeta\|\dot{w}\| \\
&\leq -\varepsilon\xi^\top(\Xi - kI)\xi - \varepsilon k\min\{\mu^2,1\}\|\zeta\|^2 + \|r\|\|\zeta\|\|\dot{w}\| \\
&\leq -\varepsilon\underline{\lambda}(\Xi - kI)\|\xi\|^2 = -\varepsilon\underline{\lambda}(\Xi - kI)|\vartheta|^2_{\mathcal{A}^*}, \quad \forall\ |z|_{\mathcal{A}^*} \geq \frac{\|r\|\|\dot{w}(t)\|_t}{\varepsilon k\min\{\mu^2,1\}}
\end{aligned} \tag{34}$$

which establishes the result by Lemma 4. ∎

5.2 Analysis of System with Static Feedback Controller under Attacks

In this section, we show an analytical derivation of the results for the stability of the system (24), which models the interconnected system with attacks to the inner control loop.

First, we analyze the stability properties of the ""stable mode", namely, $\sigma_v = s$. We define the affine mapping $(u,w) \mapsto \bar{x}(u,w)$ given by $\bar{x} = -A_s^{-1}B_s u - A_s^{-1}E_s w$; furthermore, with the change of variable $x_e = x - \bar{x}$, one arrives at the system:

$$\begin{aligned}
&\dot{x}_e = A_s x_e + A_s^{-1}B_s\dot{u} + A_s^{-1}E_s\dot{w}, \qquad \dot{u} = -\varepsilon\varphi(u,x_e) := -\varepsilon(\nabla f_u(u) + G_s^\top\nabla f_y(Cx_e + G_s u + H_s w)) \\
&G_s := -CA_s^{-1}B_s, \qquad H_s := -CA_s^{-1}E_s,
\end{aligned}$$

where $-\varphi(u,0) = -\nabla f(u,w) = -\nabla f_u(u) - G_s^\top\nabla f_y(G_s u + H_s w)$. For this system consider the auxiliary function $W := x_e^\top P x_e$, where $P \succ 0$ satisfies the algebraic condition $A_s^\top P + PA_s = -R$ for a given matrix $R \succ 0$. For $\vartheta \in D$, during the jumps we have that $W(x_e^+) = W(x_e)$. On the other hand, for $\vartheta \in C$, the derivative of W along the trajectory of the system during flows reads:

$$\begin{aligned}
\dot{W} &= \dot{x}_e^\top P x_e + x_e^\top P\dot{x}_e \\
&= \left(A_s x_e + A_s^{-1}B_s\dot{u} + A_s^{-1}E_s\dot{w}\right)^\top Px_e + x_e^\top P\left(A_s x_e + A_s^{-1}B_s\dot{u} + A_s^{-1}E_s\dot{w}\right) \\
&= x_e^\top(A_s^\top P + PA_s)x_e + 2x_e^\top P(A_s^{-1}B_s\dot{u}) + 2x_e^\top P(A_s^{-1}E_s\dot{w}) \\
&= -x_e^\top Rx_e - 2\varepsilon x_e^\top PA_s^{-1}B_s(\varphi(u,x_e) + \varphi(u,0) - \varphi(u,0)) + 2x_e^\top P(A_s^{-1}E_s\dot{w}) \\
&\leq -x_e^\top Rx_e + 2\varepsilon\|x_e\|\|PA_s^{-1}B_s\|\|\varphi(u,0) + (\varphi(u,x_e) - \varphi(u,0))\| + 2x_e^\top P(A_s^{-1}E_s\dot{w}) \\
&\leq -x_e^\top Rx_e + 2\varepsilon\|x_e\|\|PA_s^{-1}B_s\|(\|\varphi(u,0)\| + \|\varphi(u,x_e) - \varphi(u,0)\|) + 2x_e^\top P(A_s^{-1}E_s\dot{w}) \\
&\leq -\underline{\lambda}(R)\|x_e\|^2 + 2\varepsilon\|PA_s^{-1}B_s\|\|x_e\|\|\nabla f(u,w)\| + 2\varepsilon\ell_y\|PA_s^{-1}B_s\|\|C\|\|G_s\|\|x_e\|^2 + 2\|x_e\|\|PA_s^{-1}E_s\|\|\dot{w}\| \\
&\leq -\rho_s W(x_e) + 2\varepsilon\|PA_s^{-1}B_s\|\|x_e\|\|\nabla f(u,w)\| + 2\varepsilon\ell_y\|PA_s^{-1}B_s\|\|C\|\|G_s\|\|x_e\|^2 + 2\|x_e\|\|PA_s^{-1}E_s\|\|\dot{w}\|
\end{aligned}$$

where we defined $\rho_s := \frac{\underline{\lambda}(R)}{\overline{\lambda}(P)}$.

Next, for the "unstable modes". Let $\sigma_v = a_i$, for a fixed $a_i \in \Sigma_{v,a}$. We change variable to $x_e = x - \bar{x}$ where $\bar{x} := -A_{a_i}^{-1}B_{a_i}u - A_{a_i}^{-1}E_{a_i}w$, resulting in

$$\begin{aligned}
&\dot{x}_e = A_{a_i}x_e + A_{a_i}^{-1}B_{a_i}\dot{u} + A_{a_i}^{-1}E_{a_i}\dot{w}, \qquad \dot{u} = -\varepsilon\varphi(u,x_e) := -\varepsilon(\nabla f_u(u) + G_a^\top\nabla f_y(Cx_e + G_{a_i}u + H_{a_i}w)) \\
&G_{a_i} := -CA_{a_i}^{-1}B_{a_i}, \qquad H_{a_i} := -CA_{a_i}^{-1}E_{a_i},
\end{aligned}$$

and $-\varphi(u,0) = -\nabla f(u,w) = -\nabla f_u(u) - G_a^\top\nabla f_y(G_{a_i}u + H_{a_i}w)$.

Consider again the auxiliary function $W = x_e^\top Px_e$, where $P \succ 0$ and satisfies $A_{a_i}^\top P + PA_{a_i} = \hat{R}_{a_i}$ for a given $\hat{R}_{a_i}$ as stated after Assumption 4.

For $\vartheta \in D$, during jumps we have that $W(x_e^+) = W(x_e)$. Moreover, for $\vartheta \in C$, during flows we have

$$\begin{aligned}
\dot{W} &= \dot{x}_e^\top P x_e + x_e^\top P \dot{x}_e \\
&= \left(A_{a_i} x_e + A_{a_i}^{-1} B_{a_i} \dot{u} + A_{a_i}^{-1} E_{a_i} \dot{w}\right)^\top P x_e + x_e^\top P \left(A_{a_i} x_e + A_{a_i}^{-1} B_{a_i} \dot{u} + A_{a_i}^{-1} E_{a_i} \dot{w}\right) \\
&= x_e^\top (A_{a_i}^\top P + P A_{a_i}) x_e + 2 x_e^\top P (A_{a_i}^{-1} B_{a_i} \dot{u}) + 2 x_e^\top P (A_{a_i}^{-1} E_{a_i} \dot{w}) \\
&= x_e^\top \hat{R}_{a_i} x_e - 2\varepsilon x_e^\top P A_{a_i}^{-1} B_{a_i} (\varphi(u, x_e) + \varphi(u,0) - \varphi(u,0)) + 2 x_e^\top P (A_{a_i}^{-1} E_{a_i} \dot{w}) \\
&\leq x_e^\top \hat{R}_{a_i} x_e + 2\varepsilon \|x_e\| \|P A_{a_i}^{-1} B_{a_i}\| \|\varphi(u,0) + (\varphi(u,x_e) - \varphi(u,0))\| + 2 x_e^\top P (A_{a_i}^{-1} E_{a_i} \dot{w}) \\
&\leq x_e^\top \hat{R}_{a_i} x_e + 2\varepsilon \|x_e\| \|P A_{a_i}^{-1} B_{a_i}\| (\|\varphi(u,0)\| + \|\varphi(u,x_e) - \varphi(u,0)\|) + 2 x_e^\top P (A_{a_i}^{-1} E_s \dot{w}) \\
&\leq \overline{\lambda}(\hat{R}_{a_i}) \|x_e\|^2 + 2\varepsilon \|P A_{a_i}^{-1} B_{a_i}\| \|x_e\| \|\nabla f(u,w)\| + 2\varepsilon \ell_y \|P A_{a_i}^{-1} B_{a_i}\| \|C\| \|G_{a_i}\| \|x_e\|^2 + 2\|x_e\| \|P A_{a_i}^{-1} E_{a_i}\| \|\dot{w}\| \\
&\leq \rho_{a_i} W(x_e) + 2\varepsilon \|P A_{a_i}^{-1} B_{a_i}\| \|x_e\| \|\nabla f(u,w)\| + 2\varepsilon \ell_y \|P A_{a_i}^{-1} B_{a_i}\| \|C\| \|G_{a_i}\| \|x_e\|^2 + 2\|x_e\| \|P A_{a_i}^{-1} E_{a_i}\| \|\dot{w}\|,
\end{aligned}$$

where $\rho_{a_i} := \frac{\overline{\lambda}(\hat{R}_{a_i})}{\underline{\lambda}(P)}$ for a fixed a_i.

We analyze the stability of the switched plant with the following auxiliary function

$$\begin{aligned}
\bar{W}(x_e, \tau_{v,2}) = W(x_e) e^\tau, \quad \tau = \ln(\omega)\tau_{v,1} + \tau_{v,2}(\rho_s + \rho_a), \text{ where } 0 \leq \tau \leq \tau_0 := T_{0,v}(\rho_s + \rho_a) \\
\text{and} \quad \rho_s := \frac{\underline{\lambda}(R)}{\overline{\lambda}(P)} \qquad \rho_a := \frac{\overline{\lambda}(\hat{R}_a)}{\underline{\lambda}(P)}.
\end{aligned} \tag{35}$$

By Assumption 5, $G_s = G_{a_i}$, $\|P A_s^{-1} B_s\| = \|P A_{a_i}^{-1} B_{a_i}\|$, and $\|P A_s^{-1} E_s\| = \|P A_{a_i}^{-1} E_{a_i}\| \quad \forall a_i \in \Sigma_{v,a}$, hence, we drop the subindices s and a_i. Since, we will analyze the modes using the same Lyapunov function, we set $\omega = 1$; see Lemma 4 in the Appendix.

Now, see that during jumps the function $\bar{W}(x_e, \tau_{v,1}, \tau_{v,2})$ satisfies $\bar{W}(x_e^+, \tau_{v,1}^+, \tau_{v,2}^+) = \bar{W}(x_e, \tau_{v,1}, \tau_{v,2})$, and during flows it satisfies:

$$\begin{aligned}
\dot{\bar{W}}(x_e, \tau_{v,1}, \tau_{v,2}) &= \dot{W}(x_e) e^\tau + W(x_e) e^\tau \dot{\tau} \\
&= \dot{W}(x_e) e^\tau + W(x_e) e^\tau (\rho_s + \rho_a) \dot{\tau}_v \\
&\leq \dot{W}(x_e) e^\tau + W(x_e) e^\tau (\rho_s + \rho_a) \kappa_{v,2} \\
&\leq -\rho_s W(x_e) e^\tau + 2\varepsilon \|P A^{-1} B\| \|x_e\| \|\nabla f(u,w)\| e^\tau + 2\varepsilon \ell_y e^\tau \|P A^{-1} B\| \|C\| \|G\| \|x_e\|^2 \\
&\quad + 2 e^\tau \|x_e\| \|P A^{-1} B\| \|\dot{w}\| + W(x_e) e^\tau (\rho_s + \rho_a) \kappa_{v,2} \\
&\leq -\rho \bar{W}(x_e) + 2\varepsilon \|P A^{-1} B\| \|x_e\| \|\nabla f(u,w)\| e^\tau + 2\varepsilon \ell_y e^\tau \|P A^{-1} B\| \|C\| \|G\| \|x_e\|^2 \\
&\quad + 2 e^\tau \|x_e\| \|P A^{-1} B\| \|\dot{w}\| \\
&\leq -\rho \underline{\lambda}(P) \|x_e\|^2 + 2\varepsilon \|P A^{-1} B\| \|x_e\| \|\nabla f(u,w)\| e^\tau + 2\varepsilon \ell_y e^\tau \|P A^{-1} B\| \|C\| \|G\| \|x_e\|^2 \\
&\quad + 2 e^\tau \|x_e\| \|P A^{-1} B\| \|\dot{w}\|,
\end{aligned}$$

where $\rho := \rho_s - \kappa_{v,2}(\rho_s + \rho_a) > 0$, hence $\kappa_{v,2} < \frac{\rho_s}{\rho_s + \rho_a} = \frac{1}{1+\alpha}$, and $\alpha := \frac{\overline{\lambda}(\hat{R}_a)\overline{\lambda}(P)}{\underline{\lambda}(R)\underline{\lambda}(P)} > 0$. To analyze the nominal controller (5) we considered the auxiliary function $V(u,w) = f(u,w) - f(u^*)$, then we get the following upper bound:

$$\begin{aligned}
\dot{V}(u,w) &= -\varepsilon \nabla f(u,w)^\top \varphi(x_e, u) + H^\top (\nabla f_y(Gu + Hw) - \nabla f_y(Gu^* + Hw)) \dot{w} \\
&= -\varepsilon \varphi(0,u)^\top (\varphi(0,u) + \varphi(x_e,u) - \varphi(0,u)) + \ell_y \|H\| \|G\| \|u - u^*\| \|\dot{w}\| \\
&\leq -\varepsilon \|\nabla f(u,w)\|^2 + \varepsilon \|\nabla f(u,w)\| \|G\| \|\nabla f_y(C x_e + Gu + Hw) - \nabla f_y(Gu + Hw)\| \\
&\quad + \ell_y \|H\| \|G\| \|u - u^*\| \|\dot{w}\| \\
&\leq -\varepsilon \|\nabla f(u,w)\|^2 + \varepsilon \ell_y \|G\| \|C\| \|\nabla f(u,w)\| \|x_e\| + \ell_y \|H\| \|G\| \|u - u^*\| \|\dot{w}\| \\
&\leq -\varepsilon \|\nabla f(u,w)\|^2 + \varepsilon \ell_y \|G\| \|C\| \|\nabla f(u,w)\| \|x_e\| + \ell_y \|H\| \|G\| \|u - u^*\| \|\dot{w}\|.
\end{aligned}$$

For the HDS (24), we considered the Lyapunov function $U(\vartheta) = \theta\bar{W}(x_e, \tau_{v,1}, \tau_{v,2}) + (1-\theta)V(u,w)$, with $\theta \in (0,1)$. This function is bounded as follows:

$$\min\{\theta\underline{\lambda}(P), (1-\theta)\mu/2\}|z|^2_{\mathcal{A}^*} \le U(\vartheta) \le \max\{\theta\bar{\lambda}(P)e^{\tau_0}, (1-\theta)\ell/2\}|z|^2_{\mathcal{A}^*}, \quad \forall\vartheta \in C \cup D,$$

where we used the facts that $\tau_{v,1} \in [0, N_{0,v}]$, $\tau_{v,2} \in [0, T_{0,v}]$, and $\sigma_v \in \Sigma_v$. For every $\vartheta \in D$ we have that the jumps satisfy $U(\vartheta^+) = U(\vartheta)$. Moreover, for $\vartheta \in C$ the time derivative of U is upper bounded as follows:

$$\begin{aligned}
\dot{U}(\vartheta) &= \theta\dot{\bar{W}}(x_e, \tau_{v,2}) + (1-\theta)\dot{V}(u) \\
&\le -\theta\rho\underline{\lambda}(P)\|x_e\|^2 + 2\varepsilon\theta\|PA^{-1}B\|\|x_e\|\|\nabla f(u,w)\|e^{\tau} + 2\varepsilon\ell_y e^{\tau}\theta\|PA^{-1}B\|\|C\|\|G\|\|x_e\|^2 \\
&\quad + 2e^{\tau}\theta\|x_e\|\|PA^{-1}B\|\|\dot{w}\| - \varepsilon(1-\theta)\|\nabla f(u,w)\|^2 + \varepsilon(1-\theta)\ell_y\|G\|\|C\|\|\nabla f(u,w)\|\|x_e\| \\
&\quad + (1-\theta)\ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\| \\
&\le -\theta\rho\underline{\lambda}(P)\|x_e\|^2 + 2\varepsilon\theta\|PA^{-1}B\|\|x_e\|\|\nabla f(u,w)\|e^{\tau_0} + 2\varepsilon\ell_y e^{\tau_0}\theta\|PA^{-1}B\|\|C\|\|G\|\|x_e\|^2 \\
&\quad + 2e^{\tau_0}\theta\|x_e\|\|PA^{-1}B\|\|\dot{w}\| - \varepsilon(1-\theta)\|\nabla f(u,w)\|^2 + \varepsilon(1-\theta)\ell_y\|G\|\|C\|\|\nabla f(u,w)\|\|x_e\| \\
&\quad + (1-\theta)\ell_y\|H\|\|G\|\|u-u^*\|\|\dot{w}\|.
\end{aligned}$$

Now, let $\xi := [\|x_e\|, \|\nabla f(u,w)\|]^\top$, and $\zeta := [\|x_e\|, \|u-u^*\|]^\top$; hence, we obtain

$$\dot{U}(\vartheta) \le -\varepsilon\xi^\top\Xi\xi + r^\top\zeta\|\dot{w}\|, \tag{36}$$

where $r := [2e^{\tau_0}\theta\|PA^{-1}B\|, \ell_y(1-\theta)\|H\|\|G\|]^\top$, and Ξ is a symmetric matrix as in Lemma 5 with parameters $\alpha := \rho\underline{\lambda}(P)$, $\beta := 2\ell_y e^{\tau_0}\|PA^{-1}B\|\|C\|\|G\|$, $\delta := \ell_y\|G\|\|C\|$, $\chi := 2e^{\tau_0}\|PA^{-1}B\|$, $\gamma := 1$, when $\theta := \frac{\delta}{\delta+\chi}$, and ε satisfies (27).

Finally, we show that the the quadratic term dominates the linear term in (36). Notice that by the PL inequality we get $\|\xi\|^2 \ge \mu^2\|u-u^*\|^2 + \|x_e\|^2 \ge \min\{\mu^2, 1\}\|\zeta\|^2$. Hence, for $0 < k \le \underline{\lambda}(\Xi)$, we rewrite (36) as follows:

$$\begin{aligned}
\dot{U}(\vartheta) &\le -\varepsilon\xi^\top(\Xi - kI)\xi - \varepsilon k\|\xi\|^2 + r^\top\zeta\|\dot{w}\| \\
&\le -\varepsilon\xi^\top(\Xi - kI)\xi - \varepsilon k\min\{\mu^2, 1\}\|\zeta\|^2 + \|r\|\|\zeta\|\|\dot{w}\| \\
&\le -\varepsilon\underline{\lambda}(\Xi - kI)\|\xi\|^2,
\end{aligned} \tag{37}$$

the last inequality holds when $\varepsilon k\min\{\mu^2, 1\}\|\zeta\|^2 > \|r\|\|\zeta\|\|\dot{w}\|$. Notice that $\|\zeta\| = |z|_{\mathcal{A}}$ with $\mathcal{A}^*$ as defined in Theorem 2. Hence, the inequality in (37) is satisfied for $|z|_{\mathcal{A}} \ge \frac{\|r\|\|\dot{w}(t)\|_t}{\varepsilon k\min\{\mu^2,1\}}$. Using Lemma 4 in the Appendix we obtain the E-ISS result for the HDS. ■

6 Numerical Examples

In this section, we present some numerical results that validate our theoretical contributions. We simulate a linear time-invariant system with state $x \in \mathbb{R}^2$, static control input $v \in \mathbb{R}$, dynamic control input $u \in \mathbb{R}$, and disturbance $w \in \mathbb{R}$. The goal is to regulate the solutions of the plant to the solutions of the following optimization problem

$$\min_{u,y} f_u(u) + f_y(y) := \min_{u,y} u^\top Ru + (y - y_{\text{ref}})^\top Q(y - y_{\text{ref}}), \tag{38}$$

with $R = 2$, and $Q = [1\ 0; 0\ 2]$.

6.1 Attacks on the static feedback controller

We first consider the scenario where the static controller operates under two switching attacks, and the dynamic controller operates in the nominal stable mode. The plant has the following structure:

$$\dot{x} = \begin{bmatrix} 1 & 0 \\ 2 & -1.5 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} v + \begin{bmatrix} -1.06 \\ -0.62 \end{bmatrix} u + \begin{bmatrix} -0.82 \\ -0.79 \end{bmatrix} w, \qquad y = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} x, \tag{39}$$

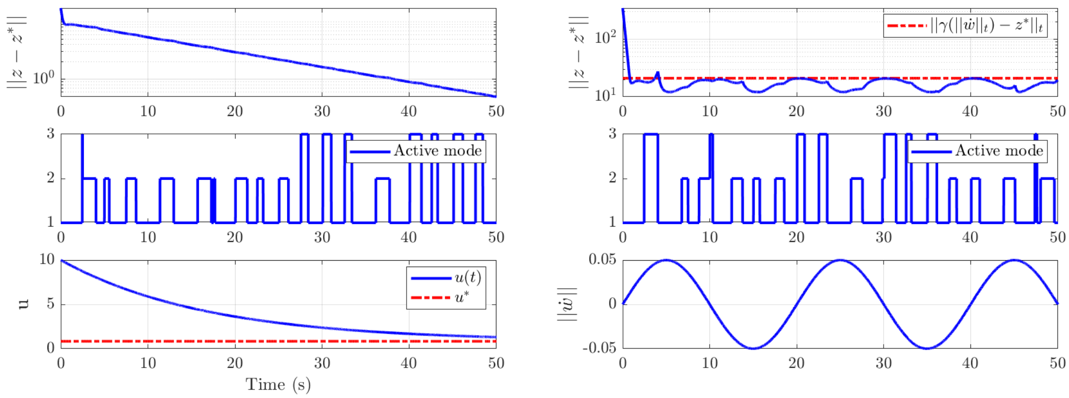

Figure 3: Attack to static feedback controller under constant and time-varying perturbations w.

For this plant, we design an internal static controller that places the poles of the closed loop system $(F + NKC)$ at -2. Hence, $K = [-40 \quad 5]$. The attacks acting on the controller are modeled by the scalars $L_{a_1} = 0$, and $L_{a_2} = -0.1$, which leads to the following matrices:

$$A_{a_1} = \begin{bmatrix} 1 & 0 \\ 2 & -1.5 \end{bmatrix} \quad B_{a_1} = \begin{bmatrix} 0.34 \\ 0.78 \end{bmatrix} \quad E_{a_1} = \begin{bmatrix} 0.30 \\ 0.34 \end{bmatrix},$$
$$A_{a_2} = \begin{bmatrix} 1.4 & -0.5 \\ 2.4 & -1.55 \end{bmatrix} \quad B_{a_2} = \begin{bmatrix} 0.48 \\ 0.92 \end{bmatrix} \quad E_{a_2} = \begin{bmatrix} 0.42 \\ 0.45 \end{bmatrix},$$

which satisfy Assumption 5. The time-ratio parameter is set as $\kappa_{v,2} = 0.365$, which induces $\alpha = 1.71$ in Proposition 3. The gain inducing the time-scale separation is set as $\varepsilon^* = 0.0149$. In our simulation, the nominal stable mode is denoted by the index 1, the unstable mode σ_{a_1} is denoted by the index 2, and σ_{a_2} is denoted by the index 3. Thus $\Sigma_v = \{1, 2, 3\}$ and $\Sigma_{v,a} = \{2, 3\}$. The theoretical value of ε^* is conservative, and to obtain faster convergence we also simulated the system with a gain of $\varepsilon = 20\varepsilon^*$. The left-hand side of Figure 3 shows the performance of the closed-loop system under a constant disturbance $w = 0.96$. As expected, the trajectories of the system converge to the set of optimal solutions. On the other hand, the the right-hand side shows the performance of the system under a time-varying disturbance modeled by $\dot{w} = a\sin(\omega t)$, with $a = 0.05$ and $\omega = 2\pi \times 0.05$. Here, we can see that the norm of the tracking error is eventually bounded by 21.

6.2 Attacks on the dynamic feedback controller

We now consider the dynamic controller operating under two switching attacks, while the nominal static controller (stable) remains free from any attack. In this case, we have the following nominal plant

$$\dot{x} = \begin{bmatrix} -3 & 0.5 \\ -2 & -1 \end{bmatrix} x + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u + \begin{bmatrix} 1 \\ 1 \end{bmatrix} w, \qquad y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x. \tag{40}$$

We simulate the system in (40) interconnected with the dynamic controller (9) under attacks. The attacks are modeled by the scalar gains $M_{a_1} = -1$, and $M_{a_2} = -2$. The time-ratio parameter is set to $\kappa_{u,2} = 0.33$, which induces $\alpha = 2$ in Proposition 1. To induce time scale separation, the gain is set as $\varepsilon^* = 0.0093$. In the simulation, the nominal stable mode is characterized by the index $\{1\}$, the attack σ_{a_1} is characterized by the mode $\{2\}$, and also σ_{a_2} is characterized by $\{3\}$. Since simulations show that the theoretical gain ε^* is very conservative, we set $\varepsilon = 20\varepsilon^*$ to achieve faster convergence in the experiments. In the left plot of Figure 4, we show the decreasing tracking error under a constant disturbance w. Similarly, in the right plot, we show the behavior of the system under a time-varying disturbance $\dot{w} = a\sin(\omega t)$ with $a = 0.05$ and $\omega = 2\pi \times 0.05$. As predicted by Theorem 2, the tracking error is eventually uniformly ultimately bounded.

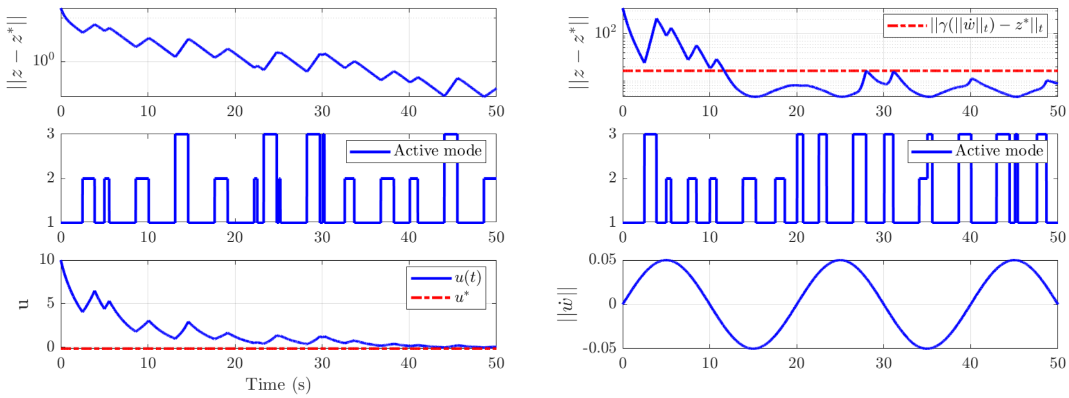

Figure 4: Attack to dynamic feedback controller under constant and time-varying perturbations w.

7 Conclusions

In this paper, we have studied for the first time the stability properties of LTI plants controlled by dynamic gradient-based controllers subject to persistent attacks. For static output-feedback controllers, we showed that exponential stability of the system can be guaranteed by constraining sufficiently often the total activation time of the attacks via defense mechanisms with a persistent rejection property. For the dynamic gradient-flow controllers, we showed that system stability can be further guaranteed by properly tuning the controller gain. To the best knowledge of the authors, our results are the first in the literature of feedback-based automatic optimization that study the convergence and stability properties of the algorithms using the framework of hybrid dynamical systems, and Lyapunov-based tools for switched systems with unstable modes. Overall, our results demonstrate for the first time that input-to-state stability can be guaranteed for the closed-loop system provided the total activation time of the unstable modes satisfies a particular upper bound, and enough time scale separation is induced between the dynamics of the plant and the controller. We also presented explicit characterizations of the ISS gains, the theoretical upper bounds on the controller gains, and the time ratio constraints needed to preserve stability. Relevant research directions that require further investigation include the analysis of systems with non-common equilibrium points and extensions to plants with nonlinear dynamics.

References

[1] A. Jokic, M. Lazar, P. P. V. D. Bosch, On constrained steady-state regulation: Dynamic KKT controllers, IEEE Transactions on Automatic Control 54 (2009) 2250–2254.

[2] F. D. Brunner, H.-B. Dürr, C. Ebenbauer, Feedback design for multi-agent systems: A saddle point approach, in: IEEE Conf. on Decision and Control, 2012, pp. 3783–3789.

[3] A. Hauswirth, S. Bolognani, G. Hug, F. Dörfler, Timescale separation in autonomous optimization, IEEE Transactions on Automatic Control 66 (2021) 611–624.

[4] L. S. P. Lawrence, Z. E. Nelson, E. Mallada, J. W. Simpson-Porco, Optimal steady-state control for linear time-invariant systems, in: IEEE Conf. on Decision and Control, Miami Beach, FL, USA, 2018, pp. 3251–3257.

[5] M. Colombino, E. Dall'Anese, A. Bernstein, Online optimization as a feedback controller: Stability and tracking, IEEE Transactions on Control of Network Systems 7 (2019) 422–432.

[6] S. Menta, A. Hauswirth, S. Bolognani, G. Hug, F. Dörfler, Stability of dynamic feedback optimization with applications to power systems, in: Annual Conf. on Communication, Control, and Computing, 2018, pp. 136–143.

[7] G. Bianchin, J. I. Poveda, E. Dall'Anese, Online optimization of switched lti systems using continuous-time and hybrid accelerated gradient flows, arXiv (2020). ArXiv:2008.03903.

[8] S. H. Low, D. E. Lapsley, Optimization flow control. i. basic algorithm and convergence, IEEE/ACM Transactions on networking 7 (1999) 861–874.

[9] T. Zheng, J. W. Simpson-Porco, E. Mallada, Implicit trajectory planning for feedback linearizable systems: A time-varying optimization approach, arXiv preprint (2019). ArXiv:1910.00678.

[10] H. Karimi, J. Nutini, M. Schmidt, Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition, in: Machine Learning and Knowledge Discovery in Databases, Riva del Garda, Italy, 2016, pp. 795–811.

[11] F. Galarza-Jimenez, J. I. Poveda, G. Bianchin, E. Dall'Anese, Extremum seeking under persistent gradient deception: A switching systems approach, IEEE Control Systems Letters (2021). To appear.

[12] F. Pasqualetti, F. Dörfler, F. Bullo, Attack detection and identification in cyber-physical systems, IEEE Transactions on Automatic Control 58 (2013) 2715–2729.

[13] D. Senejohnny, P. Tesi, C. De Persis, A jamming-resilient algorithm for self-triggered network coordination, IEEE Transactions on Control of Network Systems 5 (2018) 981–990.

[14] A. Cetinkaya, H. Ishii, T. Hayakawa, Networked control under random and malicious packet losses, IEEE Transactions on Automatic Control 62 (2016) 2434–2449.

[15] H. S. Foroush, S. Martinez, On event-triggered control of linear systems under periodic denial-of-service jamming attacks, in: IEEE Conf. on Decision and Control, IEEE, 2012, pp. 2551–2556.

[16] A. Kundu, D. E. Quevedo, Stabilizing scheduling policies for networked control systems, IEEE Transactions on Control of Network Systems 7 (2020) 163–175.

[17] N. Forti, G. Battistelli, L. Chisci, B. Sinopoli, Secure state estimation of cyber-physical systems under switching attacks, IFAC-PapersOnLine 50 (2017) 4979–4986.

[18] S. Z. Yong, M. Zhu, E. Frazzoli, Resilient state estimation against switching attacks on stochastic cyber-physical systems, in: 2015 54th IEEE Conference on Decision and Control (CDC), 2015, pp. 5162–5169. doi:10.1109/CDC.2015.7403027.

[19] Y. Wang, J. Lu, J. Liang, Security control of multiagent systems under denial-of-service attacks, IEEE Transactions on Cybernetics (2020) 1–11.

[20] X.-F. Wang, A. R. Teel, K.-Z. Liu, X.-M. Sun, Stability analysis of distributed convex optimization under persistent attacks: A hybrid systems approach, Automatica 111 (2020) 108607.

[21] Y. Qi, Y. Tang, Z. Ke, Y. Liu, X. Xu, S. Yuan, Dual-terminal decentralized event-triggered control for switched systems with cyber attacks and quantization, ISA Transactions (2020).

[22] S. Sundaram, B. Gharesifard, Distributed optimization under adversarial nodes, IEEE Transactions on Automatic Control 64 (2018) 1063–1076.

[23] L. Su, N. H. Vaidya, Byzantine-resilient multi-agent optimization, IEEE Transactions on Automatic Control (2020). In press.

[24] Y. Chen, L. Su, J. Xu, Distributed statistical machine learning in adversarial settings: Byzantine gradient descent, ACM on Measur. and Analy. of Computing Syst. 1 (2017) 1–25.

[25] B. Turan, C. A. Uribe, H.-T. Wai, M. Alizadeh, On robustness of the normalized subgradient method with randomly corrupted subgradients, arXiv preprint (2020). ArXiv:2009.13725.

[26] L. Zhai, K. G. Vamvoudakis, Data-based and secure switched cyber–physical systems, Systems and Control Letters 148 (2021) 104826.

[27] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid Dynamical Systems, Princeton University Press, Princeton, NJ, USA, 2012.

[28] A. R. Teel, F. Forni, L. Zaccarian, Lyapunov-based sufficient conditions for exponential stability in hybrid systems, IEEE Transactions on Automatic Control 58 (2013) 1591–1596.

[29] C. Cai, A. R. Teel, Characterizations of input-to-state stability for hybrid systems, Systems and Control Letters 58 (2009) 47 – 53.

[30] G. Bianchin, J. Cortés, J. I. Poveda, E. Dall'Anese, Time-varying optimization of LTI systems via projected primal-dual gradient flows, arXiv (2021). ArXiv:2101.01799.

[31] P. Kokotović, H. K. Khalil, J. O'Reilly, Singular Perturbation Methods in Control: Analysis and Design, Society for Industrial and Applied Mathematics, 1999.

[32] H. K. Khalil, Nonlinear Systems, Prentice Hall, Upper Saddle River, NJ, 2002.

[33] J. W. Simpson-Porco, On area control errors, area injection errors, and textbook automatic generation control, IEEE Transactions on Power Systems (2020).

[34] P. Chakraborty, S. Dhople, Y. C. Chen, M. Parvania, Dynamics-aware continuous-time economic dispatch and optimal automatic generation control, in: 2020 American Control Conference (ACC), IEEE, 2020, pp. 1292–1298.

[35] M. Papageorgiou, A. Kotsialos, Freeway ramp metering: An overview, IEEE Transactions on Intelligent Transportation Systems 3 (2002) 271–281.

[36] P. Grandinetti, C. Canudas-de Wit, F. Garin, Distributed optimal traffic lights design for large-scale urban networks, IEEE Transactions on Control Systems Technology 27 (2018) 950–963.

[37] S. Paternain, D. Koditschek, A. Ribeiro, Navigation functions for convex potentials in a space with convex obstacles, IEEE Transactions on Automatic Control 63 (2018) 2944–2959.

[38] C. Zhang, D. Arnold, N. Ghods, A. Siranosian, M. Krstić, Source seeking with non-holonomic unicycle without position measurement and with tuning of forward velocity, Systems and Control Letters 56 (2007) 245–252.

[39] J. I. Poveda, M. Benosman, A. R. Teel, R. G. Sanfelice, Coordinated hybrid source seeking with robust obstacle avoidance in multi-vehicle autonomous systems, IEEE Transactions on Automatic Control, 10.1109/TAC.2021.3056365 (2021) 1–16.

[40] R. M. Murray, Z. Li, S. S. Sastry, A Mathematical Introduction to Robotic Manipulation, CRC Press, 1994.

[41] J. I. Poveda, A. R. Teel, A framework for a class of hybrid extremum seeking controllers with dynamic inclusions, Automatica 76 (2017) 113–126.

[42] G. Yang, D. Liberzon, Input-to-state stability for switched systems with unstable subsystems: A hybrid Lyapunov construction, 53rd IEEE Conf. Decision Control (2014) 6240–6245.

Appendix

Consider a set-valued HDS with state $\vartheta := (\tau_1, \tau_2, \phi) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \Phi$, with $\Phi := \Phi_u \cup \Phi_s$, and dynamics:

$$\vartheta \in C_M := [0, N_0] \times [0, T_0] \times \Phi, \tag{41a}$$

$$\begin{pmatrix} \dot{\tau}_1 \\ \dot{\tau}_2 \\ \dot{\phi} \end{pmatrix} \in F_M(\vartheta) := \begin{pmatrix} [0, \kappa_1] \\ [0, \kappa_2] - I_{\Phi_u}(\phi) \\ 0 \end{pmatrix}, \tag{41b}$$

$$\vartheta \in D_M := [1, N_0] \times [0, T_0] \times \Phi, \tag{41c}$$

$$\begin{pmatrix} \tau_1^+ \\ \tau_2^+ \\ \phi^+ \end{pmatrix} \in G_M(\vartheta) := \begin{pmatrix} \tau_1 - 1 \\ \tau_2 \\ \Phi \backslash \{\phi\} \end{pmatrix}, \tag{41d}$$

where $T_0 \geq 0$, $N_0 \in \mathbb{Z}_{\geq 1}$, $\kappa_1 > 0$, and $\kappa_2 \in (0, 1)$. Denote $\mathcal{T} := [0, N_0] \times [0, T_0] \times \Phi$, and consider a HDS with state $\xi = (\vartheta, (\zeta, s))$, where $\vartheta \in \mathbb{R}^3$, $\zeta \in \mathbb{R}^p$, $\upsilon \in \mathbb{R}^m$ and $s \in \mathbb{R}$; having continuous-time dynamics given by

$$\xi \in C_M \times C, \quad \dot{\vartheta} \in F_M(\vartheta), \;\; \dot{\zeta} = F_\phi(\zeta, s, \upsilon), \;\; \dot{s} = \varsigma, \tag{42}$$

where $\varsigma > 0$, $F_M : \mathbb{R}^3 \rightrightarrows \mathbb{R}^3$, $C := \mathbb{R}^p \times [\underline{s}, \bar{s}]$ with $\bar{s} > \underline{s} > 0$, and $t \mapsto \upsilon(t)$ is a continuously differentiable external input. The discrete-time dynamics are given by

$$\xi \in D_1 \cup D_2, \qquad \xi^+ \in G_{1,2}(\xi), \tag{43}$$

where $D_1 := D_M \times C$, $D_2 := C_M \times D$, $D := \mathbb{R}^p \times \{\bar{s}\}$, D_M and C_M are defined in (41), and

$$G_{1,2}(\xi) := \begin{cases} G_1(\xi), & \text{if } \xi \in D_1 \\ G_2(\xi), & \text{if } \xi \in D_2 \\ G_1(\xi) \cup G_2(\xi), & \text{if } \xi \in D_1 \cap D_2, \end{cases} \tag{44}$$

with set-valued maps $G_1, G_2 : \mathbb{R}^{4+p} \rightrightarrows \mathbb{R}^{4+p}$ defined as

$$G_1(\xi) = G_M(\vartheta) \times \{\zeta\} \times \{s\}, \;\; G_2(\xi) = \{\vartheta\} \times \{G(\zeta)\} \times \{\underline{s}\},$$

where $G : \mathbb{R}^p \to \mathbb{R}^p$, and G_M is defined in (41d). The following lemma provides a minor extension of [11, Lemma 6] for systems with inputs in the flow map via the mapping F_ϕ. It is similar to [42, Prop 3.], with the difference that the states (τ_1, τ_2, ϕ) are generated by the hybrid automaton (41), and it allows for extra periodic jumps triggered by the timer s.

Lemma 4 *Suppose that G and F_ϕ are continuous functions for each $\phi \in \Phi := \Phi_s \cup \Phi_u \subset \mathbb{Z}_{\geq 1}$, where (Φ_s, Φ_u) satisfy $\Phi_s \cap \Phi_u = \emptyset$. Let $\psi := (\zeta, s)$ and $\mathcal{A} \subset C \cup D$ be compact. Suppose there exist continuously differentiable functions $V_\phi : (C \cup D) \to \mathbb{R}_{\geq 0}$ such that:*

1. *There exists $\alpha_1(r) := c_1 r^2$, and $\alpha_2(r) := c_2 r^2$ such that:*
$$\alpha_1(|\psi|_{\mathcal{A}}) \leq V_\phi(\psi) \leq \alpha_2(|\psi|_{\mathcal{A}}), \;\; \forall \; (\psi, \phi) \in (C \cup D) \times \Phi.$$

2. *There exists $\varphi \in \mathcal{K}_\infty$, and $\rho_s, \rho_u > 0$ such that if $|\psi|_{\mathcal{A}} \geq \varphi(\|\upsilon\|)$, then*
$$\langle \nabla V_{\phi_s}(\psi), F_{\phi_s}(\psi, \upsilon) \rangle \leq -\rho_s V_{\phi_s}(\psi), \;\; \forall \; (\psi, \phi_s) \in C \times \Phi_s.$$
$$\langle \nabla V_{\phi_u}(\psi), F_{\phi_u}(\psi, \upsilon) \rangle \leq \rho_u V_{\phi_u}(\psi), \;\; \forall \; (\psi, \phi_u) \in C \times \Phi_u.$$

3. *There exists $\omega \geq 1$ such that*
$$V_{\phi'}(\psi) \leq \omega V_\phi(\psi), \;\; \forall \; (\psi, \phi', \phi) \in (C \cup D) \times \Phi \times \Phi.$$

4. *There exists $\varrho \in (0, 1) > 0$ such that*
$$V_\phi(\psi^+) - V_\phi(\psi) \leq -\varrho V_\phi(\psi), \;\; \forall \; (\psi, \phi) \in D \times \Phi.$$

Then, if $\rho_s > \kappa_1 \ln(\omega) + \kappa_2(\rho_s + \rho_u)$, there is $\alpha \in \mathcal{K}_\infty$, and $\beta \in \mathcal{KL}$ such that for any complete solution $\psi(t, k)$ to the HDS (42)-(43) *with initial conditions $\psi(0, 0)$ and input υ then $|\psi(t, k)|_{\mathcal{A} \times \mathcal{T}} \leq \beta(|\psi(t, k)|_{\mathcal{A} \times \mathcal{T}}, t - 0) + \alpha(\|u\|_t)$, for all $t \geq 0$.* □

Proof: Define $\tau := \ln(\omega)\tau_1 + (\rho_s+\rho_u)\tau_2$, and $V(\vartheta) = V_\phi(\psi)e^\tau$. Using (41b), it follows that during flows we have $\dot\tau \in \ln(\omega)[0,\kappa_1] + (\rho_s+\rho_u)([0,\kappa_2] - I_{\Phi_u}(\phi)) = [0,\gamma] - (\rho_s+\rho_u)I_{\Phi_u}(\phi)$, where $\gamma := \kappa_2(\rho_s+\rho_u) + \kappa_1\ln(\omega)$. It follows that if $\phi \in \Phi_s$ and $\psi \in C$, then

$$\begin{aligned}\dot V(\vartheta) &\le V_\phi(\psi)e^\tau\dot\tau - \rho_s V_\phi(\psi)e^\tau \\ &= -(\rho_s-\gamma)V_\phi(\psi)e^\tau = -\rho V(\vartheta),\end{aligned} \tag{45}$$

where $\rho := \rho_s - \gamma > 0$ whenever $\rho_s > \kappa_2(\rho_s+\rho_u) + \kappa_1\ln(\omega)$. Similarly, if $\phi \in \Phi_u$ and $\psi \in C$, then

$$\begin{aligned}\dot V(\vartheta) &\le V_\phi(\psi)e^\tau\dot\tau + \rho_u V_\phi(\psi)e^\tau \\ &\le V_\phi(\psi)e^\tau(\gamma - (\rho_s+\rho_u)) + \rho_u V_\phi(\psi)e^\tau \le -\rho V(\vartheta).\end{aligned}$$

During jumps of the form $\vartheta^+ \in G_2(\vartheta)$, we have that

$$V(\vartheta^+) = V_\phi(\psi^+)e^\tau \le (1-\varrho)V_\phi(\psi)e^\tau = (1-\varrho)V(\vartheta).$$

for all $\xi \in D_2$. Similarly, since $\tau^+ = \tau - \ln(\omega)$, during jumps of the form $\vartheta^+ \in G_1(\vartheta)$, we have

$$\begin{aligned}V(\vartheta^+) = V_{\phi^+}(\psi)e^{\tau^+} &\le \max_{\phi^+\in\Phi} V_{\phi^+}(\psi)e^\tau e^{-\ln(\omega)} \\ &\le \omega V_\phi(\psi)e^\tau e^{-\ln(\omega)} = V(\vartheta).\end{aligned} \tag{46}$$

for all $\xi \in D_1$. Combining inequalities (45)-(46), the result follows by [42, Proposition 3] with $\alpha(r) := \alpha_1^{-1}(\alpha_2(\varphi(r)))$, and $\beta(r,t) := \alpha_1^{-1}(\alpha_2(r)\exp(-\rho t))$. ■

Lemma 5 *Let $\alpha, \beta, \delta, \chi, \gamma$ be positive scalars, let $\theta \in (0,1)$ be tunable parameters, and let*

$$\Xi = \begin{bmatrix} \theta\left(\frac{\alpha}{\varepsilon} - \beta\right) & -\frac{1}{2}\left((1-\theta)\delta + \theta\chi\right) \\ -\frac{1}{2}\left((1-\theta)\delta + \theta\chi\right) & (1-\theta)\gamma \end{bmatrix}.$$

If $0 < \varepsilon < \alpha\gamma/(\beta\gamma + \delta\chi)$, then there exists $\theta < \delta/(\delta+\chi)$, such that Ξ is positive definite.

Proof: The matrix Ξ is positive definite if and only if the leading principal minors are positive. In this case $(1-\theta)\gamma > 0$ and $\theta(1-\theta)\left(\frac{\alpha}{\varepsilon} - \beta\right)\gamma > \frac{1}{4}((1-\theta)\delta + \theta\chi)^2$. The first inequality is guarantee by the definition of θ and γ. The second inequality can be rewritten as:

$$\varepsilon < \frac{\alpha\gamma}{\beta\gamma + \frac{((1-\theta)\delta + \theta\chi)^2}{4\theta(1-\theta)}} = \hat\varepsilon(\theta).$$

The function $\hat\varepsilon$ attains its maximum at $\theta = \theta^* := \frac{\delta}{\delta+\chi}$, with $\varepsilon^* := \hat\varepsilon(\theta^*) = \frac{\alpha\gamma}{\beta\gamma + \delta\chi}$. ■