Online Optimization of Switched LTI Systems Using Continuous-Time and Hybrid Accelerated Gradient Flows

Gianluca Bianchin, Jorge I. Poveda, Emiliano Dall'Anese

Department of Electrical, Computer, and Energy Engineering, University of Colorado, Boulder, CO, 80309 USA.

Abstract

This paper studies the design of feedback controllers to steer the output of a switched linear time-invariant system towards the solution of a time-varying optimization problem. We propose two classes of control policies: an online gradient-descent controller, and an online hybrid controller that can be interpreted as a regularized version of Nesterov's accelerated gradient method. By design, the controllers continuously steer the system towards the time-varying minimizer of the underlying optimization problem without requiring knowledge of exogenous inputs or disturbances. For cost functions that are smooth and satisfy the Polyak-Łojasiewicz inequality, we demonstrate that the online gradient descent controller ensures uniform global exponential stability when the time-scales of the system and controller are sufficiently separated and the switching signal of the system is slow on the average. Under a strong convexity assumption, we show that the online hybrid accelerated controller guarantees tracking of optimal trajectories, and outperforms controllers based on gradient descent. Interestingly, the proposed hybrid accelerated controller resolves the lack of robustness suffered by standard accelerated gradient methods when coupled with a dynamical system. When the cost function is not strongly convex, we establish global practical asymptotic stability for the hybrid accelerated method. Our results suggest the existence of a trade-off between acceleration rate and convergence accuracy for online controllers using dynamic momentum. Our findings are illustrated through several numerical simulations.

Key words: Optimization, Online Optimization, Switched Systems, Hybrid Dynamical Systems.

1 Introduction

This paper focuses on the design of feedback controllers that steer the output of a dynamical system towards the solution of an optimization problem that embeds performance metrics associated with the steady-state inputs and outputs of the system. The setting considered here – as well as in the prior works [1–7] – advocates a departure from conventional architectures that rely on a distinct separation between steady-state optimization and control, where: an optimization problem is solved offline based on a steady-state input-output map of the system to compute reference points and, subsequently, proportional and/or integral controllers are used to steer the system to the setpoint while coping with disturbances. The premise for this class of architectures is that disturbances and uncontrollable exogenous inputs are predictable and slowly varying, so that offline optimization can be used to produce effective reference points. In contrast, we consider the design of control laws that rely on principled first-order optimization methods, thus incorporating the optimization and control layers in a joint scheme that can quickly react to unmodeled and dynamically-varying disturbances. By design, this class of online optimization controllers steers the system to a set-point that is implicitly defined as a solution of an optimization problem, without requiring knowledge of exogenous disturbances. This design strategy is pertinent for a number of domains where physical systems operate in dynamic settings, and where metering of exogenous disturbances is impractical. Prominent examples include power grids [2, 4] and transportation systems [8, 9].

Related Work. The considered setting is aligned with works on feedback-based optimization [10–12] and autonomous optimization [2,3,6]. The works [10–12] model the underlying (physical) system as a steady-state input-output map, thus implicitly assuming that the dynamics of the system are significantly faster than the algorithmic updates. The authors in [4] considered stable linear time-invariant (LTI) dynamical systems, and designed a gradient descent controller to drive the system output to the critical points of an unconstrained problem with constant exogenous disturbances. The approach was extended to stable plants with nonlinear dynamics and to more general controllers in [3]. Joint stabilization of an LTI dynamical system and regulation of its output to the solution of a convex optimization problem was considered in [5, 6]. LTI dynamical systems with time-varying exogenous inputs were considered in [2], along with the problem of tracking an optimal solution trajectory of a *time-varying* problem with a strongly convex cost. The authors leveraged integral quadratic constraints to provide linear matrix inequality conditions (to be checked numerically) that guarantee exponential stability and a bounded tracking error. Time-varying optimization problems and feedback-linearizable systems were considered in [7]. We also mention the works [13, 14] on regret minimization for discrete-time time LTI systems.

Contributions. In this paper, we address the steady state optimization problem in a class of systems that have not been studied before, namely, *switched LTI systems*. This type of systems emerges naturally in many engineering applications with multiple operation modes, as well as in feedback control systems that are stabilized by supervisory controllers. For a comprehensive introduction to LTI switched systems and its applications, we refer the reader to [15] and [16]. When the target optimization problem is time-varying – due to time-varying cost functions or to the presence of time-varying disturbances – the problem at hand becomes that of tracking an optimal trajectory for the switched system, preserving suitable stability properties in the closed-loop system. We propose two design strategies: *(i)* an online gradient method (or gradient flow), similar to those considered in [3, 4]; and, *(ii)* a novel hybrid feedback controller based on Nesterov's accelerated gradient flows [17–19], which incorporates dynamic momentum in order to induce acceleration in the closed-loop system, without sacrificing stability and robustness properties that are fundamental in feedback control. The two alternatives offer a number of trade-offs between implementation complexity, achievable convergence rates, and conditions for exponential stability. Based on this, the main contributions of this work are the following:

- First, for the gradient flow controller, we leverage arguments from singular perturbation analysis [20] and input-to-state stability (ISS) for hybrid systems [21] in order to provide sufficient conditions for uniform global exponential stability for problems where the steady state cost function has a Lipschitz-continuous gradient and satisfies the Polyak-Łojasiewicz inequality [22]. Besides considering *switching* plants and providing sufficient stability conditions in terms of average dwell-times, the contribution relative to [3–6] consists in providing conditions for exponential stability and ISS, with a explicit quantification of rates of convergence and linear gains.
- Second, when the optimization problem is time-varying, we show that the gradient flow controller achieves tracking of the optimal trajectories with tracking error that is bounded by the time-variability of the unknown exogenous disturbances. Relative to [2], our conditions for stability are easier to check as they do not require to numerically solve a linear matrix inequality.
- Third, we propose an online hybrid controller that is based on Nesterov's accelerated gradient method [18, 23]. When the cost function is strongly convex, we show that this controller ensures uniform global exponential stability of the optimal points and tracking of time-varying trajectories. Our results provide a possible solution to the instability and lack of robustness suffered by standard online accelerated gradient flows interconnected with dynamic plants, discussed in [3, Sec. IV.B]. When the function is only convex we show that the online accelerated gradient method achieves asymptotic convergence to a small neighborhood of the optimizers.

The rest of this paper is organized as follows: Section 2 describes the statement of the problem. Sections 3 and 4 contain the main stability results, while Section 5 contains the technical proofs. Section 6 presents some numerical examples, and Section 7 concludes the paper.

Notation. Given a compact set $\mathcal{A} \subset \mathbb{R}^n$ and a vector $x \in \mathbb{R}^n$, we define $\|x\|_{\mathcal{A}} := \min_{y\in\mathcal{A}} \|y - x\|_2$. Given $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$, we let $\text{vec}\,(v, w) := [v^\top, w^\top]^\top$. For a matrix $M \in \mathbb{R}^{n\times n}$, we let $\bar{\lambda}(M)$ and $\underline{\lambda}(M)$ denote its largest and smallest eigenvalues, respectively. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{KLL}$ if it is nondecreasing in its first argument, nonincreasing in its second and third argument, $\lim_{r\to 0^+} \beta(r, s_1, s_2) = 0$ for each pair $(s_1, s_2) \in \mathbb{R}_{\geq 0}$, and $\lim_{s_i\to\infty} \beta(r, s_1, s_2) = 0$ for each $r \in \mathbb{R}_{\geq 0}$, and $i \in \{1, 2\}$. A continuous function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is of class $\mathcal{K}_\infty$ if $\gamma(0) = 0$, it is nondecreasing, and $\lim_{r\to\infty} \gamma(r) = \infty$.

2 Problem Statement and Preliminaries

Consider a switched linear dynamical system of the form:

$$\begin{aligned} \dot{x} &= A_\sigma x + B_\sigma u + E_\sigma w, \\ y &= Cx + Dw, \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the input, $w \in \mathbb{R}^q$ is an unknown (and possibly time-varying) exogenous input or disturbance, $y \in \mathbb{R}^p$ is the measurable output, and $\sigma : \mathbb{R}_+ \to \mathcal{S} := \{1, 2, ..., S\} \subset \mathbb{Z}_+$ is a piecewise-constant function that specifies the index of the active subsystem. Each value of $\sigma \in \mathcal{S}$ specifies a LTI system in (1), which is also referred to as a *mode* of the plant, with matrices $(A_\sigma, B_\sigma, E_\sigma, C, D)$ of appropriate dimension. In order to have well-defined feedback optimization problems, we make the following assumptions.

Assumption 1 *For each $\sigma \in \mathcal{S}$, there exist positive definite matrices $P_\sigma \in \mathbb{R}^{n\times n}$ and $Q_\sigma \in \mathbb{R}^{n\times n}$, such that $A_\sigma^\top P_\sigma + P_\sigma A_\sigma \preceq -Q_\sigma$.*

Assumption 2 *All modes of* (1) *have common equilibrium points; that is, for any $w \in \mathbb{R}^q$ there exists $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ such that $0 = A_p x + B_p u + E_p w$, $\forall\, \sigma \in \mathcal{S}$.*

Assumption 1 implies that each mode is exponentially stable [20]. If a mode is not stable *a priori*, an inner stabilizing controller can be used as in [5, 6]. Assumption 2 is standard in the literature of switched systems (see, e.g., [16, 21]), and it is further discussed in Remark 1.

In this paper we will restrict our attention to switching signals for which the number of discontinuities in every open interval $(s, t) \subset \mathbb{R}_+$, denoted by $N(t, s)$, satisfies $N(t, s) \leq N_0 + \frac{t-s}{\tau_d}$, where $\tau_\mathrm{d} > 0$ is the so-called "average dwell-time" and $N_0 \in \mathbb{Z}_{>0}$ is the "chatter bound" [16]. As discussed in [15], this is a fairly general class of switching signals that do not preclude a finite number of arbitrarily fast switches in the plant (1).

Such behavior is often desirable in many practical applications where the switching signal is generated by a supervisory control system.

We focus on the problem of regulating the state of system (1) to the solutions of the steady-state optimization problem:

$$\min_{u,x} \quad h(u) + g(y) \tag{2a}$$
$$\text{s.t.} \quad 0 = A_\sigma x + B_p u + E_\sigma w, \quad \forall \sigma \in \mathcal{S}, \tag{2b}$$
$$y = Cx + Dw, \tag{2c}$$

where $h : \mathbb{R}^m \to \mathbb{R}$ is a cost associated with the steady-state control input, and $g : \mathbb{R}^p \to \mathbb{R}$ is a cost associated with the steady-state output. The solutions of (2) are the set of equilibrium points of (1) that minimize the cost specified by $h(\cdot)$ and $g(\cdot)$. We make the following assumption on the cost functions.

Assumption 3 *The functions $u \mapsto h(u)$ and $y \mapsto g(y)$ are continuously differentiable and there exist $\ell_u, \ell_y > 0$ such that for every $u, u' \in \mathbb{R}^m$ and $y, y' \in \mathbb{R}^p$, we have $\|\nabla h(u) - \nabla h(u')\| \leq \ell_u \|u - u'\|$, and $\|\nabla g(y) - \nabla g(y')\| \leq \ell_y \|y - y'\|$, respectively.*

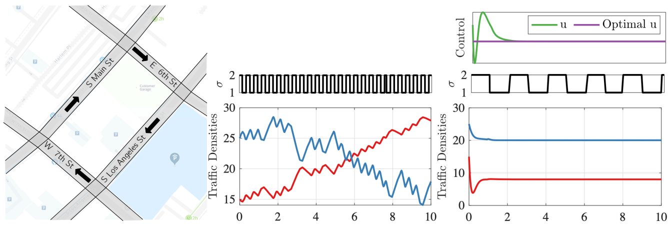
In order to solve problem (2), we next restate it as an optimization problem on the steady-state outputs. By combining (2b)-(2c) we can rewrite the system output as $y = G_\sigma u + H_\sigma w$, where $G_\sigma := -CA_\sigma^{-1}B_\sigma$ and $H_\sigma := D - CA_\sigma^{-1}E_\sigma$ denote the steady-state input-output maps, and where the invertibility of A_σ follows from Assumption 1. Moreover, Assumption 2 implies that the input-output maps are common across the modes and, by letting $G := G_\sigma$ and $H := H_\sigma$, problem (2) can be reformulated as the following unconstrained problem:

$$\min_{u} \quad f(u) := h(u) + g(Gu + Hw), \tag{3}$$

where again we see w as an exogenous signal that parametrizes f. Therefore, when w is time-varying, the point-wise optimization problem (2) becomes time-varying (similarly to, e.g., [2]), and the optimal points become optimal *trajectories*. Finally, two technical observations are in order. First, Assumption 3 implies that $u \mapsto f(u)$ has a Lipschitz-continuous gradient with Lipschitz constant $\ell = \ell_u + \ell_y \|G\|^2$. Second, every solution of (2) is also a solution of (3); however, the inverse implication holds only under the assumption that the pairs (A_σ, C_σ) are observable for all $\sigma \in \mathcal{S}$. Nonetheless, this assumption is not necessary for the subsequent analysis.

Remark 1 Since problem (3) is independent of σ, for continuous functions $t \mapsto w(t)$ the optimal trajectories $t \mapsto u^*(t)$ generated by solving (3), point-wise in time, are well-defined continuous functions. □

Remark 2 When the matrices G_σ and H_σ are not common across modes, and system (1) is at equilibrium for constant values of u and w, one can define the average map $y = G_{\text{av}} u + H_{\text{av}} w$, where $G_{\text{av}} := \sum_{p \in \mathcal{S}} \alpha_p G_p$ and $H_{\text{av}} := \sum_{p \in \mathcal{S}} \alpha_p H_p$, with $0 \leq \alpha_p \leq 1$, $\sum_p \alpha_p = 1$. In this case, G_{av} and H_{av} represent the average transfer functions; one may then consider the problem $\min_u h(u) + g(G_{\text{av}} u + H_{\text{av}} w)$. This case is relevant in the context of transportation systems (see, e.g., [8]). □



(a) (b) (c)

Fig. 1. (a) Block with two signalized intersections in Downtown Los Angeles. (b) Suboptimal switching and control result in unstable trajectories. (c) Stable trajectories.

Remark 3 When the cost function in problem (2) is parametrized by a set of time-varying parameters, the optimization problem becomes time-varying. Two simple examples are: $h(u,t) = \frac{1}{2} u^\top Q(t) u$ where $Q(t) \in \mathbb{R}^{m \times m}$ models time-varying weights, and $g(y,t) = \frac{1}{2}(y - \bar{y}(t))^\top R(t)(y - \bar{y}(t))$. Although in this paper we focus only on time-varying vectors w, the results can be naturally extend to this class of time-varying costs. □

2.1 Problem Statement

We define the problem addressed in this work.

Problem 1 *Design a dynamical feedback controller for the input u that uses only real-time measurements of the system output y, such that for any bounded and continuously differentiable function w the following holds for all trajectories $z := \text{vec}(x, u)$ of the closed-loop system:*

$$\limsup_{t \to \infty} \|z(t) - z^*(t)\| \leq \gamma(\sup_{\tau \geq 0} \|\dot{w}(\tau)\|),$$

where $\gamma \in \mathcal{K}_\infty$ and $z^* := \text{vec}(x^*, u^*)$ denotes an optimal trajectory of (2), i.e., a trajectory that satisfies $0 = A_\sigma x^*(t) + B_\sigma u^*(t) + E_\sigma w(t)$, and $0 = \nabla h(u^*(t)) + \nabla g(Cx^*(t) + Dw(t))$, at all times. □

In words, Problem 1 specifies that, for any well-behaved disturbance w, the error between the trajectories of (1) and an optimal trajectory of (2) must be bounded at all times, and if $\dot{w}(t) \to 0$ as $t \to \infty$ (i.e., the disturbance w becomes a constant), then $\lim_{t \to \infty} \|z(t) - z^*\| = 0$. Notice that when $\dot{w} = 0$, Problem 1 simplifies to the standard optimal *output regulation* problem.

Problem 1 is motivated, in part, by the following example, which shows that plants where all modes are stable may result in unstable behaviors when the switching signal and/or the controller are not properly designed.

Example 1 *Consider the traffic network in Fig. 1, which models a urban block in Downtown Los Angeles.*

We model with a variable x_1 the combined traffic density in W 7th St and S Main St, and with a variable x_2 the combined traffic density in E 6th St and S Los Angeles St, where densities of shorter roads have been combined for simplicity of illustration. We model the dynamics of the states by adopting the Cell Transmission Model (CTM):

$$\dot{x}_1 = -\min\{d_1x_1, s_2x_2\} + r_{21}\min\{d_2x_2, s_1x_1\} + w_1,$$
$$\dot{x}_2 = -\min\{d_2x_2, s_1x_1\} + r_{12}\min\{d_1x_1, s_2x_2\} + w_2,$$

where $d_1, d_2 > 0$ model linear demand functions, $s_1, s_2 > 0$ model linear supply functions, $r_{12}, r_{21} \in [0, 1]$ describe the routing preferences (routing ratios), and $w_1, w_2 > 0$ model exogenous inflows and outflows to the network. The minimization functions in the CTM describe regimes of free-flow (in this case, $\min\{d_ix_i, s_jx_j\} = d_ix_i$), or regimes of congestion (in this case, $\min\{d_ix_i, s_jx_j\} = s_jx_j$). We assume that the system alternates between regimes of free-flow and regimes of congestion, so that the system is described by two modes:

$$\begin{aligned}
\textit{Mode 1:}\quad & \dot{x}_1 = -d_1x_1 + r_{21}s_1x_1 + w_1,\\
& \dot{x}_2 = -s_1x_1 + r_{12}d_1x_1 - (1 - r_{21})d_2x_2 + w_2,\\
\textit{Mode 2:}\quad & \dot{x}_1 = -s_2x_2 + r_{21}d_2x_2 - (1 - r_{12})d_1x_1 + w_1,\\
& \dot{x}_2 = -d_2x_2 + r_{12}s_2x_2 + w_2,
\end{aligned}$$

where the terms $(1 - r_{21})d_2x_2$ and $(1 - r_{12})d_1x_1$ model vehicles that leave the network when the downstream road is congested (by using, e.g., an uncongested lane). Let $d_1 = 0.79$, $d_2 = 0.67$, $s_1 = 1.33$, $s_2 = 0.71$, $r_{12} = 0.79$, and $r_{21} = 0.47$. Then, the system can be modeled as s linear switched system of the form (1), where each mode is stable (Mode 1 has eigenvalues $\{-0.46, -0.17\}$, Mode 2 has eigenvalues $\{-0.17, -0.10\}$). Although all modes are stable, Fig. 1(b) illustrates a case where the system trajectories engage in oscillations that are amplified over time. This behavior suggests that when the CTM switches between regimes of free-flow and congestion in an uncontrolled fashion, the network may have unstable behaviors even when all modes are stable. On the other hand, assume that the exogenous inflows can be partially controlled so that $w_i = u + \tilde{w}_i$, $i = \{1, 2\}$, where u_i denotes the portion of controllable flow and $\tilde{w}_i$ denotes the portion of uncontrollable flow. Fig. 1(c) demonstrates that the system converges to a stable equilibrium when the control and switching signals are properly designed. □

2.2 Modeling Framework

In order to address Problem 1, we will use the framework of set-valued hybrid dynamical systems (HDS), introduced in [21]. In general, a set-valued hybrid dynamical system with exogenous inputs w is modeled by a state $\varphi \in \mathbb{R}^n$, and the equations [21]:

$$\varphi \in C,\ \dot{\varphi} \in F(\varphi, \dot{w}), \qquad \varphi \in D,\ \varphi^+ \in G(\varphi), \qquad (4)$$

where $F : \mathbb{R}^n \times \mathbb{R}^q \rightrightarrows \mathbb{R}^n$, and $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are the flow and jump map, respectively, and the sets $C \subset \mathbb{R}^n$ and $D \subset \mathbb{R}^n$ are the flow set and jump set, respectively, and the signal $\dot{w}$ represents an input that in our case will correspond to the time-derivative of w in (1). Systems of the form (4) generalize purely continuous time systems ($D = \emptyset$) and purely discrete-time systems ($C = \emptyset$), and their solutions z are parametrized by two time indexes: a continuous-time index $t \in \mathbb{R}_{\geq 0}$ that increases continuously whenever $\varphi(t, j) \in C$ and the system can evolve as $\dot{\varphi} := \frac{d\varphi(t,j)}{dt} \in F(\varphi(t, j), w(t, j))$; and a discrete-time index $j \in \mathbb{Z}_{\geq 0}$ that increases by one whenever $\varphi(t, j) \in D$ and the system evolves as $\varphi^+ := \varphi(t, j + 1) \in G(\varphi(t, j), w(t, j))$. Therefore, solutions to (4) are defined on hybrid time-domains [21, Ch. 2], which are subsets of $\mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ defined as a union of intervals $[t_j, t_{j+1}] \times \{j\}$, where $0 = t_0 \leq t_1 \leq \ldots$, where the last interval can be closed or open to the right. For a precise definition of solutions to HDS, we refer the reader to [21, Ch. 2]. Applications of HDS in standard static optimization can be found in [19] and [24].

There are four main motivations behind the use of the modeling framework (4) for the solution of Problem 1. (a) First, the framework of HDS allows us to model switching LTI systems under average dwell-time constraints as time-invariant dynamical systems, thus facilitating our stability analysis. Indeed, as shown in [25, Prop. 1.1], every switching signal satisfying an average dwell-time constraint with (τ_d, N_0) can be generated by the following HDS:

$$(\tau, \sigma) \in \mathcal{T}_C := [0, N_0] \times \mathcal{S},\ \dot{\sigma} = 0,\ \dot{\tau} \in [0, 1/\tau_d] \qquad (5a)$$
$$(\tau, \sigma) \in \mathcal{T}_D := [1, N_0] \times \mathcal{S},\ \sigma^+ \in \mathcal{S},\ \tau^+ = \tau - 1. \qquad (5b)$$

Moreover, every solution of system (5) satisfies an average dwell-time constraint with (τ_d, N_0). (b) Second, the framework of HDS allows us to use Lyapunov tools for the analysis of *nonlinear* hybrid controllers with set-valued dynamics. (c) Third, by working with *well-posed* HDS we will be able to establish suitable robustness properties for the closed-loop system. (d) Fourth, the powerful notion of input-to-state stability (ISS) for HDS allows us to establish a relationship between Problem 1 and the notion of ISS, which is summarized next.

Definition 2.1 *System (4) is said to render a compact set $\mathcal{A} \subseteq \mathbb{R}^n$ ISS if there exists a function $\gamma \in \mathcal{K}_\infty$, and a function $\beta \in \mathcal{KLL}$ such that all solutions satisfy:*

$$\|\varphi(t, j)\|_{\mathcal{A}} \leq \beta(\|\varphi(0, 0)\|_{\mathcal{A}}, t, j) + \gamma(\sup_{\tau\geq 0} \|\dot{w}(\tau)\|). \qquad (6)$$

for all $(t, j) \in dom(\varphi)$. Moreover, system (4) is said to render the set $\mathcal{A}$ exponentially ISS (E-ISS) [26, Def. 5] if there exists positive constants a_0, b_0, c_0, d_0, such that

$$\|\varphi(t, j)\|_{\mathcal{A}} \leq a_0(e^{-\frac{1}{2}(b_0t+c_0j)}\|\varphi(0, 0)\|_{\mathcal{A}} + d_0 \sup_{\tau\geq 0} \|\dot{w}(\tau)\|)$$

for all $(t, j) \in dom(\varphi)$.

Since the function β in (6) is a class-$\mathcal{KLL}$ function, it follows that $\limsup_{t+j\to\infty} \|\varphi(t, j)\|_{\mathcal{A}} \leq \gamma(\sup_{\tau\geq 0} \|\dot{w}(\tau)\|)$.

Moreover, when $\dot{w} = 0$, the bound (6) reduces to the standard characterization of uniform global asymptotic stability of $\mathcal{A}$, [21, Ch.3]. It then follows that ISS is a suitable property to certify the solution of Problem 1. Note that when $d_0 = 0$, E-ISS reduces to the standard notion of exponential stability for HDS [27]. Since in this paper the exogenous input $t \mapsto w(t)$ is assumed to be continuous and only affects the flow map of the HDS, we omit the notation of j in the argument of w to simplify our presentation.

We focus our attention on two main feedback control architectures, shown in Fig. 2. First, in Section 3, we will study the solution of Problem 1 in switched LTI systems by using a gradient descent-based controller (Fig.2-(a)), similar to those studied in [3–6]. For this controller, we will establish stronger and novel results in terms of exponential stability and E-ISS. After this, in Section 4, we introduce a novel hybrid feedback-controller (Fig.2-(b)) inspired by the accelerated Nesterov's gradient flows of [18, 23], which will induce better transient performance in terms of rate of convergence.

3 Feedback Control Using Gradient Flows

When w and H are *known*, a feedback-based gradient flow for the solution of Problem 1 is a dynamical system on $\mathbb{R}^m$ of the form:

$$\dot{u} = -\nabla h(u) - G^\top \nabla g(Gu + Hw). \tag{7}$$

Under Assumption 3, the above dynamical equation admits a unique solution that converges to the critical points of (2). To solve Problem 1, we interconnect the plant dynamics (1) with a gradient-flow algorithm of the form (7) where the input-output map $Gu + Hw$ is replaced by instantaneous measurements of the system output [2,4,6].The resulting interconnection is given by:

$$\varepsilon_\sigma \dot{x} = A_\sigma x + B_\sigma u + E_\sigma w, \qquad y = Cx + Dw, \tag{8a}$$

$$\dot{u} = -\nabla h(u) - G^\top \nabla g(y), \quad (x, u) \in \mathbb{R}^n \times \mathbb{R}^m, \tag{8b}$$

where $\varepsilon_\sigma \in \mathbb{R}_{>0}$ is a (small) parameter that characterizes the difference in timescale between the (faster) plant and the (slower) controller; see also Figure 2(a).

The following assumption will be used to strengthen our convergence results.

Assumption 4 *The function f, defined in* (3)*, satisfies the Polyak-Łojasiewicz (PL) inequality, i.e., $\exists\ \mu > 0$ such that $\frac{1}{2}\|\nabla f(u)\|^2 \geq \mu(f(u) - f(u^*))$, $\forall\ u \in \mathbb{R}^m$.*

The PL inequality implies that every stationary point is a global minimum of $f(u)$. Furthermore, it is possible to show that it implies the following quadratic growth (QG) condition [22]: $f(u) - f(u^*) \geq \frac{\mu}{2}\|u - u^*\|^2$, $\forall\ u \in \mathbb{R}^m$, where u^* is a global minimizer of f. For simplicity of exposition, in this section we assume that u^* is *unique*.

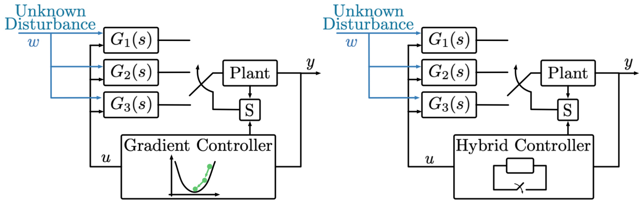


Fig. 2. (a) A switched system in feedback with a gradient flow controller. (b) A switched system in feedback with a hybrid controller. "S" denotes a supervisory controller that actuates the switching between the modes of the plant.

Remark 4 The PL inequality is a weak condition used in the analysis of linear convergence in gradient methods [22]. If a function is μ-strongly convex, then it also satisfies the PL inequality. However, the inverse implication does not hold: functions that satisfy the PL inequality are not necessarily convex, instead, they satisfy a notion of invexity [22]. □

3.1 Main Results

We begin by characterizing the performance of system (8) when the disturbance w is constant. To this end, for each $\sigma \in \mathcal{S}$, we define the following quantities:

$$\bar{a}_\sigma := \max\left\{(1 - \theta_\sigma)\frac{\ell}{2}, \theta_\sigma \bar{\lambda}(P_\sigma)\right\}, \tag{9a}$$

$$\underline{a}_\sigma := \min\left\{(1 - \theta_\sigma)\frac{\mu}{2}, \theta_\sigma \underline{\lambda}(P_\sigma)\right\}, \tag{9b}$$

where we recall that $\ell = \ell_u + \ell_y\|G\|^2$, and θ_σ is given by

$$\theta_\sigma := \frac{\ell_y\|C\|\|G\|}{\ell_y\|C\|\|G\| + 2\|P_\sigma A_\sigma^{-1} B_\sigma\|}.$$

For convenience we also define the quantities $\bar{a} := \max_{\sigma\in\mathcal{S}} \bar{a}_\sigma$, and $\underline{a} := \min_{\sigma\in\mathcal{S}} \underline{a}_\sigma$.

The following results provides a sufficient condition to ensure output regulation under the assumption that the switching signal σ is constant. All the proofs are based on Lyapunov tools for hybrid dynamical systems, and are postponed to Section 5.

Proposition 3.1 *Suppose that Assumptions 1-4 hold. Consider the closed-loop system given by* (8) *with $\sigma(t) = p \in \mathcal{S}$, $\forall\ t \geq 0$, and ε_σ satisfying the inequality*

$$\varepsilon_p < \bar{\varepsilon}_p := \frac{\underline{\lambda}(Q_p)}{4\ell_y\|C\|\|G\|\|P_p A_p^{-1} B_p\|}. \tag{10}$$

Then, for any constant w, the closed-loop system is exponentially stable with respect to the set $\{z^*\}$, with parameters: $a_0 = (\bar{a}_p/\underline{a}_p)^{0.5}$, $b_0 = 2\mu^2/\ell$, $c_0 = d_0 = 0$, where $\bar{a}_p$ and $\underline{a}_p$ are defined in (9)*.*

Proposition 3.1 gives an explicit characterization, for each mode of the plant, on the necessary time scale separation between the plant dynamics and the control dynamics in order to have E-ISS. We point out that the conditions of Proposition 3.1 are computationally-lighter to check relative to [2], where a linear matrix inequality must be solved and no explicit bound for ε_p is available. As shown in (10), and not surprisingly, the time scale separation is proportional to the smallest eigenvalue of the matrix Q_σ of Assumption 1. Note that in this case the rate of convergence is of order $\mathcal{O}(e^{-\mu^2/\ell})$. Differently, when $f(u)$ is strongly convex, it can be shown that the rate of convergence is $\mathcal{O}(e^{-\kappa})$, where κ is the condition number of $f(u)$. Next, we provide a solution to the optimal regulation problem for the switched LTI plant.

Proposition 3.2 *Let Assumptions 1-4 be satisfied, and let each mode implement $\varepsilon_\sigma \in (0, \bar{\varepsilon}_\sigma)$ where $\bar{\varepsilon}_\sigma$ is as in* (10) *for all $\sigma \in \mathcal{S}$. If τ_d in* (5) *satisfies*

$$\tau_d > \frac{\ell}{2\mu^2} \ln\left(\frac{\bar{a}}{\underline{a}}\right), \quad N_0 \in \mathbb{Z}_{>0}, \tag{11}$$

then for any constant w and any positive ϱ satisfying $\ln(\bar{a}/\underline{a}) < \varrho < 2\mu^2\tau_d/\ell$, the closed-loop system given by (5) *and* (8) *with regulation error $\tilde{z} := z - z^*$, and states $(\tilde{z}, \sigma, \tau)$ renders globally exponentially stable the set $\mathcal{A} = \{0\} \times \mathcal{S} \times \mathcal{T}_c$, with parameters:*

$$a_0 = \sqrt{\frac{\bar{a}e^{N_0 k}}{\underline{a}}}, \quad b_0 = \frac{2\mu^2}{\ell} - \frac{\varrho}{\tau_d}, \quad c_0 = \varrho - \ln\left(\frac{\bar{a}}{\underline{a}}\right),$$

and $d_0 = 0$.

Theorem 3.2 asserts that when the individual modes are exponentially stable and the dwell time of the switching signal of the LTI plants satisfies (11), then the gradient flow controller guarantees exponential regulation to the optimal solution of (2). The coefficients (b_0, c_0) characterize the effect of the dwell-time condition on the rate of convergence of the closed-loop system. As $\tau_d \to \infty$, we recover the transient performance of Proposition 3.1.

Remark 5 When Assumption 4 does not hold, [3, 4] showed that under conditions similar to (10) the interconnection of the gradient-flow controller with a linear system (not switching) converges to critical points of (3) and strict local minimizers are asymptotically stable. Suppose that these conditions hold for each of the modes in (1); then, one can use [16, Theorem 4] to establish convergence to critical points and asymptotic stability of strict local minimizers also for switched systems. However, in general in this case it is not possible to obtain exponential stability and the explicit characterizations of the coefficients (a_0, b_0, c_0, d_0). We do not include this result for space constraints. □

Having established a sufficient condition to guarantees E-ISS for the switched closed-loop system (8) with constant disturbances, we are now ready to present the first main result of this paper, which provides a solution to the general tracking Problem 1.

Theorem 3.3 *Let Assumptions 1-4 be satisfied, and consider the closed-loop system* (5) *and* (8)*, with state $(\tilde{z}, \tau, \sigma)$, where $\tilde{z} = z - z^*$ is the tracking error. Let each mode implement $\varepsilon_\sigma \in (0, \bar{\varepsilon}_\sigma)$ where $\bar{\varepsilon}_p$ is as in* (10) *for all $\sigma \in \mathcal{S}$.*

(i) If $\sigma(t) = p$, $\forall\, t \geq 0$, then the set $\mathcal{A} = \{0\} \times \mathcal{T}_C$ is E-ISS with parameters:

$$a_0 = \sqrt{\frac{\bar{a}_p}{\underline{a}_p}}, \quad b_0 = \frac{2\mu^2}{\ell}, \quad d_0 = \frac{\sqrt{2}\|r_p\|}{\tilde{d}\min\{1, \mu^2\}},$$

and $c_0 = 0$, where $r_p^\top := [2\theta_p\|P_pA_p^{-1}E_p\|, (1 - \theta_p)\ell_y\|H\|\|G\|]$, and $\tilde{d} > 0$.

(ii) If $t \mapsto \sigma(t)$ satisfies (11)*, then for each positive ϱ satisfying $\ln(\bar{a}/\underline{a}) < \varrho < 2\mu^2\tau_d/\ell$, the set $\mathcal{A} = \{0\} \times \mathcal{S} \times \mathcal{T}_C$ is E-ISS with parameters:*

$$a_0 = \sqrt{\frac{\bar{a}e^{N_0\varrho}}{\underline{a}}}, \qquad b_0 = \frac{2\mu^2}{\ell} - \frac{\varrho}{\tau_d},$$

$$c_0 = \varrho - \ln\left(\frac{\bar{a}}{\underline{a}}\right), \qquad d_0 = \frac{\sqrt{2}\max_p\|r_p\|}{\tilde{d}\min\{1, \mu^2\}}$$

where $\tilde{d} > 0$.

The characterization of the coefficients (a_0, b_0, c_0, d_0) in items *(i)-(ii)* of Theorem 3.3 can be expressed in terms of three main types of parameters: (a) the parameters (μ, ℓ), which characterize the condition number of the steady state cost f, and which are common in standard optimization problems; (b) the parameters (τ_d, N_0) that characterize the switching signal of the LTI system; and (c) the parameters $(\theta_\sigma, \bar{a}_\sigma, \underline{a}_\sigma, r_\sigma)$, which are intrinsically related to the internal dynamics of each mode $\sigma \in \mathcal{S}$ of the LTI system (1).

The proof of Theorem 3.3 exploits singular perturbation tools for hybrid dynamical systems. However, the result does not follow directly by applying existing singular perturbation results in the literature. This is the case because unlike [28] and [29, Sec. 4.4], the closed-loop system (8) is not linear. Moreover, the general results of [29] and [30] assume *a priori* that the states of the fast dynamics (in this case, the plant) evolve in a predefined compact set, and do not establish convergence results for the fast states. Therefore, in order to prove Theorem 3.3, we have extended to the hybrid setting the singular perturbation approach presented in [20, Ch.11] by using a suitable Lyapunov construction that exploits some of the ideas of [31] in order to directly establish the desired ISS results under average dwell-time constraints and strong decrease conditions during flows and jumps for the Lyapunov function.

4 Feedback Control Using Hybrid Accelerated Gradient Flows

Whereas gradient descent methods have a long tradition in the design of feedback controllers, recent developments in the areas of optimization and machine learning have triggered a new interest in a different class of algorithms with acceleration properties induced by the incorporation of dynamic momentum, see for instance [18,32, 33]. However, even though the convergence properties of these algorithms in standard optimization problems are now well-understood, their application in problems of feedback control remain mostly unexplored. Moreover, as shown by the counter examples presented in [3, Sec. IV.B] and [19, Prop. 1], some of the dynamics with time-varying momentum that emerge from continuous-time approximations of accelerated algorithms can lead to unstable systems once they are slightly perturbed or interconnected with a dynamic plant. Indeed, in many cases, these dynamics are not even suitable for the direct application of singular perturbation theory. Nevertheless, given that accelerated optimization algorithms generate trajectories that minimize the cost at a more efficient rate compared to gradient descent flows, their application in optimization problems is of interest.

In this section, we address this problem by introducing a hybrid feedback controller inspired by Nesterov's accelerated gradient method [18,33]. In particular, we now consider the closed-loop system with continuous-time dynamics given by

$$\varepsilon_\sigma \dot{x} = A_\sigma x + B_\sigma u_1 + E_\sigma w, \tag{12a}$$

$$\begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \dot{u}_3 \end{pmatrix} = \begin{pmatrix} \dfrac{\rho}{u_3}(u_2 - u_1) \\ -\dfrac{\kappa u_3}{\rho}\left(\nabla h(u_1) + G^\top \nabla g(y)\right) \\ 1 \end{pmatrix}, \tag{12b}$$

which evolves whenever the states are in the set

$$(x, u_1, u_2, u_3) \in C := \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times [\delta, \Delta], \tag{13}$$

where $\Delta > \delta > 0$ and $\kappa, \rho > 0$ are tunable parameters. The closed-loop system also implements discrete-time dynamics, given by

$$x^+ = x, \qquad u^+ = R_0 u, \tag{14}$$

with $u = \text{vec}(u_1, u_2, u_3)$ and matrix R_0 given by

$$R_0 = \begin{bmatrix} I & 0 & 0 \\ r_0 I & (1 - r_0)I & 0 \\ 0 & 0 & \frac{\delta}{\Delta} \end{bmatrix}, \quad r_0 \in \{0, 1\}, \tag{15}$$

The updates (14) occur whenever the states satisfy:

$$(x, u_1, u_2, u_3) \in D := \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \{\Delta\}. \tag{16}$$

The hybrid controller is then characterized by equations (12b) and the updates of u in (14), and it is interconnected with the switching LTI system by using $u = u_1$ as output, and $y = Cx + Dw$ as input. In this controller, the state $u_2 \in \mathbb{R}^m$ acts as a momentum state, $u_3 \in \mathbb{R}_+$ models a timer governed by the constants $\Delta, \delta > 0$, and the Boolean variable $r_0 \in \{0, 1\}$ selects a resetting policy for the momentum state u_2. The hybrid controller can be seen as a regularized version of Nesterov's ODE [18], with additional discrete-time dynamics that periodically reset the momentum and the timer. In order to study the closed-loop system (12)-(16), we will work with the following two assumptions.

Assumption 5 *The function f is convex and radially unbounded, and $\exists\ \ell_0 > 0$ and $\nu_0 > 0$ such that $\|u - u^*\| > \nu_0 \implies \|\nabla f(u)\| > \ell_0 \|u - u^*\|$, $\forall\ u \in \mathbb{R}^m$ and $\forall\ u^* \in$ arg min $f(u)$.*

Assumption 6 *$\exists\ \mu > 0$ such that $f(u) \geq f(u') + \nabla f(u')^\top (u - u') + \frac{\mu}{2}\|u - u'\|^2$, for all $u, u' \in \mathbb{R}^m$.*

The last property of Assumption 5 is a type of reverse Lipschitz condition. This property will only be needed to guarantee global convergence in our results. However, it can be omitted if one is interested only on semi-global convergence results. On the other hand, Assumption 6, is in general more restrictive than the PL inequality used to analyze the gradient-flow controller, and it states that the cost is strongly convex. Assumption 6 implies Assumption 5 when f is twice continuously differentiable.

4.1 Main Results

Our first result concerns the stability properties of (12)-(16) under constant values of σ and w. The result does not ask for strong convexity, but rather assumes only convexity and the reverse Lipschitz condition. Since it is well-known that well-posed hybrid systems with *asymptotically* stable individual modes retain the main stability properties under *slow* average switching [21, Cor. 7.28], we present the result only for fixed values of σ. To do this, for each $\sigma \in \mathcal{S}$ and each constant w we define the optimal set $\mathcal{A}_f := \arg\min\ f(u)$, and the "regulation" set $\mathcal{A}_r := \{z^* = (x^*, u_1^*) \in \mathbb{R}^{n+m} : u_1^* \in \mathcal{A}_f, x^* = -A_\sigma^{-1} B_\sigma u_1^*\}$, which is compact due to Assumption 5.

Theorem 4.1 *Let Assumptions 1-3, and 5 hold. Suppose that $\sigma(t) = p$, $\forall\ t \geq 0$, and let the controller parameters satisfy $\rho \leq 4$ and $r_0 = 0$. Then, for every $\nu > 0$ there exits $\varepsilon_p > 0$ satisfying*

$$\varepsilon_p < \frac{\underline{\lambda}(Q_p)\delta}{12\Delta \ell_y \rho \|C_p\| \|G_p\| \|P_p A_p^{-1} B_p\|} \min\left\{\frac{\rho}{\kappa\Delta}, \frac{\delta \ell_0}{2\rho\ell}\right\}, \tag{17}$$

such that, for any constant w, any solution z of (12)-(16), and all $(t, j) \in dom(z)$, the following holds:

- $\limsup_{t+j\to\infty} \|z(t,j)\|_{\mathcal{A}_r} \leq \nu, \quad \forall\ (t,j) \in dom(z),$
- $f(u_1(t,j)) - f(u_1^*) \leq \dfrac{\alpha(s,j)}{u_3^2(t,j)} + \nu, \ \forall\ (s,j) \in dom(z),\ t > s,$

where $\alpha := \theta\alpha_p + (1-\theta)\alpha_c$ with $\alpha_p := \tilde{x}^\top P_\sigma \tilde{x} e^{\varrho u_3}$, $\tilde{x} = x - x^*$, $\varrho > 0$, $\theta \in (0,1)$, and $\alpha_c := \frac{1}{4}|u_1 - u_2|^2 + \frac{1}{4}|u_2|^2_{\mathcal{A}_f} + \kappa\delta^2(f(u_1) - f(u^*))$.

The convergence result of Theorem 4.1 is global, but "practical". Namely, convergence is only achieved to a small ν-neighborhood of the solution z^* of Problem 1, where ν can be made arbitrarily small by the choice of ε_p. To our knowledge, this is the first result in the literature that establishes convergence and stability properties for online feedback controllers based on accelerated gradient flows interconnected with a dynamic plant, let alone in switched systems. Here, two comments comments are in order. First, the bound (17) shows that the feasible time-scale separation (parameterized by ε_p) shrinks to zero when either $\delta \to 0^+$ or $\Delta \to \infty$. Since $\delta = 0$ and $\Delta = \infty$ imply that the controller (12b) simplifies to Nesterov's second order ODE studied in [18, 33], our analysis suggests that in this case there is no $\varepsilon > 0$ that guarantees stability and convergence properties for the closed-loop system. This is consistent with some unstable behaviors observed in numerical experiments, e.g., [3, Sec. IV.B]. Second, the result establishes that the steady-state sub-optimality measure decreases (outside a neighborhood of the optimal point) a rate of order $\mathcal{O}(c_j/u_3^2)$ during the intervals of flow. It follows that the larger the difference $\Delta-\delta$, the larger the size of the intervals where this bound holds. Since in the hybrid controller the state u_3 models a timer, the decrease on the sub-optimality measure can be seen as a (semi) sub-linear acceleration property, similar to those studied in [18, 23, 33] for static optimization problems, which here only holds during each interval of flow. In contrast to existing results, the function α now also contains information related to the LTI dynamics (1), particularly via the matrix P_σ (c.f. Assumption 1). Finally, the bound (17) also reveals the effect of the Lipschitz parameters (ℓ_0, ℓ) on the required time-scale separation, as well as the role of the smallest eigenvalue of the stability matrix Q_σ.

Next, we study the properties of the closed-loop system (12)-(16) under the additional assumption that the steady-state cost (2) is strongly convex. In this case, we use the resetting policy $r_0 = 1$ and, for ease of notation, we first introduce the following constants defined for each $\sigma \in \mathcal{S}$:

$$\begin{aligned}
\bar{a}_\sigma &:= \max\left\{(1-\theta)\kappa\ell\Delta^2/2\rho, \theta\bar{\lambda}(P_\sigma)\exp(\Delta)\right\}, \\
\underline{a}_\sigma &:= \min\left\{(1-\theta_\sigma)/2, (1-\theta_\sigma)\kappa\mu\delta^2/4\rho, \theta_\sigma\underline{\lambda}(P_\sigma)\exp(\delta)\right\}, \\
b &:= \min\left\{\delta\mu/4\ell\Delta^2, \rho^2/2\kappa\ell\Delta\right\}, \\
c &:= \max\{\ln(\Delta^2\kappa\mu^2/(\delta\kappa\mu^2 - 2\rho)), \Delta - \delta\},
\end{aligned} \tag{18}$$

where now $\theta_\sigma := \eta_\sigma/(\eta_\sigma + \delta_\sigma)$, and

$$\eta_\sigma := \frac{2\sqrt{2}\kappa\Delta\ell_y\|C_\sigma\|\|G_\sigma\|}{\rho}, \quad \delta_\sigma := \frac{2\exp(\Delta)\rho\|P_\sigma A_\sigma^{-1}B_\sigma\|}{\delta},$$

As before, we define $\bar{a} := \max_\sigma \bar{a}_\sigma$ $\underline{a} := \min_\sigma \underline{a}_\sigma$, and:

$$\underline{\tau} := \frac{\ln\bar{a} - \ln\underline{a}}{\underline{b}}, \tag{19a}$$

$$\bar{\varepsilon}_\sigma := \frac{\exp(\delta-\Delta)\gamma\underline{\lambda}(Q_\sigma)\delta}{\gamma\delta\bar{\lambda}(P_\sigma) + 2\sqrt{2}\kappa\Delta\ell_y\|C_\sigma\|\|G_\sigma\|\|P_\sigma A_\sigma^{-1}B_\sigma\|}. \tag{19b}$$

where $\gamma = \min\{\rho/4\Delta, \kappa\delta\mu/8\rho\}$. Again, note that $\bar{\varepsilon}_\sigma \to 0^+$ as either $\delta \to 0^+$ or $\Delta \to \infty$.

The following theorem, corresponding to the second main result of this paper, establishes sufficient conditions to solve Problem 1, in terms of the required time-scale separation in the closed-loop system, the dwell-time conditions of the LTI plants, and the restarting frequency of the hybrid controller. We state the result with respect to the set $\mathcal{A} = \{0\} \times \{0\} \times [\delta, \Delta] \times \mathcal{T}_C$.

Theorem 4.2 *Suppose that Assumptions 1-3 and 6 hold, and consider the closed-loop system* (5) *and* (12)-(16)*, with state* $(\tilde{z}, u_2, u_3, \sigma, \tau)$*, where* $\tilde{z} = (x - x^*, u_1 - u_1^*)$ *is the tracking error. Let the parameters of the hybrid controller satisfy* $\Delta^2 - \delta^2 > \frac{2\rho}{\kappa\mu}$, $\rho \le 4$ *and* $r_0 = 1$*. If* $\varepsilon_\sigma \in (0, \bar{\varepsilon}_\sigma)$*, then the following holds:*

(i) If $\sigma(t) = p \in \mathcal{S}$, $\forall\, t \ge 0$*, then the set* $\mathcal{A}$ *is E-ISS with:* $a_0 = (\bar{a}_p/\underline{a}_p)^{0.5}$, $b_0 = b_p$, $c_0 = c_p$, $d_0 = \|r_p\|/\tilde{d}$*, where* $r_p := [2\theta\|P_p A_p^{-1}E_p\|, (1-\theta_p)\frac{\sqrt{2}\kappa\ell_y\Delta^2}{2\rho}\|H_p\|\|G_p\|]^\top$ *and* $\tilde{d} > 0$.

(ii) If $\tau_d > \underline{\tau}$*, then for any* ϱ *satisfying* $\ln(\bar{c}\bar{a}/\underline{a}) < \varrho < b\tau_d$*, the set* $\mathcal{A}$ *is E-ISS with constants* $d_0 = \sqrt{2}e^{-\varrho+\ln(\bar{a}/\underline{a})}$, $a_0 = (\bar{a}e^{N_0\varrho}/\underline{a})^{0.5}$, $b_0 = b - \varrho/\tau_d$*, and* $c_0 = -\varrho + \ln(c\bar{a}/\underline{a})$.

Theorem 4.2 establishes an exponential ISS result for the HDS system (12)-(16) with switching signal generated by (5). The ISS bound (6) guarantees convergence of z to a neighborhood of the optimal trajectory z^* (i.e., tracking), where the size of the residual set grows linearly with the dimension of the derivative of w. Note that the "quadratic dwell-time" condition $\Delta^2 - \delta^2 > \frac{2\rho}{\kappa\mu}$ imposed on the updates of the controller is similar to the bound used in [23] for the solution of static optimization problems, and it regulates, via the strong convexity parameter μ, how fast the *controller* can reset its internal states. Therefore, the convergence result of Theorem 4.2 holds under two different switching conditions: an average dwell-time condition imposed on the switches of plant, and a dwell-time condition imposed on the resets of the hybrid controllers. The result of Theorem 4.2 also characterizes rates of convergence and ISS-gains via the coefficients (a_0, b_0, c_0, d_0). An important observation is that the hybrid controller further allows to tune the rate of convergence of the system via the parameters (Δ, δ). This mechanism is analogous to "restarting" techniques studied in the static optimization literature [23, 32]. However, to our knowledge, these techniques have not been studied before in the context of online optimization.

Next, we specialize Theorem 4.2 to two particular cases that are of interest on their own due to their prevalence in engineering systems. Namely, steady-state output *regulation* problems in LTI standard and switched systems:

Corollary 4.3 *Let Assumptions 1-3 and 6 hold, and let* $\sigma(t) = p$, $\forall t \ge 0$, $\rho \le 4$, $r_0 = 1$, *and* $\Delta^2 - \delta^2 > \frac{2\rho}{\kappa\mu}$*. If,* $\varepsilon_\sigma < \bar{\varepsilon}_\sigma$*, then for any constant* w *the set* $\mathcal{A}$ *is E-ISS with:* $a_0 = (\bar{a}_p/\underline{a}_p)^{0.5}$, $b_0 = b_p$, $c_0 = c_p$, $d_0 = 0$.

Corollary 4.4 *Let Assumptions 1-3 and 6 hold. Let each mode implement* $\varepsilon_p \in (0, \bar{\varepsilon}_p)$ *where* $\bar{\varepsilon}_p$ *is as in Proposition 4.3, and let the parameters of the controller satisfy* $\Delta^2 - \delta^2 > \frac{2\rho}{\kappa\mu}$*, and* $\rho \le 4$, $r_0 = 1$*. If* $\tau > \underline{\tau}_d$*, then for any constant* w*, and* ϱ *satisfying* $\ln(c\bar{a}/\underline{a}) < \varrho < b\tau_d$*, the set* $\mathcal{A}$ *is exponentially stable with* $a_0 = (\bar{a}e^{N_0\varrho}/\underline{a})^{0.5}$, $b_0 = b - \varrho/\tau_d$, $c_0 = -\varrho + \ln(c\bar{a}/\underline{a})$.

We emphasize that Theorem 4.2 and Corollaries 4.3-4.4, cannot be trivially obtained when standard continuous-time accelerated methods with vanishing damping are interconnected with a dynamic plant (let alone a switched system), even if the steady state cost function (4.4) is strongly convex, e.g., [3, Sec. IV.B] and [19, Prop. 1]. Indeed, as shown in [19] and [34], without further damping terms, the optimization dynamics may lack uniformity (with respect to initial time) in the convergence properties of system, leading dynamical systems that can be rendered unstable by using arbitrarily small disturbances. This issue can be solved by inducing a *persistence of excitation* (PE) condition on the damping parameter of the controller, which in our case corresponds to the inverse of the timer u_3. In this way, the hybrid controller with reset policy $r_0 = 0$ in (15) can be seen as an online optimization algorithm with *dynamic* momentum that periodically resets the timer u_3 in order to induce the PE property.

On the other hand, the hybrid controller with reset policy $r_0 = 1$ has similarities to (periodic) restarting heuristics used in the standard optimization literature [32], and hybrid reset controllers that have been studied in the context of standard stabilization and tracking [27], [35], [36]. However, the design and application of these types of controllers in *online* optimization problems where acceleration properties are desirable seem to be mostly unexplored.

5 Proofs of the Results

In this section, we present the proofs of the main results of the paper.

5.1 Proof of Proposition 3.1

Since the system does not switch we omit the dependence on σ. Define the change of variables $\tilde{x} := x - \mathcal{M}(u)$, where $\mathcal{M}(u) := -A^{-1}Bu - A^{-1}Ew$. The interconnected system (8) in the new variables reads as:

$$\begin{aligned} \varepsilon\dot{\tilde{x}} &= \psi_p(\tilde{x},u) := A\tilde{x} + \varepsilon A^{-1}B\psi_c(\tilde{x},u) \\ \dot{u} &= \psi_c(\tilde{x},u) := -\nabla h(u) - G^\top \nabla g(C\tilde{x} + Gu + Hw). \end{aligned} \tag{20}$$

Next, consider the following Lyapunov function:

$$V(z) = (1-\theta)V_1(u) + \theta V_2(\tilde{x}), \tag{21}$$

where $V_1(u) := f(u) - f(u^*)$, $V_2(\tilde{x}) = \tilde{x}^\top P\tilde{x}$, and $\theta \in (0,1)$ is a free parameter. By Assumptions 1 and 4, the uniqueness of u^*, and the QG condition, V is positive definite with respect to $z^* = (0,u^*)$ and radially unbounded. Moreover, by using the inequalities $V(z) \leq (1-\theta)\frac{\ell}{2}\|u-u^*\|^2 + \theta\bar{\lambda}(P)\|\tilde{x}\|^2$ and $V \geq (1-\theta)\frac{\mu}{2}\|u-u^*\|^2 + \theta\underline{\lambda}(P)\|\tilde{x}\|^2$, it follows that $\underline{a}|z|^2 \leq V(z) \leq \bar{a}|z|^2$, with $\bar{a} = \max\left\{(1-\theta)\frac{\ell}{2}, \theta\bar{\lambda}(P)\right\}$, $\underline{a} = \min\left\{(1-\theta)\frac{\mu}{2}, \theta\underline{\lambda}(P)\right\}$.

Next, note that the time-derivative of V_1 satisfies:

$$\begin{aligned} \dot{V}_1(z) &= (\nabla h(u) + G^\top \nabla g(Gu + Hw))^\top \psi_c(\tilde{x},u), \\ &= -\psi_c(0,u)^\top(\psi_c(\tilde{x},u) + \psi_c(0,u) - \psi_c(0,u)) \\ &\leq -\|\psi_c(0,u)\|^2 + \|\psi_c(0,u)\|\|\psi_c(\tilde{x},u) - \psi_c(0,u)\| \\ &\leq -\|\nabla f(u)\|^2 + \ell_y\|C\|\|G\|\|\nabla f(u)\|\|\tilde{x}\|. \end{aligned} \tag{22}$$

where the last inequality follows by $\psi_c(0,u) = \nabla f(u)$ and Assumption 3. On the other hand, the time-derivative of V_2 satisfies:

$$\begin{aligned} \dot{V}_2(z) &= 2\tilde{x}^\top P\psi_p(\tilde{x},u) = -\frac{1}{\varepsilon}\tilde{x}^\top Q\tilde{x} + 2\tilde{x}^\top PA^{-1}B\psi_c(\tilde{x},u) \\ &\leq -\frac{\underline{\lambda}(Q)}{\varepsilon}\|\tilde{x}\|^2 + 2\|PA^{-1}B\|\|\tilde{x}\|\|\psi_c(\tilde{x},u)\|, \end{aligned}$$

where the third identity follows from Assumption 1. The last factor in the second term can be bounded as:

$$\begin{aligned} \|\psi_c(\tilde{x},u)\| &= \|\psi_c(\tilde{x},u) + \psi_c(0,u) - \psi_c(0,u)\| \\ &\leq \|\psi_c(0,u)\| + \|\psi_c(\tilde{x},u) - \psi_c(0,u)\| \\ &\leq \|\psi_c(0,u)\| + \ell_y\|C\|\|G\|\|\tilde{x}\|, \end{aligned}$$

Therefore, the time-derivative of V_2 satisfies:

$$\begin{aligned} \dot{V}_2(z) \leq &-\frac{\underline{\lambda}(Q)}{\varepsilon}\|\tilde{x}\|^2 + 2\|PA^{-1}B\|\|\tilde{x}\|\|\nabla f(u)\| \\ &+ 2\ell_y\|C\|\|G\|\|PA^{-1}B\|\|\tilde{x}\|^2. \end{aligned} \tag{23}$$

Next, note that combining the PL inequality and the QC inequality we have $\|u-u^*\|^2 \leq 1/\mu^2\|\nabla f(u)\|^2$. It follows that

$$-(1-\theta)\frac{\ell}{2\mu^2}\|\nabla f(u)\|^2 - \theta\bar{\lambda}(P)\|\tilde{x}\|^2 \leq -V(z). \tag{24}$$

Let $\xi := [\|\tilde{x}\|, \|\nabla f(u)\|]^\top$. Combining the bounds (22)-(24), we obtain $\dot{V} \leq -bV - \xi^\top M\xi$, where M is a symmetric matrix with the form described in Lemma A.2 in the Appendix with parameters: $\alpha = \underline{\lambda}(Q)$, $\beta = 2\ell_y\|C\|\|G\|\|PA^{-1}B\|$, $\eta = \ell_y\|C\|\|G\|$, $\delta = 2\|PA^{-1}B\|$, $\varphi = \bar{\lambda}(P)$, $\nu = \ell/2\mu^2$, and $\gamma = 1$. By using Lemma A.2, we conclude that when ε satisfies (10), then there exists $b < 2\mu^2/\ell$ such that M is positive definite. Finally, the bound in the statement follows from Lemma A.1 in the Appendix, with $c = 0$ and considering only one mode (see also Lemma 4.6 in [20]). ■

5.2 Proof of Theorem 3.2

The proof leverages arguments from [21, Ex. 3.22] and [31]. More precisely, we prove this claim by showing that there exists a Lyapunov function for the HDS (5)-(8) that satisfies the assumptions of Lemma A.1 stated in the Appendix. We consider the Lyapunov function $W(z) = e^{\varrho\tau}V_\sigma(z)$, where $V_\sigma(z)$ is defined as in (21) for each $\sigma \in \mathcal{S}$, and $\varrho > 0$. First, we show that $W(z)$ satisfies (A.1). To this end, notice that

$$\bar{a}_\sigma e^{\varrho N_0}\|z\|^2 \geq e^{\varrho N_0}V_\sigma(z \geq W(z) \geq e^{\varrho\cdot 0}V_\sigma(z) \geq \underline{a}_\sigma\|z\|^2.$$

Hence, (A.1) holds. Next, we show that $W(z)$ satisfies (A.2). Indeed, the time-derivative of W is

$$\dot{W}(z) = e^{\varrho\tau}(\varrho V_\sigma(z)\dot{\tau} + \dot{V}_\sigma(z)) \leq \left(\frac{\varrho}{\tau_d} - \frac{2\mu^2}{\ell}\right) W(z),$$

where we used $\dot{\tau} \leq 1/\tau_d$, and $\dot{V}_\sigma(z) \leq -(2\mu^2/\ell)V_\sigma(z)$ shown in the proof of Proposition 3.1. Hence, (A.2) is satisfied when $\varrho < 2\mu^2\tau_d/\ell$. Next, we show that $W(z)$ satisfies (A.3). During switches of the plant, we have that $z^+ = z$, $u^+ = u$, $\sigma^+ \in \mathcal{S}$, and $\tau^+ = \tau - 1$. Therefore,

$$\begin{aligned} W(z^+) &\leq e^{\varrho\tau^+} \max_\sigma V_\sigma(z^+) \leq e^{\varrho(\tau-1)+\ln \bar{a}_\sigma}\|z\|^2 \\ &\leq e^{-\varrho+\ln \bar{a}_\sigma - \ln \underline{a}_\sigma} W(z), \end{aligned} \tag{25}$$

Hence, $W(z)$ satisfies (A.3) when $\varrho > \ln \bar{a}_\sigma - \ln \underline{a}_\sigma = \ln(\bar{a}_\sigma/\underline{a}_\sigma)$. By combining the upper and lower inequalities on ρ we conclude that $W(z)$ satisfies the assumptions of Lemma A.1 when $\tau_d > \frac{\ell}{2\mu^2}\ln(\bar{a}_\sigma/\ln \underline{a}_\sigma)$ for all $\sigma \in \mathcal{S}$, which establishes the result. ■

5.3 *Proof of Theorem 3.3*

First, we shift the equilibrium points of (8a) to the origin by using $\tilde{x} := x - \mathcal{M}(u)$, where $\mathcal{M}(u) := -A^{-1}Bu - A^{-1}Ew$. System (8) in the new variables reads as:

$$\varepsilon_\sigma \dot{\tilde{x}} = \psi_\sigma(\tilde{x}, u, w), \quad \dot{u} = \psi_c(\tilde{x}, u, w),$$

with $\psi_p(\tilde{x}, u, w) = A_\sigma\tilde{x} + \varepsilon A_\sigma^{-1}B_\sigma\psi_c(\tilde{x}, u, w) + \varepsilon A^{-1}E\dot{w}$, $\psi_c(\tilde{x}, u, w) := -\nabla f(u) - G^\mathsf{T}\nabla g(C\tilde{x} + Gu + Hw)$.

Case (i). Since we focus on a single mode, in what follows we drop the dependence on the subscript σ. We prove this claim by showing that $V(z)$ defined in (21) satisfies the assumptions of Lemma A.1 in the Appendix when applied to a single mode (and setting $c = 0$); see also [20, Lemma 4.6]. To this end, notice that $V(z)$ satisfies (A.1) as previously shown in the proof of Proposition 3.1. Next, we show that $V(z)$ satisfies (A.2). The time-derivative of V_1 reads as:

$$\begin{aligned} \dot{V}_1(z) = \nabla f(u)^\mathsf{T}\psi_c(\tilde{x}, u, w) - \underbrace{\nabla f(u^*)^\mathsf{T}}_{=0} \frac{\partial u^*}{\partial t} \\ + H^\mathsf{T}\left(\nabla g(Gu + Hw) - \nabla g(Gu^* + Hw)\right)\dot{w}. \end{aligned}$$

By using the bound (22) and by recalling that $\psi_c(0, u, w) = \nabla f(u)$, the first term satisfies

$$\nabla f(u)^\mathsf{T}\psi_c(\tilde{x}, u, w) \leq -\|\nabla f(u)\|^2 + \ell_y\|C\|\|G\|\|\nabla f(u)\|\|\tilde{x}\|.$$

By using Assumption 3, the second term can be bounded by $\ell_y\|H\|\|G\|\|u - u^*\|\|\dot{w}\|$. Hence, the following bound on $\dot{V}_1(z)$ can be derived:

$$\begin{aligned} \dot{V}_1(z) \leq -\|\nabla f(u)\|^2 + \ell_y\|C\|\|G\|\|\nabla f(u)\|\|\tilde{x}\| \\ + \ell_y\|H\|\|G\|\|u - u^*\|\|\dot{w}\|. \end{aligned} \tag{26}$$

On the other hand, the time-derivative of V_2 reads:

$$\begin{aligned} \dot{V}_2(z) &= 2\tilde{x}^\mathsf{T}P\left(\frac{1}{\varepsilon}A\tilde{x} + A^{-1}B\psi_c(\tilde{x}, u, w) + A^{-1}E\dot{w}\right) \\ &\leq -\frac{\underline{\lambda}(Q)}{\varepsilon}\|\tilde{x}\|^2 + 2\|PA^{-1}B\|\|\tilde{x}\|\|\psi_c(\tilde{x}, u, w)\| \\ &\quad + 2\|PA^{-1}E\|\|\tilde{x}\|\|\dot{w}\|, \end{aligned}$$

where the inequality follows from Assumption 1. Using Assumption 3 we have:

$$\begin{aligned} \|\psi_c(\tilde{x}, u, w)\| &= \|\psi_c(\tilde{x}, u, w) + \psi_c(0, u, w) - \psi_c(0, u, w)\| \\ &\leq \|\psi_c(0, u, w)\| + \ell_y\|C\|\|G\|\|\tilde{x}\|. \end{aligned}$$

It follows that the time-derivative of $V_2(\tilde{x})$ satisfies:

$$\begin{aligned} \dot{V}_2(z) \leq -\frac{\underline{\lambda}(Q)}{\varepsilon}\|\tilde{x}\|^2 + 2\ell_y\|C\|\|G\|\|PA^{-1}B\|\|\tilde{x}\|^2 \\ + 2\|PA^{-1}B\|\|\tilde{x}\|\|\nabla f(u)\| + 2\|PA^{-1}E\|\|\tilde{x}\|\|\dot{w}\|. \end{aligned} \tag{27}$$

Let $\xi := [\|\tilde{x}\|, \|\nabla f(u)\|]^\mathsf{T}$, $\zeta := [\|\tilde{x}\|, \|u - u^*\|]^\mathsf{T}$; by combining (26)-(27) with (24), we obtain:

$$\dot{V} \leq -bV - \xi^\mathsf{T}\hat{M}\xi + \|\dot{w}\|r^\mathsf{T}\zeta, \tag{28}$$

where $r := [2\theta\|PA^{-1}E\|, (1-\theta)\ell_y\|H\|\|G\|]^\top$, and $\hat{M}$ is a symmetric matrix as in Lemma A.2 with parameters $\alpha = \underline{\lambda}(Q)$, $\beta = 2\ell_y\|C\|\|G\|\|PA^{-1}B\|$, $\eta = \ell_y\|C\|\|G\|$, $\delta = 2\|PA^{-1}B\|$, $\varphi = \bar{\lambda}(P)$, $\nu = \ell/2\mu^2$, and $\gamma = 1$. By using Lemma A.2 and by substituting, we conclude that when ε satisfies (10) then there exists $0 < b < 2\mu^2/\ell$ such that $\hat{M}$ is positive definite. Next, we show that that the quadratic term in (28) dominates the linear term for large $\|\zeta\|$. To this end, we bound the vector ξ as follows $\|\xi\|^2 = \|\nabla f(u)\|^2 + \|\tilde{x}\|^2 \geq \mu^2\|u - u^*\|^2 + \|\tilde{x}\|^2 \geq \min\{\mu^2, 1\}\|\zeta\|^2$, where we used the PL inequality. By using the above bound we rewrite (28) as follows:

$$\begin{aligned} \dot{V}_p + bV_p &\leq -\xi^\mathsf{T}(\hat{M} - kI)\xi - k\|\xi\|^2 + \|\dot{w}\|r^\mathsf{T}\zeta \\ &\leq -\underline{\lambda}(\hat{M} - kI)\|\xi\|^2 - k\min\{\mu^2, 1\}\|\zeta\|^2 \\ &\quad + \|\dot{w}\|\|r\|\|\zeta\| \\ &\leq -\underline{\lambda}(\hat{M} - kI)\|\xi\|^2, \end{aligned} \tag{29}$$

where $0 < k < \underline{\lambda}(\hat{M})$, and the last inequality holds when $-k\min\{\mu^2, 1\}\|\zeta\|^2 + \|\dot{w}\|\|r\|\|\zeta\| < 0$. By recalling that $\|\zeta\| \geq 1/\sqrt{2}\|z\|$, we conclude that (29) is satisfied for all $\|z\| \geq \frac{\sqrt{2}\|r\|\sup_{t\geq 0}\|\dot{w}\|}{k\min\{\mu^2, 1\}}$, which proves (A.2). Finally, the claim follows by application of Theorem A.1. in the Appendix when applied to a single-mode system.

Case (ii). Consider the HDS system (5)-(8) and the Lyapunov function $W(z) = e^{\varrho\tau}V_\sigma(z)$, where $\varrho > 0$ and $V_\sigma(z)$ is as in (21) with V_2 now depending on the mode $\sigma \in \mathcal{S}$ via the matrix P_σ. First, by using $W(z) \geq \underline{a}_\sigma\|z\|^2$

and $W(z) \leq \bar{a}_\sigma e^{\rho N_0}\|z\|^2$ we conclude that $W(z)$ satisfies (A.1) with $\underline{a}_\sigma\|z\|^2 \leq W(z) \leq \bar{a}_\sigma e^{\varrho N_0}\|z\|^2$. Second, we show that $W(z)$ satisfies (A.2). To this end note that $\dot{W}(z) = e^{\varrho\tau}(\varrho V_\sigma(z)\dot{\tau} + \dot{V}_\sigma(z))$. Therefore, $\dot{W}(z) \leq (\frac{\varrho}{\tau_d} - \frac{2\mu^2}{\ell})W(z)$, where the second inequality follows from (29), and holds when $\|z\| \geq \frac{\sqrt{2}\|r\|\sup_{t\geq 0}\|\dot{w}\|}{k\min\{\mu^2,1\}}$. Therefore, (A.2) is satisfied when $\varrho < \frac{2\mu^2}{\ell}\tau_d$. Third, we show that $W(z)$ satisfies (A.3). To this aim, by iterating the bound (25) we conclude that $W(z^+) \leq e^{-\varrho + \ln\bar{a}_p - \ln\underline{a}_p}W(z)$, which shows that $W(z)$ satisfies (A.3) when $\varrho > \ln\bar{a}_\sigma/\underline{a}_\sigma$, for all $\sigma \in \mathcal{S}$. By combining the upper and lower bounds on ρ we conclude that $W(z)$ satisfies the assumptions of Lemma A.1 when $\tau_d > \frac{\ell}{2\mu^2}\ln(\bar{a}_p/\ln\underline{a}_p)$ for all $p \in \mathcal{S}$. ■

5.4 Proof of Theorem 4.1

Since we focus on a single mode, in we drop the subscript σ. We prove this statement by using a positive definite and radially unbounded Lyapunov function V that decreases during flows, outside a neighborhood of the optimal set, and also does not increase during jumps. We begin by shifting the equilibrium points of the plant via the change of variables $\tilde{x} := x - \mathcal{M}(u_1)$, where $\mathcal{M}(u_1) := -A^{-1}Bu_1 - A^{-1}Ew$. Let $\tilde{z} = \mathrm{vec}(\tilde{x}, u)$. In the new variables, the closed-loop continuous-time dynamics are given by:

$$\varepsilon\dot{\tilde{x}} = A\tilde{x} + \varepsilon\frac{\rho}{u_3}A^{-1}B(u_2 - u_1), \quad \dot{u}_1 = \frac{\rho}{u_3}(u_2 - u_1),$$
$$\dot{u}_2 = -\kappa\frac{u_3}{\rho}\psi_c(\tilde{x}, u_1), \qquad \dot{u}_3 = 1. \tag{30}$$

where $\psi_c(\tilde{x}, u_1) = \nabla h(u) + G^\top\nabla g(C\tilde{x} + Gu_1 + Hw)$. Since $u_1^+ = u_1$ and $x^+ = x$ we have that $\tilde{x}^+ = \tilde{x}$. Define $\mathcal{A} := \{0\} \times \mathcal{A}_c \times [\delta, \Delta]$, where $\mathcal{A}_c = \{(u_1, u_2) : u_1 = u_2,\ u_1 \in \arg\min\ f(u)\}$. Next, we consider the Lyapunov function $V(\tilde{z}) = (1-\theta)V_1(u) + \theta V_2(\tilde{x})$, where $0 < \theta < 1$ is a free parameter, $V_1(u) = \frac{1}{2}\big(\|u_2 - u_1\|^2 + \|u_2\|_{\mathcal{A}}^2 + \frac{\kappa u_3^2}{\rho}(f(u_1) - f(u_1^*))\big)$, and $V_2(\tilde{x}) = \tilde{x}^\top P\tilde{x}$. We have that

$$\frac{\partial V_1(u)}{\partial u_1}\dot{u}_1 = -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{\kappa u_3}{2}(u_2 - u_1)^\top\nabla f(u)$$
$$\leq -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{2\kappa u_3}{\rho}(u_2 - u_1)^\top\nabla f(u),$$

where the inequality follows from $\rho \leq 4$. We also have:

$$\frac{\partial V_1(u)}{\partial u_2}\dot{u}_2 = -\frac{\kappa u_3}{\rho}((u_2 - u_1) + (u_2 - u^*))^\top\psi_c(\tilde{x}, u)$$
$$\leq \frac{\kappa u_3}{\rho}\|2u_2 - u_1 - u^*\|\|\psi_c(\tilde{x}, u) - \psi_c(0, u)\|$$
$$- \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\psi_c(0, u),$$

Since $\|2u_2 - u_1 - u^*\| \leq 3(\|u_2 - u_1\| + \|u_1 - u^*\|)$, by using $\|\psi_c(\tilde{x}, u) - \psi_c(0, u)\| \leq \ell_y\|C\|\|G\|\|\tilde{x}\|$, we obtain:

$$\frac{\partial V_1(u)}{\partial u_2}\dot{u}_2 \leq \underbrace{\frac{3\kappa\Delta}{\rho}\ell_y\|C\|\|G\|}_{:=\eta}(\|u_2 - u_1\| + \|u_1\|_{\mathcal{A}})\|\tilde{x}\|$$
$$- \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\psi_c(0, u).$$

Since $\frac{\partial V_1(u)}{\partial u_3}\dot{u}_3 = \frac{\kappa u_3}{\rho}(f(u) - f(u^*))$, recalling that $\psi_c(0, u) = \nabla f(u)$, the time-derivative of V_1 satisfies:

$$\begin{aligned}\dot{V}_1(\tilde{z}) \leq& -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{2\kappa u_3}{\rho}(u_2 - u_1)^\top\nabla f(u)\\ &+ \eta(\|u_2 - u_1\| + \|u_1\|_{\mathcal{A}})\|\tilde{x}\|\\ &- \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\nabla f(u)\\ &+ \frac{\kappa u_3}{\rho}(f(u) - f(u^*))\\ \leq& -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \eta(\|u_2 - u_1\| + \|u_1\|_{\mathcal{A}})\|\tilde{x}\|\\ &+ \frac{\kappa u_3}{\rho}\big((u^* - u_1)^\top\nabla f(u) + f(u) - f(u^*)\big)\\ \leq& -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \eta(\|u_2 - u_1\| + \|u_1\|_{\mathcal{A}})\|\tilde{x}\|\\ &- \omega\|\nabla f(u)\|^2,\end{aligned} \tag{31}$$

where $\omega = \frac{\kappa\delta}{2\rho\ell}$, and where the last inequality follows by using $(u^* - u_1)^\top\nabla f(u) + f(u) - f(u^*) \leq -\frac{1}{2\ell}\|\nabla f(u)\|^2$, which follows from Lipschitz smoothness of the cost (Assumption 3). The time-derivative of V_2 reads as:

$$\begin{aligned}\dot{V}_2(\tilde{z}) =& 2\tilde{x}^\top P\left(\frac{1}{\varepsilon}A\tilde{x} + \frac{\rho}{u_3}A^{-1}B(u_2 - u_1)\right)\\ \leq& -\frac{1}{\varepsilon}\underline{\lambda}(Q)\|\tilde{x}\|^2 + \frac{2\rho}{u_3}\|PA^{-1}B\|\|u_2 - u_1\|\|\tilde{x}\|\\ \leq& -\frac{1}{2\varepsilon}\underline{\lambda}(Q)\|\tilde{x}\|^2 - \frac{1}{2\varepsilon}\underline{\lambda}(Q)\|\tilde{x}\|^2\\ &+ \frac{2\rho}{u_3}\|PA^{-1}B\|\|u_2 - u_1\|\|\tilde{x}\|.\end{aligned} \tag{32}$$

By letting $\hat{\xi} = \mathrm{vec}(\|\tilde{x}\|, \|u_2 - u_1\|)$, and by combining (31) with (32) we conclude that $\dot{V}(\tilde{z})$ can be bounded by a quadratic function of $\hat{\xi}$ as follows:

$$\begin{aligned}\dot{V}(\tilde{z}) \leq& -\hat{\xi}^\top M\hat{\xi} - \theta\frac{\alpha}{\varepsilon}\|\tilde{x}\|^2\\ &+ (1-\theta)(\eta\|\tilde{x}\|\|u_1\|_{\mathcal{A}} - \omega\|\nabla f(u)\|^2),\end{aligned} \tag{33}$$

where M is a symmetric matrix with the form described in Lemma A.2 with parameters: $\alpha = \underline{\lambda}(Q)/2$, $\beta = 0$, $\gamma = \rho/\Delta$, $\eta = 3\kappa\Delta\ell_y\|C\|\|G\|/\rho$, $\delta = 2\rho\|PA^{-1}B\|/\delta$, $b = 0$. By using Lemma A.2 and by substituting we conclude that when ε satisfies (17) the matrix M is positive definite. Next, we distinguish two cases.

Case 1: Suppose $\theta\frac{\alpha}{\varepsilon}\|\tilde{x}\| > (1-\theta)\eta\|u_1\|_{\mathcal{A}_f}$. In this case, $\dot{V}(z) < 0$ whenever $\|u_1\|_{\mathcal{A}_f} \neq 0$.

Case 2: Suppose $\theta\frac{\alpha}{\varepsilon}\|\tilde{x}\| \leq (1-\theta)\eta\|u_1\|_{\mathcal{A}_f}$. In this case, we have $\|\tilde{x}\| \leq \frac{\varepsilon(1-\theta)\eta}{\theta\alpha}\|u_1\|_{\mathcal{A}_f}$, and the inequality (33) can be rewritten as:

$$\begin{aligned}\dot{V}(\tilde{z}) &\leq -\frac{\theta\alpha}{\varepsilon}\|\tilde{x}\|^2 + (1-\theta)(\eta\|\tilde{x}\|\|u_1\|_{\mathcal{A}_f} - \omega\|\nabla f(u)\|^2)\\ &\leq -\frac{\theta\alpha}{\varepsilon}\|\tilde{x}\|^2 + \frac{(1-\theta)^2\eta^2\varepsilon}{\alpha\theta}\|u_1\|^2_{\mathcal{A}_f}\\ &\quad - (1-\theta)\omega\|\nabla f(u)\|^2\\ &\leq -\frac{\theta\alpha}{\varepsilon}\|\tilde{x}\|^2 + \left(\frac{(1-\theta)^2\eta^2\varepsilon}{\alpha\theta} - (1-\theta)\omega\ell_0\right)\|u_1\|^2_{\mathcal{A}_f},\end{aligned}$$

where the last inequality follows from the reverse Lipschitz condition (Assumption 5). Hence, $\dot{V}(\tilde{z}) < 0$ for all $\|\tilde{x}\| \neq 0$ and $\|\tilde{z}\|_{\mathcal{A}} > \delta_0 := \nu_0$ whenever $\varepsilon < \theta\ell_0\alpha\omega/(1-\theta)\eta^2$, which holds when (17) is satisfied. Finally, note that during jumps we have:

$$V(\tilde{z}^+) - V(\tilde{z}) = \frac{\kappa}{2\rho}(\delta^2 - \Delta^2)\left(f(u_1) - f(u^*)\right), \quad (34)$$

which is non-positive since $\delta < \Delta$. The strong decrease of V during flows (outside a δ_0 neighborhood of $\mathcal{A}$), the non-increasing condition (34), and the fact that the jumps are periodic separated by constant intervals of flow, guarantee uniform convergence from compact sets to a δ_0-neighborhood of $\mathcal{A}$ [21, Ch.8]. The convergence bound for $f - f^*$ follows directly by the structure of V and the fact that $V(t,j) \leq V(s,j)$ during flows provided $|\tilde{z}|_{\mathcal{A}} > \nu_0$. ■

5.5 *Proof of Corollary 4.3*

In order to simplify the proof of Theorem 4.2, we first present the proof of Corollaries 4.3 and 4.4. Since we focus on a single mode, we drop again the subscript σ. We establish the result by using a Lyapunov function V that satisfies the assumptions of Lemma A.1 stated in the Appendix. We begin by shifting again to the origin the equilibrium points of the system, using the change of variables (30). Next, we consider the Lyapunov function $V(\tilde{z}) = (1-\theta)V_1(u) + \theta V_2(\tilde{x})$, where $0 < \theta < 1$ is a free parameter, $u = \text{vec}(u_1, u_2, u_3)$, $V_2(\tilde{x}) = \exp(u_3)\ \tilde{x}^\top P\tilde{x}$, and

$$V_1(u) = \frac{1}{2}\big(\|u_2 - u_1\|^2 + \|u_2 - u^*\|^2 + \frac{\kappa u_3^2}{\rho}(f(u_1) - f(u_1^*))\big).$$

By strong convexity of f, it is easy to see that V satisfies (A.1) with $(\bar{a}, \underline{a})$ given by (18). Next, note that

$$\begin{aligned}\frac{\partial V_1(u)}{\partial u_1}\dot{u}_1 &= -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{\kappa u_3}{2}(u_2 - u_1)^\top\nabla f(u)\\ &\leq -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{2\kappa u_3}{\rho}(u_2 - u_1)^\top\nabla f(u),\end{aligned}$$

where the inequality follows from $\rho \leq 4$. Also, note that:

$$\begin{aligned}\frac{\partial V_1(u)}{\partial u_2}\dot{u}_2 &= -\frac{\kappa u_3}{\rho}((u_2 - u_1) + (u_2 - u^*))^\top\psi_c(\tilde{x}, u)\\ &\leq \frac{\kappa u_3}{\rho}\|2u_2 - u_1 - u^*\|\|\psi_c(\tilde{x}, u) - \psi_c(0, u)\|\\ &\quad - \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\psi_c(0, u),\end{aligned}$$

where we decomposed $\psi_c(\tilde{x}, u) = \psi_c(\tilde{x}, u) - \psi_c(0, u) + \psi_c(0, u)$. Let $\tilde{\mathcal{A}}_c = \{u^*\} \times u^* \times [\delta, \Delta]$. During flows the factor in the first term can be bounded as $\|2u_2 - u_1 - u^*\| \leq 2\sqrt{2}\|u\|_{\tilde{\mathcal{A}}_c}$. By substituting the above bound and by using $\|\psi_c(\tilde{x}, u) - \psi_c(0, u)\| \leq \ell_y\|C\|\|G\|\|\tilde{x}\|$ we obtain:

$$\begin{aligned}\frac{\partial V_1(u)}{\partial u_2}\dot{u}_2 &\leq \underbrace{\frac{2\sqrt{2}\kappa\Delta}{\rho}\ell_y\|C\|\|G\|}_{:=\eta}\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c}\\ &\quad - \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\psi_c(0, u). \quad (35)\end{aligned}$$

Since $\frac{\partial V_1(u)}{\partial u_3}\dot{u}_3 = \frac{\kappa u_3}{\rho}(f(u) - f(u^*))$, and $\psi_c(0, u) = \nabla f(u)$, it follows that

$$\begin{aligned}\dot{V}_1(u) &\leq -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \frac{2\kappa u_3}{\rho}(u_2 - u_1)^\top\nabla f(u)\\ &\quad + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c} - \frac{\kappa u_3}{\rho}(2u_2 - u_1 - u^*)^\top\nabla f(u)\\ &\quad + \frac{\kappa u_3}{\rho}(f(u) - f(u^*))\\ &\leq -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c}\\ &\quad + \frac{\kappa u_3}{\rho}\Big((u^* - u_1)^\top\nabla f(u) + f(u) - f(u^*)\Big)\\ &\leq -\frac{\rho}{u_3}\|u_2 - u_1\|^2 + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c} - \frac{\kappa u_3\mu}{2\rho}\|u_1 - u^*\|^2\\ &\leq -\min\left\{\frac{\rho}{u_3}, \frac{\kappa u_3\mu}{2\rho}\right\}(\|u_2 - u_1\|^2 + \|u_1 - u^*\|^2)\\ &\quad + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c}.\end{aligned}$$

Since $\|u_2 - u_1\|^2 + \|u_1 - u^*\|^2 \geq \frac{1}{4}\|u\|^2_{\tilde{\mathcal{A}}_c}$, we obtain:

$$\dot{V}_1(u) \leq -\frac{1}{4}\min\left\{\frac{\rho}{\Delta}, \frac{\kappa\delta\mu}{2\rho}\right\}\|u\|^2_{\tilde{\mathcal{A}}_c} + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c}. \quad (36)$$

The time-derivative of $V_2(z)$ reads as:

$$\begin{aligned}\dot{V}_2(z) &= e^{u_3}\left(2\tilde{x}^\top P\dot{\tilde{x}} + \tilde{x}^\top P\tilde{x}\dot{u}_3\right)\\ &\leq e^{u_3}\bigg(-\frac{1}{\varepsilon}\underline{\lambda}(Q)\|\tilde{x}\|^2\\ &\quad + 2\frac{\rho}{u_3}\|PA^{-1}B\|\|\tilde{x}\|\|u_2 - u_1\| + \bar{\lambda}(P)\|\tilde{x}\|^2\bigg)\\ &\leq -\frac{1}{\varepsilon}\underline{\lambda}(Q)\exp(\delta)\|\tilde{x}\|^2\\ &\quad + e^{\Delta}\left(2\frac{\rho}{\delta}\|PA^{-1}B\|\|\tilde{x}\|\|u\|_{\mathcal{A}} + \bar{\lambda}(P)\|\tilde{x}\|^2\right) \quad (37)\end{aligned}$$

where the last inequality follows from $\|u_2 - u_1\| = \|u_2 - u^* + u^* - u_1\| \leq \|u_2 - u^*\| + \|u_1 - u^*\|$. By letting $\xi := \text{vec}\left(\|\tilde{x}\|, \|u\|_{\tilde{\mathcal{A}}_c}\right)$, and by combining (36), (37), and the quadratic bound on V we obtain $\dot{V}(z) - bV(z) \leq -\xi^\top M\xi$, where M is a symmetric matrix as in Lemma A.2 with parameters: $\alpha = e^{\delta}\underline{\lambda}(Q)$, $\beta = e^{\Delta}\bar{\lambda}(P)$, $\eta = 2\sqrt{2}\kappa\Delta\ell_y\|C\|\|G\|/\rho$, $\delta = 2e^{\Delta}\rho\|PA^{-1}B\|/\delta$, $\varphi = e^{\Delta}\bar{\lambda}(P)$, $\nu = \kappa\ell\Delta^2/2\rho$,

$\gamma = \min\{\rho/4\Delta, \kappa\delta\mu/8\rho\}$. By Lemma A.2 it follows that M is positive definite when $\varepsilon < \bar{\varepsilon}_\sigma$ for some $b < \gamma/\nu = \min\{\delta\mu/4\ell\Delta^2, \rho^2/2\kappa\ell\Delta T\}$, which proves (A.2). Next, we show (A.3). During jumps we have:

$$\begin{aligned}
V_1(u^+) - V_1(u) &= \frac{1}{2}\bigg(\|u_1 - u^*\|^2 + \frac{\kappa\delta^2}{\rho}(f(u) - f(u^*)) \\
&\quad - \|u_2 - u_1\|^2 - \|u_2 - u^*\|^2 - \frac{\kappa\Delta^2}{\rho}(f(u) - f(u^*))\bigg) \\
&\leq -\frac{1}{2}\bigg(\|u_2 - u_1\|^2 + \|u_2 - u^*\|^2 \\
&\quad - \frac{\kappa\Delta T^2}{\rho}\underbrace{\left(1 - \frac{\delta^2}{\Delta^2} - \frac{2\rho}{\kappa\mu\Delta^2}\right)}_{:=\gamma_0}(f(u) - f(u^*))\bigg),
\end{aligned}$$

where we used $\|u_1 - u^*\| \leq 2/\mu(f(u) - f(u^*))$, which follows from the strong convexity assumption (Assumption 6). Notice that $\gamma_0 < 1$ by construction, and the choice of (Δ, δ) guarantees that $\gamma_0 > 0$. Hence, during jumps $V_1(u^+) - V_1(u) \leq -\gamma_0 V_1(u)$. Similarly,

$$\begin{aligned}
V_2(\tilde{x}^+) - V_2(\tilde{x}) &= (e^\delta - e^\Delta)\tilde{x}^\mathsf{T} P \tilde{x} \\
&= (e^{\delta - \Delta + \Delta} - e^\Delta)\tilde{x}^\mathsf{T} P \tilde{x} \\
&= -\underbrace{(1 - e^{\delta - \Delta})}_{:=\eta_0} V_2(\tilde{x}),
\end{aligned} \tag{38}$$

where $0 < \eta_0 < 1$ because $\Delta > \delta$. Using the bounds during jumps for V_1 and V_2 we conclude

$$V(\tilde{z}^+) \leq \max\{1 - \gamma_0, 1 - \eta_0\} V(\tilde{z}), \tag{39}$$

which proves that condition (A.3) is satisfied with $c = \max\{\ln(\frac{\Delta^2\kappa\mu^2}{\delta\kappa\mu^2 - 2\rho}), \Delta - \delta\}$. Finally, the statement follows by application of Lemma A.1 in the Appendix. ∎

5.6 Proof of Corollary 4.4

The closed-loop system (5) and (12)-(16). Consider the Lyapunov function $W(\tilde{z}) = e^{\varrho\tau} V_\sigma(\tilde{z})$, where $V_\sigma(\tilde{z})$ is defined as in (21) for each $\sigma \in \mathcal{S}$, and $\varrho > 0$. First, we show that $W(\tilde{z})$ satisfies (A.1). To this aim, note that

$$\bar{a}_\sigma e^{N_0\varrho}\|\tilde{z}\|^2 \geq e^{N_0\varrho} V_\sigma(\tilde{z}) \geq W(\tilde{z}) \geq e^{0\varrho} V_\sigma(\tilde{z}) \geq \underline{a}_\sigma \|z\|^2,$$

where $\underline{a}_\sigma$ and $\bar{a}_\sigma$ are as in (18). Second, we show that $W(\tilde{z})$ satisfies (A.2). To this aim, note that

$$\dot{W}(\tilde{z}) = \varrho e^{\varrho\tau} V_\sigma(\tilde{z})\dot{\tau} + e^{\mu\tau}\dot{V}_\sigma(\tilde{z}) \leq (\varrho/\tau_d - b_\sigma) W(\tilde{z})$$

where we used $\dot{\tau} \leq 1/\tau_d$, and $\dot{V}_\sigma(\tilde{z}) \leq -b_\sigma V_\sigma(\tilde{z})$ with $b_\sigma = \min\{\delta\mu/4\ell\Delta^2, \rho^2/2\kappa\ell\Delta\}$, which follows from the proof of Corollary 4.3. Hence, (A.2) is satisfied when $\varrho < b_\sigma\tau_d$. Third, we show that W satisfies (A.3). During switches of the plant we have:

$$\begin{aligned}
W(\tilde{z}^+) &\leq e^{\varrho\tau^+} \max_\sigma V_\sigma(\tilde{z}^+) \leq e^{\varrho(\tau - 1)} \max_\sigma V_\sigma(\tilde{z}) \\
&\leq e^{\varrho(\tau-1) + \ln \bar{a}_\sigma}\|\tilde{z}\|^2 \leq e^{-\varrho + \ln \bar{a}_\sigma - \ln \underline{a}_\sigma} W(\tilde{z}).
\end{aligned} \tag{40}$$

Similarly, during jumps of the controller:

$$\begin{aligned}
W(\tilde{z}^+) &\leq e^{\varrho\tau} \max_\sigma V_\sigma(\tilde{z}^+) \leq e^{\varrho\tau} \max_\sigma e^{-c_\sigma} V_\sigma(\tilde{z}) \\
&\leq W(\tilde{z}) \max_\sigma e^{-c_\sigma} \leq W(\tilde{z}),
\end{aligned} \tag{41}$$

where the second inequality follows from (39) with $c_\sigma = \max\{\ln(\frac{\Delta^2\kappa\mu^2}{\delta\kappa\mu^2 - 2\rho}), \Delta - \delta\}$ the third inequality follows from the definition of $W(\tilde{z})$, and the last inequality follows from $c_\sigma > 0$. Hence, $W(z)$ satisfies (A.3) when $\varrho > \ln \bar{a}_\sigma - \ln \underline{a}_\sigma$. By combining the upper and lower bounds on ρ we conclude that W satisfies the assumptions of Lemma A.1 when $\tau_d > (\ln \bar{a}_\sigma - \ln \underline{a}_\sigma)/b_\sigma$ for all $\sigma \in \mathcal{S}$. The result follows by iterating the bound in Lemma A.1 to both switches in the plant and to restarts of the algorithm. ∎

5.7 Proof of Theorem 4.2

We begin by shifting the equilibrium points of the plant to the origin by using the change of variables $\tilde{x} := x - \mathcal{M}(u)$, where $\mathcal{M}(u) := -A^{-1}Bu - A^{-1}Ew$. In the new variables $\tilde{z} = \mathrm{vec}(\tilde{x}, u)$, the plant dynamics read as:

$$\dot{\tilde{x}} = A\tilde{x} + \varepsilon\frac{\rho}{u_3}A^{-1}B(u_2 - u_1) + \varepsilon A^{-1}E\dot{w},$$

where we note that w is a time-varying quantity.

Proof of (i) Since we focus on a single mode, in what follows we drop the dependence on the subscript σ. We prove this claim by showing that the Lyapunov function V defined the proof of Corollary 4.3 satisfies the assumptions of Lemma A.1 in the Appendix. First, note that during flows:

$$\begin{aligned}
\frac{\partial V_1(u)}{\partial w}\dot{w} &= \frac{\kappa u_3^2}{2\rho} H^\mathsf{T}(\nabla g(Gu + Hw) - \nabla g(Gu^* + Hw))\dot{w} \\
&\leq \frac{\kappa u_3^2 \ell_y}{2\rho}\|H\|\|G\|\|\dot{w}\|\|u_1\|_{\mathcal{A}_f} \\
&\leq \frac{\sqrt{2}\kappa u_3^2 \ell_y}{2\rho}\|H\|\|G\|\|\dot{w}\|\|u\|_{\tilde{\mathcal{A}}_c},
\end{aligned}$$

where the second inequality follows from Assumption 3 and for the last inequality we used $\|u_1 - u^*\| \leq \|u_1 - u^*\| + \|u_2\|_\mathcal{A} \leq \sqrt{2}(\|u_1 - u^*\|^2 + \|u_2 - u^*\|^2)^{0.5}$. Moreover, by using inequality (36) for the remaining terms we conclude that during flows $\dot{V}_1(u)$ satisfies

$$\begin{aligned}
\dot{V}_1(u) \leq &-\frac{1}{4}\min\{\frac{\rho}{\Delta}, \frac{\kappa\delta\mu}{2\rho}\}\|u\|^2_{\tilde{\mathcal{A}}_c} + \eta\|\tilde{x}\|\|u\|_{\tilde{\mathcal{A}}_c} \\
&+ \frac{\sqrt{2}\kappa\ell_y u_3^2}{2\rho}\|H\|\|G\|\|\dot{w}\|\|u\|_{\tilde{\mathcal{A}}_c}.
\end{aligned} \tag{42}$$

The time-derivative of $V_2(\tilde{x})$ satisfies:

$$\begin{aligned}\dot{V}_2(\tilde{x}) &\leq -\frac{1}{\varepsilon}\underline{\lambda}(Q)e^{\delta}\|\tilde{x}\|^2 \\ &\quad + e^{\Delta}\left(2\frac{\rho}{\delta}\|PA^{-1}B\|\|\tilde{x}\|\|u\|_{\mathcal{A}} + \bar{\lambda}(P)\|\tilde{x}\|^2\right) \\ &\quad + 2\|PA^{-1}E\|\|\tilde{x}\|\|\dot{w}\|, \end{aligned} \tag{43}$$

where we used (37). Let $\xi = \operatorname{vec}\left(\|x\|, \|u\|_{\tilde{\mathcal{A}}_c}\right)$; by combining (42)-(43) and the quadratic bounds of V, we have:

$$\dot{V}_\sigma + bV_\sigma \leq -\xi^{\mathsf{T}}M\xi + \|\dot{w}\|r^{\mathsf{T}}\xi \tag{44}$$

where $r := [2\theta\|PA^{-1}E\|, (1-\theta)\frac{\sqrt{2}\kappa\ell_y\Delta^2}{2\rho}\|H\|\|G\|]^{\mathsf{T}}$ and M is a symmetric matrix with form as in Lemma A.2 with parameters: $\alpha = e^{\delta}\underline{\lambda}(Q)$, $\beta = e^{\Delta}\bar{\lambda}(P)$, $\eta = 2\sqrt{2}\kappa\Delta\ell_y\|C\|\|G\|/\rho$, $\delta = 2e^{\Delta}\rho\|PA^{-1}B\|/\delta$, $\varphi = e^{\Delta}\bar{\lambda}(P)$, $\nu = \kappa\ell\Delta^2/2\rho$, $\gamma = \min\{\rho/4\Delta T, \kappa\delta\mu/8\rho\}$. By using Lemma A.2, we conclude that when $\varepsilon < \bar{\varepsilon}_\sigma$, then there exists $b < \gamma/\nu$ such that M is positive definite. Next, we show that the quadratic term in (44) dominates the linear term for large $\|\xi\|$. To this aim, we rewrite (44) as follows:

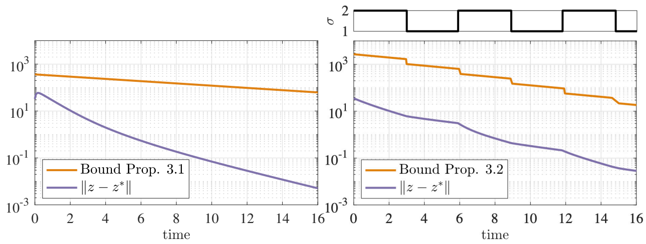
$$\begin{aligned}\dot{V}_\sigma + bV_\sigma &\leq -\xi^{\mathsf{T}}(M-kI)\xi - k\|\xi\|^2 + \|\dot{w}\|\|r\|\|\xi\|, \\ &\leq -\xi^{\mathsf{T}}(M-kI)\xi, \end{aligned} \tag{45}$$

where $0 < k < \underline{\lambda}(M)$, and where the last inequality holds when $-k\|\xi\|^2 + \|\dot{w}\|\|r\|\|\xi\| < 0$. By recalling that $\|\xi\| \geq 1/\sqrt{2}\|\tilde{z}\|$, we conclude that (A.2) is satisfied for all $\|\tilde{z}\| \geq \sqrt{2}\|r\|/k$. Finally, (A.2) follows from (39) since the disturbance w is continuous during jumps. To conclude, the statement follows by application of Lemma A.1 in the Appendix.

Proof of (ii) Consider the Lyapunov function $W(\tilde{z}) = e^{\varrho\tau}V_p(\tilde{z})$, where $V_p(\tilde{z})$ is defined as in (21) for some $\sigma \in \mathcal{S}$, and $\varrho > 0$. By the quadratic upper and lower bounds of V_σ we conclude that W satisfies (A.1) with $\underline{a}_\sigma\|\tilde{z}\|^2 \leq W(\tilde{z}) \leq \bar{a}_\sigma e^{N_0\varrho}\|\tilde{z}\|^2$. Second, we show that W satisfies (A.2). To this aim, note that:

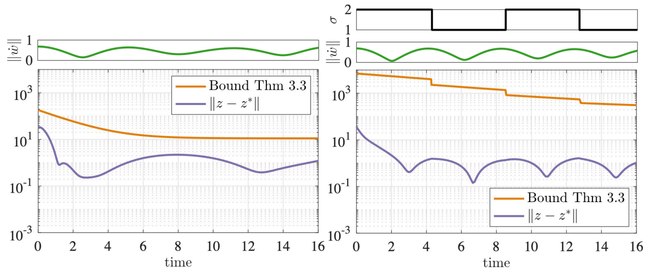
$$\begin{aligned}\dot{W}(\tilde{z}) = \varrho e^{\varrho\tau}(V_\sigma(\tilde{z})\dot{\tau} + \dot{V}_\sigma(\tilde{z})) &\leq e^{\varrho\tau}(\frac{\varrho}{\tau_d}V_\sigma(\tilde{z}) + -bV_\sigma(\tilde{z})) \\ &\leq (\varrho/\tau_d - b_p)\,W(z), \end{aligned} \tag{46}$$

where the second inequality follows from (45), and holds when $\|\tilde{z}\| \geq \sqrt{2}\|r\|/k$. Thus, the above inequality proves that (A.2) is satisfied when $\varrho < b_\sigma\tau_d$. Third, we show that $W(\tilde{z})$ satisfies (A.3). During switches of the plant we have $W(\tilde{z}^+) \leq e^{-\varrho + \ln\bar{a}_\sigma - \ln\underline{a}_\sigma}W(\tilde{z})$, where we used (40) and the fact that the disturbance w is continuous during the jumps. Similarly, when $u_3 = \Delta$ w have $W(\tilde{z}^+) \leq e^{c + \ln\bar{a}_\sigma - \ln\underline{a}_\sigma}W(\tilde{z})$, where we used (41) and the fact that w is continuous at controller restarts. Thus, $W(\tilde{z})$ satisfies (A.3) when $\varrho > \ln\bar{a}_\sigma - \ln\underline{a}_\sigma$. By combining the upper and lower bounds on ϱ we conclude that W satisfies the assumptions of Lemma A.1 when $\tau_d > (\ln\bar{a}_\sigma - \ln\underline{a}_\sigma)/b_\sigma$. The result follows by applying the bound in Lemma A.1 to both switches in the plant and to restarts of the control algorithm. ■



(a) (b)

Fig. 3. Regulation error of gradient flow controller (8). (a) No plant switching. (b) Plant switches between two modes.



(a) (b)

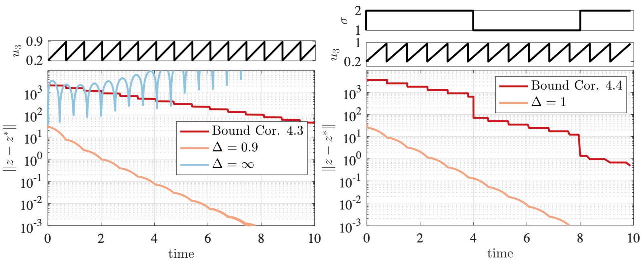
Fig. 4. Tracking error of gradient flow controller (8). (a) No plant switching. (b) Plant switches between two modes.

6 Simulations Results

In this section, we provide illustrative numerical results that validate our theoretical findings. We consider a plant with two modes (i.e. $\mathcal{S} = \{1,2\}$), $n = 10$ states, $m = 5$ control inputs, $p = 5$ outputs, $q = 6$ exogenous disturbances, and we focus on an instance of (3) characterized by the following cost function $h(u) = u^{\mathsf{T}}Ru$, $g(y) = (y - y_{\mathrm{ref}})^{\mathsf{T}}Q(y - y_{\mathrm{ref}})$ where $R \in \mathbb{R}^{m\times m}$, $R \succ 0$, $Q^{p\times p}$, $Q \succ 0$, and $y_{\mathrm{ref}} \in \mathbb{R}^p$ is a constant reference signal.

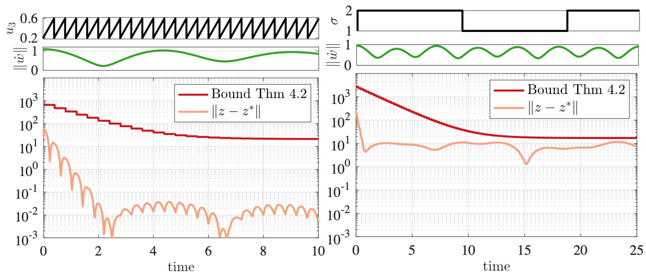
Gradient flow controller. We first illustrate the performance of the gradient flow controller discussed in Section 3. Fig. 3(a) illustrates the regulation bound in Theorem 3.1 when the disturbance w and the switching signal σ are constant. Fig. 3(b) illustrates the regulation bound in Theorem 3.2 when the switching signal is time-varying. We observe that in this case the bound in Theorem 3.2 is a function that is decreasing in both time t and the number of jumps of σ. Fig. 4(a)-(b) illustrate the tracking error characterized in Theorem 3.3 for constant σ and for time-varying σ, respectively.

Accelerated Gradient controller. We now illustrate the performance of the accelerated gradient controller discussed in Section 4. Fig. 5(a) illustrates the regulation bound in Theorem 4.3 when the switching signal



(a) (b)

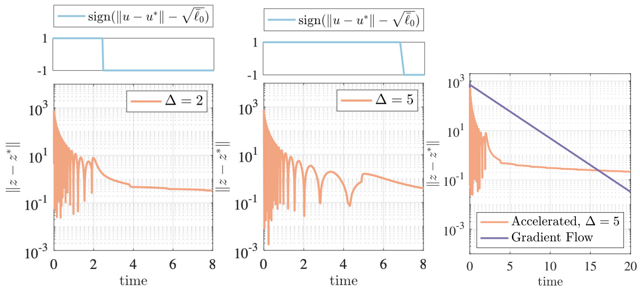
Fig. 5. Regulation error of accelerated gradient controller. (a) No plant switching. (b) Switched plant.



(a) (b)

Fig. 6. Regulation error of accelerated gradient controller. (a) No plant switching. (b) Switched plant.

σ is constant. The figure illustrates two restarting policies: finite Δ chosen as in the statement of Theorem 4.2, and $\Delta = \infty$. As illustrated by the figure (and previously pointed out in [3] via a numerical example), accelerated gradient controllers that are in feedback with a dynamical plant can generate unstable unstable when $\Delta = \infty$. Differently, the presence of restarts ($\Delta < \infty$) allows us to prove uniform stability of the accelerated gradient controller proposed in this work. Fig. 5(b) illustrates the regulation bound in Theorem 4.4 for time-varying switching signals. Fig. 4(a)-(b) illustrates the tracking error characterized in Theorem (4.2) for constant σ and time-varying σ, respectively. We note that, in this simulation, the time horizon has been rescaled in order to illustrate the behavior of the controller under switching of the plant. Finally, Fig. 7 illustrates the convergence of the accelerated gradient method when the cost function satisfies the reverse Lipschitz condition (Assumption 5). To this aim, we consider a scalar plant ($n = m = p = 1$) composed of a single mode ($|\mathcal{S}| = 1$) and cost function described by: $f(u) = 1/4(Gu + Hw - y_{\text{ref}})^4$, where $y_{\text{ref}} \in \mathbb{R}$ is a constant reference. Notice that the gradient of $f(u)$ satisfies Assumption 3 on compact sets. Two important observations follow from Fig. 7. First, the simulations suggest that admissible smaller restarting times Δ, in general, result in faster convergence. Second, the simulation illustrates that the controller ensures converges only to a neighborhood of the optimal points, thus showing that the set of optimal points is only practically asymptotically stable. The residual neighborhood can be characterized as follows. Assumption 5 implies $(u - u_{\text{ref}})^2 > \nu_0^2$, which shows that $\ell_0 = \nu_0^2$, i.e., Assumption 5 holds. Finally, Fig. 7(c) compares the regulation errors of the gradient flow controller and of the Accelerated Gradient controller. The figure illustrates that: (i) the accelerated gradient controller, in general, ensures *faster* convergence, but only to a neighborhood of the optimal points, and (ii) the gradient flow controller converges *exactly* to the actual optimal points, thus overcoming the lack of asymptotic stability suffered by the accelerated controller. However, in this case the convergence is slower. Thus, our results seem to uncover the existence of a trade-off between acceleration and precise convergence properties in online optimization with dynamic momentum.



(a) (b) (c)

Fig. 7. Regulation error of accelerated gradient controller with polynomial cost function. (a) $\Delta = 2$. (b) $\Delta = 5$. (c) Comparison with Gradient Flow.

7 Conclusions

This paper introduced two classes of controllers to steer the output of a switched dynamical system to the solution of an optimization problem despite unknown and time-varying disturbances. We demonstrated how online gradient flow controllers can be properly adapted to cope with the switched nature of the underlying dynamical plant. We showed that a restarting mechanism can overcome the lack of robustness suffered by accelerated optimization methods when applied to an online setting. Overall, our results show that the controllers ensure exponential tracking of the optimal solutions when the controller operates at a timescale that is sufficiently slower than the plant, and the switching is slow on the average. Future research directions will focus on the development of accelerated control algorithms for online optimization on Riemannian manifolds

A Auxiliary Lemmas

The following Lemmas will be instrumental for our results. The first one provides sufficient Lyapunov conditions to certify exponential ISS in a class of hybrid systems with jumps triggered by timers and inputs only affecting the flows. The proof follows similar ideas as in [37], [26], and [38]. To apply the Lemma in the proofs of Section 3 the state s can be omitted and G can taken as the identity function.

Lemma A.1 *Consider a HDS with states* $\Lambda = (\sigma, \tau, \phi, s)$*, where the dynamics of* (σ, τ) *are given by* (5)*,*

$\dot{\phi} = F_\sigma(\phi, s, u)$, $\phi^+ = G(\phi, s)$, $\dot{s} = 1$ and $s^+ = \delta$, having flow set and jump set given by $C := \mathcal{T}_C \times \mathbb{R}^r \times [\delta, \Delta]$, $D := D_1 \cup D_2$, $D_1 := \mathcal{T}_D \times \mathbb{R}^r \times [\delta, \Delta]$, $D_2 := \mathcal{T}_C \times \mathbb{R}^r \times \{\Delta\}$, where F_σ and G are continuous on $\mathbb{R}^r \times [\delta, \Delta] \times \mathbb{R}^m$ and $\mathbb{R}^r \times \{\Delta\}$, respectively. Let $\mathcal{A} = \mathcal{T}_c \times \mathcal{A}_\phi \times [\delta, \Delta]$, where $\mathcal{A}_\phi \subset \mathbb{R}^r$ is compact. Suppose there exists a continuously differentiable function $V : \mathbb{R}^n \to \mathbb{R}$ such that

$$\underline{a}\|\Lambda\|^2_{\mathcal{A}} \leq V(\Lambda) \leq \bar{a}\|\Lambda\|^2_{\mathcal{A}}, \quad \forall\, \Lambda \in C \cup D, \tag{A.1}$$

$$\dot{V} \leq -bV(\Lambda), \qquad \forall \Lambda \in C \text{ s.t } \|\Lambda\| \geq b_0\|u\|, \tag{A.2}$$

$$V(x) \leq \exp(-c)V(x), \quad \forall \Lambda \in D, \tag{A.3}$$

where $\underline{a}$,$\bar{a}$,b,c,$b_0 > 0$. Then every complete solution of the complete hybrid dynamical system satisfies $\|\Lambda(t,j)\|_{\mathcal{A}} \leq \sqrt{\bar{a}/\underline{a}} \exp\left(-\frac{1}{2}(bt + cj)\right) \|\Lambda(0,0)\|_{\mathcal{A}} + \sqrt{\bar{a}/\underline{a}}\; b_0\|u\|$.

Proof: The proof extends the technical arguments of [37, Prop. 2]. We distinguish among three cases.
(Case 1) $\|\Lambda(s,k)\|_{\mathcal{A}} > b_0\|u\|$ for all $0 \leq s \leq t$ and $0 \leq k \leq j$. Since (A.2) is a linear differential equation, the time-evolution of $V_p(\Lambda)$ during flows (i.e. $\Lambda \in C$) reads $V(\Lambda(t,j)) \leq \exp(-bt)V(\Lambda(0,j))$. During jumps (i.e. $\Lambda \in D$), (A.3) implies $V(\Lambda(t,j)) \leq \exp(-c)V(\Lambda(t,j-1))$. By combining the two inequalities we obtain $V(\Lambda(t,j)) \leq \exp(-bt-cj)V(\Lambda(0,0))$; then, by using (A.1), we obtain:

$$\begin{aligned}\|\Lambda(t,j)\|^2 &\leq \frac{1}{\underline{a}}V(\Lambda(t,j)) \leq \frac{1}{\underline{a}}\exp(-bt-cj)V(\Lambda(0,0)) \\ &\leq \frac{\bar{a}}{\underline{a}}\exp(-bt-cj)\|\Lambda(0,0)\|^2. \end{aligned}\tag{A.4}$$

(Case 2) $\|\Lambda(s,k)\| = b_0\|u\|$ for some $s \geq 0$ and $k \geq 0$. We show this condition implies $\|\Lambda(t,j)\| \leq b_0\|u\|$ for all $t \geq s$ and $j \geq k$. First, (A.2) implies $\dot{V}(\Lambda(s,k)) < 0$, which shows that $\|x(s,k)\|$ is decreasing during flow and thus $\|\Lambda(t,k)\| \leq b_0\|u\|$ for all $t \geq s$. Second, (A.3) implies $V(\Lambda(s,k+1)) \leq e^{-c}V(\Lambda(s,k))$ which shows that $\Lambda(s,j) \leq b_0$ for all $j \geq k$.
(Case 3) $\|\Lambda(t,j)\| \leq b_0\|u\|$ $\forall\, t \geq s$ and $j \geq k$. Then,

$$\|\Lambda(t,j)\|^2 \leq \frac{1}{\underline{a}}V(\Lambda(t,j)) \leq \frac{\bar{a}}{\underline{a}}b_0^2\|u\|^2. \tag{A.5}$$

Finally, the claim follows by combining (A.4)-(A.5). ■

Lemma A.2 *Let $\alpha, \beta, \eta, \varphi, \delta, \gamma, \nu$ be positive scalars, let $0 < \theta < 1$ and $b > 0$ be free parameters, and let*

$$M = \begin{bmatrix} \theta(\frac{\alpha}{\varepsilon} - \beta - b\varphi) & -\frac{1}{2}((1-\theta)\eta + \theta\delta) \\ -\frac{1}{2}((1-\theta)\eta + \theta\delta) & (1-\theta)(\gamma - b\nu) \end{bmatrix}.$$

If $0 < \varepsilon < \alpha\gamma/(\beta\gamma + \eta\delta)$, then there exists $b < \gamma/\nu$ and $\theta \leq \eta/(\eta + \delta)$, such that M is positive definite.

Proof: Matrix M is positive definite if and only if its leading principal minors are positive, that is, $(1-\theta)(\gamma - b\nu) > 0$ and $\theta(1-\theta)(\gamma - b\nu)(\frac{\alpha}{\varepsilon} - \beta - b\varphi) > \frac{1}{4}((1-\theta)\eta + \theta\delta)^2$. The first inequality implies that $b < \gamma/\nu$ is a necessary condition for M to be positive definite. The second inequality can be rewritten as:

$$\varepsilon < \frac{\alpha(\gamma - b\nu)}{(\gamma - b\nu)(\beta + b\varphi) + \frac{((1-\theta)\eta + \theta\delta)^2}{4\theta(1-\theta)}} := \hat{\varepsilon}(\theta, b).$$

The function $\hat{\varepsilon}(\theta, b)$ achieves its maximum at $\theta = \bar{\theta} := \eta/(\eta + \delta)$ and $b = \bar{b} = 0$, with $\hat{\varepsilon}(\bar{\theta}, \bar{b}) = \alpha\gamma/(\beta\gamma + \eta\delta)$. Further, $\hat{\varepsilon}(\theta, b)$ achieves a minimum at $\theta = \underline{\theta} := 0$ and $b = \underline{b} := \gamma/\nu$, with $\hat{\varepsilon}(\underline{\theta}, \underline{b}) = 0$. Finally, the statement follows from continuity of $\bar{\varepsilon}(\theta, b)$ in its parameters. ■

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 (9) (2009) 2250–2254.

[2] M. Colombino, E. Dall'Anese, A. Bernstein, Online optimization as a feedback controller: Stability and tracking, IEEE Transactions on Control of Network Systems 7 (1) (2020) 422–432.

[3] A. Hauswirth, S. Bolognani, G. Hug, F. Dörfler, Timescale separation in autonomous optimization, IEEE Transactions on Automatic Control(In press) (2020).

[4] 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.

[5] 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.

[6] L. S. Lawrence, J. W. Simpson-Porco, E. Mallada, Linear-convex optimal steady-state control, arXiv preprint (2018) arXiv:1810.12892.

[7] 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.

[8] G. Bianchin, F. Pasqualetti, Gramian-based optimization for the analysis and control of traffic networks, IEEE Transactions on Intelligent Transportation Systems 21 (7) (2020) 3013–3024.

[9] R. Kutadinata, W. Moase, C. Manzie, L. Zhang, T. Garoni, Enhancing the performance of existing urban traffic light control through extremum-seeking, Transp. Research Pt. C: Emerging Technologies 62 (2016) 1–20.

[10] S. Bolognani, S. Zampieri, A distributed control strategy for reactive power compensation in smart microgrids, IEEE Transactions on Automatic Control 58 (11) (2013) 2818–2833.

[11] A. Bernstein, E. Dall'Anese, and A. Simonetto, Online primal-dual methods with measurement feedback for time-varying convex optimization, IEEE Transactions on Signal Processing 67 (8) (2019) 1978–1991.

[12] C.-Y. Chang, M. Colombino, J. Cortés, E. Dall'Anese, Saddle-flow dynamics for distributed feedback-based optimization, IEEE Control Systems Letters 3 (4) (2019) 948–953.

[13] M. Nonhoff, M. A. Müller, An online convex optimization algorithm for controlling linear systems with state and input constraints, arXiv preprint (2020) arXiv:2005.11308.

[14] N. Agarwal, B. Bullins, E. Hazan, S. M. Kakade, K. Singh, Online control with adversarial disturbances, arXiv preprint (2019) arXiv:1902.08721.

[15] D. Liberzon, Switching in systems and control, Springer Science, 2003.

[16] J. P. Hespanha, A. S. Morse, Stability of switched systems with average dwell-time, in: IEEE Conf. on Decision and Control, Phoenix, AZ, USA, 1999, pp. 2655–2660.

[17] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Oklady Akademii Nauk SSSR 269 (1983) 543–547.

[18] W. Su, S. Boyd, E. Candes, A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights, in: Advances in Neural Information Processing Systems 27, Curran Associates, 2014, pp. 2510–2518.

[19] J. I. Poveda, N. Li, Inducing uniform asymptotic stability in non-autonomous accelerated optimization dynamics via hybrid regularization, in: IEEE Conf. on Decision and Control, Nice, France, 2019, pp. 3000–3005.

[20] H. K. Khalil, Nonlinear Systems, 2nd Edition, Prentice Hall, Upper Saddle River, NJ, 2002.

[21] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid dynamical systems: modeling stability, and robustness, Princeton University Press, Princeton, NJ, 2012.

[22] 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.

[23] J. I. Poveda, N. Li, Robust hybrid zero-order optimization algorithms with acceleration via averaging in time, AutomaticaTo appear (2020).

[24] D. M. Hustig-Schultz, R. G. Sanfelice, A robust hybrid heavy ball algorithm for optimization with high performance, American Control Conference (2019) 151–156.

[25] C. Cai, A. R. Teel, R. Goebel, Smooth Lyapunov functions for hybrid systems part ii: (Pre) asymptotic stable compact sets, IEEE Transactions on Automatic Control 53 (53) (2008) 734–748.

[26] D. Nešić, A. R. Teel, G. Valmorbida, L. Zaccarian, Finite-gain $\mathcal{L}_p$ stability for hybrid dynamical systems, Automatica 49 (2013) 2384–2396.

[27] F. Forni, D. Nešić, L. Zaccarian, Reset passivation of nonlinear controllers via a suitable time-regular reset map, Automatica (2011) 2099–2106.

[28] J. B. Rejeb, I. Morarescu, A. Girard, J. Daafouz, Stability analysis of a general class of singularly perturbed linear hybrid systems, Automatica (2018) 98–108.

[29] R. G. Sanfelice, A. R. Teel, On singular perturbations due to fast actuators in hybrid control systems, Automatica 47 (4) (2011) 692–701.

[30] W. Wang, A. R. Teel, D. Nešić, Averaging in singularly perturbed hybrid systems with hybrid boundary layer systems, IEEE Conf. on Decision and Control 6855-6860 (2012).

[31] D. Liberzon, D. Nešić, A. R. Teel, Lyapunov-based small gain theorems for hybrid systems, IEEE Transactions on Automatic Control (2014) 1395–1410.

[32] B. O'Donoghue, E. Candes, Adaptive restart for accelerated gradient schemes, Foundations of computational mathematics 15 (3) (2015) 715–732.

[33] A. Wibisono, A. C. Wilson, M. I. Jordan, A variational perspective on accelerated methods in optimization, Proceedings of the National Academy of Sciences 113 (47) (2016) E7351–E7358.

[34] J. I. Poveda, A. R. Teel, The heavy-ball ode with time-varying damping: Persistence of excitation and uniform asymptotic stability, American Control Conference (2020) 773–778.

[35] C. Prieur, I. Queinnec, S. Tarbouriech, L. Zaccarian, Analysis and synthesis of reset control systems, Foundations and Trends in Systems and Control 6 (2018) 117–338.

[36] D. Nešić, L. Zaccarian, A. R. Teel, Stability properties of reset systems, Automatica 44 (2008) 2019–2026.

[37] G. Yang, D. Liberzon, Input-to-state stability for switched systems with unstable subsystems: A hybrid lyapunov construction, in: IEEE Conf. on Decision and Control, 2014, pp. 6240–6245.

[38] C. Cai, A. R. Teel, Characterizations of input-to-state stability for hybrid systems, Systems & Control Letters 58 (1) (2009) 47 – 53.

Gianluca Bianchin is a Postdoctoral Scholar in the Department of Electrical, Computer, and Energy Engineering at the University of Colorado at Boulder. He received the Ph.D. degree in Mechanical Engineering at The University of California at Riverside in 2020, the Laurea degree in Information Engineering and the Laurea Magistrale degree (Summa Cum Laude) in Controls Engineering from the University of Padova, Italy, in 2012 and 2014.

Jorge I. Poveda is an Assistant Professor in the Department of Electrical, Computer, and Energy Engineering at the University of Colorado, Boulder. He received in 2016 and 2018 the M.Sc. and Ph.D. degrees in Electrical and Computer Engineering from the University of California at Santa Barbara. He was a Postdoctoral Fellow at Harvard University in 2018. He has received the CCDC Outstanding Scholar Fellowship and the National Science Foundation Career Research Initiation Award (CRII) in 2020.

Emiliano Dall'Anese is an Assistant Professor in the Department of Electrical, Computer, and Energy Engineering at the University of Colorado, Boulder. He received the Ph.D. in Information Engineering from the Department of Information Engineering, University of Padova, Italy, in 2011. From January 2011 to November 2014, he was a Postdoctoral Associate at the Department of Electrical and Computer Engineering of the University of Minnesota, and from December 2014 to July 2018 he was a Senior Researcher at the National Renewable Energy Laboratory. He received the National Science Foundation CAREER Award in 2020.