Annual Reviews in Control 56 (2023) 100926

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Full length article

Multi-time scale control and optimization via averaging and singular perturbation theory: From ODEs to hybrid dynamical systems

Mahmoud Abdelgalil, Daniel E. Ochoa, Jorge I. Poveda [*,1]

Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, 92093, USA

ARTICLE INFO

Keywords:
Averaging and singular perturbation theory
Adaptive systems
Stability analysis
Hybrid dynamical systems
Extremum-seeking systems

ABSTRACT

Multi-time scale techniques based on singular perturbations and averaging theory are among the most powerful tools developed for the synthesis and analysis of feedback control algorithms. This paper introduces some of the recent advances in singular perturbation theory and averaging theory for continuous-time dynamical systems modeled as ordinary differential equations (ODEs), as well as for hybrid dynamical systems that combine continuous-time dynamics and discrete-time dynamics. Novel multi-time scale analytical tools based on higher-order averaging and singular perturbation theory are also discussed and illustrated via different examples. In the context of hybrid dynamical systems, a class of sufficient Lyapunov-based conditions for global stability results are also presented. The analytical tools are illustrated through various new architectures and algorithms within the context of adaptive and extremum-seeking systems. These tools are suitable for the study of model-free optimization and stabilization problems that require the synergistic use of continuous-time and discrete-time feedback. The paper aims to acquaint the reader with a range of modern tools for studying multi-time scale phenomena in optimization and control systems, providing some guidelines for future research in this field.

1. Introduction

Singular perturbation and averaging methods are widely recognized and extensively employed perturbation techniques in the study of multiple time scales within control theory. In this introductory section, we first provide a brief survey of the recent literature on singular perturbation and averaging theory in the control's literature, with a focus on highlighting new themes and novel applications. Detailed discussions on particular technical developments (stability techniques, regularity assumptions, etc.) are deferred to the main sections of the paper.

1.1. Singular perturbation theory

The popularization of singular perturbation methods in control theory, at least in the West, can be fairly attributed to the seminal work of Kokotovic and collaborators during the 1960s and 1970s (Kokotović, O'Malley, & Sannuti, 1976; Kokotović & Sannuti, 1968; Sannuti & Kokotović, 1969), who showed that the behavior of complex feedback architectures with multiple (stable) time scales could be approximately predicted by a "reduced" model operating in the slowest time scale. Since then, the use of singular perturbation in control theory has steadily grown. A historical account of the development of singular perturbation theory is given in the recent monograph by O'Malley (2014), and a comprehensive collection of developments in singular perturbation for control systems between 2002 and 2012 is given in the article (Zhang, Naidu, Cai, & Zou, 2014), and also in the recent monograph (Narang-Siddarth & Valasek, 2014). Singular perturbation techniques remain one of the most indispensable tools for the analysis and design of control systems.

Classically, singular perturbation theory has a well-established pedigree in the analysis and design of high-gain feedback control and high-gain observers, an enduring subject of investigation in control theory (Khalil, 2017). In this context, it has been recently applied to regulate nonminimum phase systems (Huang, Khalil, & Song, 2019), as well as to address output feedback challenges encountered in systems with sensors modeled by partial differential equations (Ahmed-Ali, Lamnabhi-Lagarrigue, & Khalil, 2023). Another classical application of singular perturbation techniques is in (near-)optimal control problems (Sannuti & Kokotović, 1969). In a recent development, a synergy between singular perturbation techniques and reinforcement learning for solving optimal control problems emerged, see for

* Corresponding author.
E-mail addresses: mabdelgalil@ucsd.edu (M. Abdelgalil), dochotamayo@ucsd.edu (D.E. Ochoa), poveda@ucsd.edu (J.I. Poveda).

[1] This work was supported in part by the Air Force Office of Scientific Research grant FA9550-22-1-0211 and the National Science Foundation grants CAREER ECCS 2305756 and CMMI 2228791.

https://doi.org/10.1016/j.arcontrol.2023.100926
Received 31 July 2023; Received in revised form 7 November 2023; Accepted 7 November 2023
Available online 24 November 2023

example (Mukherjee, Bai, & Chakrabortty, 2020; Xue et al., 2019). In multi-agent systems, singular perturbation remains an important analysis and design tool. It has been applied to the study of consensus in sparse large-scale networks (Chowdhury & Khalil, 2017; Martin, Morărescu, & Nešić, 2016), in distributed resource allocation (Liang, Zeng, & Hong, 2018), in characterizing the convergence rates of distributed consensus over cluster networks (Dutta, Boker, & Doan, 2022), in the study of synchronization (Kong, Ni, Zhu, Hu, & Huang, 2023), and most recently in non-convex distributed optimization (Carnevale & Notarstefano, 2023). A fascinating avenue where singular perturbation has found new horizons is within the control-theoretic approach to synthetic biology (Del Vecchio, Dy, & Qian, 2016). In this context, it has been effectively employed to study modularity within biological systems (Rivera-Ortiz & De Vecchio, 2014). Additionally, it has proven invaluable in the design of bio-molecular feedback controllers (Grunberg & Del Vecchio, 2019) and the analysis and design of bio-molecular periodic circuits (Cuba Samaniego, Giordano, & Franco, 2020). Moreover, singular perturbation theory has played a vital role in the analysis and design of PID implementations in biochemical networks (Whitby et al., 2022), examining antithetic integral feedback (Zand, Tavazoei, & Kuznetsov, 2022), and for model order reduction in bio-molecular systems (Herath & Vecchio, 2019; Nakakuki & Imura, 2020; Pandey & Murray, 2023) and biochemical networks (Sootla & Anderson, 2017). Furthermore, it has been applied to the long-term regulation of epidemic outbreaks (Al-Radhawi, Sadeghi, & Sontag, 2022; Jaison & Naidu, 2019), and to the study of inhibition and recruitment in thalamocortical networks in the brain (McCreesh & Cortés, 2023).

1.2. Averaging theory

On the other hand, the method of averaging is a cornerstone in modern dynamical systems theory, with roots dating back to the work of Laplace and Lagrange on secular motion in celestial mechanics. A historical account of the development of averaging techniques can be found in the classic book (Sanders, Verhulst, & Murdock, 2007). We also mention the classic survey (Volosov, 1962). The core idea of averaging is very natural; the behavior of a nonautonomous dynamical system can be approximated by the behavior of an autonomous dynamical system in which the time-varying behavior is replaced with its "average" effect. Nevertheless, this simple idea is far-reaching and has a wide range of applications across scientific and engineering disciplines. Naturally, control systems analysis and design has substantially benefited from the methodology of averaging. From stability and robustness analysis to nonlinear and adaptive control design, techniques based on averaging have been widely utilized to solve challenging control problems.

A key result in the theory of averaging is that, under suitable regularity conditions on the dynamics, if the averaged system possesses a locally exponentially stable equilibrium point, then the original time-varying system possesses a locally exponentially stable attractor, a (quasi-)periodic orbit in the (quasi-)periodic case. In the traditional averaging literature, the assumption of exponential stability (or more generally, normal hyperbolicity) has been a necessity. Nevertheless, relatively recent results in the control community, based on advanced Lyapunov techniques, showed that it is possible to relax the assumption of exponential stability of the averaged system to only asymptotic stability (Teel, Peuteman, & Aeyels, 1999), with the price of a weaker notion of stability for the original time-varying system, i.e., semi-global practical stability. Unlike the classical averaging literature, the focus in Teel, Moreau, and Nešić (2003) and Teel et al. (1999) is not on characterizing the nature of the attractor near the equilibrium of the averaged system. In particular, practical stability implies the existence of a compact attracting neighborhood of the origin whose size can be made arbitrarily small, and the basin of its attraction arbitrarily large, by choosing a sufficiently small value for the perturbation parameter of the problem. However, without further conditions on the dynamics, nothing is asserted about the structure of this attracting compact set. A key challenge to note here is that, when normal hyperbolicity fails, the main tool that can be used for establishing the existence and uniqueness of attractors, i.e., the implicit function theorem, is not applicable. However, from a practical point of view, the nature of the attractor is irrelevant as long as it is guaranteed to be bounded by an arbitrarily small upper bound. Hence, the power of the results in Teel et al. (1999). Moreover, it was recognized that the result in Teel et al. (1999) can be characterized in terms of a *closeness-of-trajectories* property (Moreau & Aeyels, 2000; Teel & Nešić, 2000). When two systems satisfy the conditions for the closeness-of-trajectories property, the (uniform) stability properties of one of the systems can be inherited by the other in a practical sense, attaining practical stability. This observation exploits structural robustness results for ODEs and stability bounds in terms of $\mathcal{KL}$ functions (Teel & Praly, 2000). It is an interesting question as to how the closeness-of-trajectories property compares to the more classical "shadowing" property in dynamical systems (Sanders et al., 2007). For instance, under stronger conditions on the dynamics, the latter allows for infinite intervals of validity of solution approximations, even when asymptotic stability does not hold.

The development of practical stability tools in averaging and singular perturbation theory opened the door for various theoretical advancements in the control community. In particular, results from the geometric control literature (Bullo, 2002; Liu, 1997a, 1997b; Sussmann & Liu, 1991) became particularly useful for establishing the closeness-of-trajectories property between a time-varying system and a suitably averaged system. Henceforth, practical stability became a staple in the stability theory of time-varying nonlinear systems. In addition, inspired by the notion of practical stability, various generalizations of control-oriented stability notions emerged, e.g. practical input-to-state stability (ISS)-like properties (Moreau & Aeyels, 2000; Nešić & Teel, 1999; Suttner & Dashkovskiy, 2022), singular practical stability (Dürr, Krstić, Scheinker, & Ebenbauer, 2015), partial practical stability (Grushkovskaya & Zuyev, 2019), fixed-time practical stability (Poveda & Krstić, 2021), etc.

1.3. Applications to control and optimization

The potential of employing high-frequency, open-loop, or feedback-based, oscillatory inputs to induce stability around otherwise unstable equilibria has been long recognized. In fact, a large class of smooth systems cannot be stabilized via time-invariant smooth feedback, as shown in Brockett's seminal paper (Brockett et al., 1983), but can be stabilized via smooth time-varying feedback (Coron, 1995). We distinguish between strictly open-loop stabilization and feedback-based stabilization using oscillatory inputs. The classical example of the former, also known as *vibrational stabilization*, is the Kapitza's pendulum (Meerkov, 1977). In this type of time-varying control, strictly open-loop high-amplitude high-frequency input signals can induce stability, without any feedback. This phenomenon was the subject of significant interest in the control community during the 20th century (Bellman, Bentsman, & Meerkov, 1986a, 1986b; Bentsman, 1987; Meerkov, 1977; Shujaee & Lehman, 1997). On the other hand, feedback-based highly oscillatory control inputs have been extensively studied in the context of non-holonomic driftless control-affine systems (Morin, Pomet, & Samson, 1999), and for the practical stabilization of control-affine systems with drift (Moreau & Aeyels, 2000), see also Bullo (2002) and Vela and Burdick (2003).

One particular control paradigm has substantially benefited from a combination of singular perturbation and averaging: adaptive and learning systems (Amelina, Granichin, & Fradkov, 2019; Anderson et al., 1986; Ariyur & Krstić, 2003; Astrom & Wittenmark, 1989; Benosman, 2016; Borkar, 2009; Kosut, Anderson, & Mareels, 1987; Krilašević, 2023; Sastry & Bodson, 1989; Scheinker & Krstić, 2017). In fact, traditional and two-time scale averaging theory have been used to establish

stability properties in identifiers and parameter estimation dynamics under suitable persistence of excitation conditions (Sastry & Bodson, 1989, Ch. 4). Averaging theory and singular perturbation theory have also been used in the analysis of extremum-seeking systems, a subclass of adaptive systems invented in the early 1920's (Leblanc, 1922). This technique aims to optimize the steady-state input-to-output map of a dynamical system with unknown model, using only output measurements of the plant. The first stability proof of extremum-seeking control for general nonlinear plants was possible largely due to this combination (Krstić & Wang, 2000). Since then, a variety of algorithms, schemes, and methodologies have been developed for the solution of model-free optimization problems in ODEs (Dürr, Stanković, Ebenbauer, & Johansson, 2013; Grushkovskaya, Zuyev, & Ebenbauer, 2018; Guay, Vandermeulen, Dougherty, & McLellan, 2015; Moase & Manzie, 2012; Nešić, Tan, Manzie, Mohammadi, & Moase, 2012; Suttner & Dashkovskiy, 2017; Suttner & Krstić, 2022; Tan, Nešić, & Mareels, 2006; Zhu & Fridman, 2022), PDEs and systems with delays (Feiling, Koga, Krstić, & Oliveira, 2018; Oliveira & Krstić, 2022; Oliveira, Krstić, & Tsubakino, 2016; Tsubakino, Oliveira, & Krstić, 2023; Yu, Koga, & Oliveira, 2021), Nash equilibrium-seeking problems in games (Frihauf, Krstić, & Basar, 2012; Krilašević & Grammatico, 2023; Poveda & Quijano, 2015; Ye, Han, Ding, & Xu, 2023), and hybrid systems (Krilašević, 2023; Kutadinata, Moase, Manzie, Zhang, & Garoni, 2016; Poveda, 2023; Poveda, Benosman, Teel, & Sanfelice, 2021; Poveda et al., 2018; Poveda & Li, 2021; Poveda & Teel, 2017a), to name just a few. As shown in Scheinker and Krstić (2012), inspired by principles of vibrational control, extremum-seeking algorithms can also be used for the solution of model-free *stabilization* problems in settings where a control-like Lyapunov function is known a priori for the system, see Scheinker and Krstić (2017) for a monograph on this subject. When the model of the plant is known and exploration/adaptation is not needed, extremum-seeking techniques reduce to steady-state optimization-based controllers, which can be studied via singular perturbation theory (Bianchin, Poveda and Dall'Anese, 2022; Colombino, Dall'Anese, & Bernstein, 2020; Hauswirth, Bolognani, Hug, & Dorfler, 2020). Such techniques have found important applications in transportation (Bianchin, Cortes, Poveda and DallAnese, 2022) and power systems literature (Ortmann, Maeght, Panciatici, Dörfler, & Bolognani, 2022).

While singular perturbation and averaging theory were initially developed for dynamical systems modeled as ordinary differential equations, they have been recently extended to a class of hybrid dynamical systems (HDS) that incorporate continuous-time dynamics and discrete-time dynamics (Goebel, Sanfelice, & Teel, 2012). Hybrid dynamical systems are ubiquitous in modern engineering and control systems, and they enable the systematic incorporation of logic states, timers, clocks, resets, and other non-smooth phenomena in closed-loop systems (Cassandras & Lygeros, 2010; Goebel, Sanfelice, & Teel, 2009; Sanfelice, 2021). Hybrid systems also emerge naturally in mechanical systems, such as robotic systems, where mechanical contacts are usually modeled as "jumps" (Westervelt, Grizzle, & Koditschek, 2003). Hybrid controllers that incorporate logic states have been shown to overcome some of the fundamental limitations that emerge when using smooth feedback controllers. Applications include the robust global stabilization of a point on smooth compact manifolds and other topologically-obstructed systems (Mayhew, 2010; Prieur, Goebel, & Teel, 2007; Sanfelice, 2021), the robust global solution of synchronization problems (Poveda & Teel, 2019), the improvement of transient performance of nonlinear systems via reset control (Prieur, Queinnec, Tarbouriech, Zaccarian, et al., 2018) and event-triggered control (Poveda & Teel, 2017b), and more recently the development of hybrid optimization algorithms that implement restarting and/or switching policies (Baradaran, Poveda, & Teel, 2018; Poveda & Li, 2021; Teel, Poveda, & Le, 2019). Averaging tools for hybrid systems with fast-varying states acting in the continuous-time dynamics were also introduced in Teel and Nešić (2010), and later generalized in Wang, Teel, and Nešić (2012a) using singular perturbation tools. To a significant extent, these developments were made possible by the robustness results developed in Goebel et al. (2012), which enabled the extension of the closeness-of-trajectories property (in a graphical sense) from solutions of ODEs to solutions of hybrid systems, which are often non-smooth or discontinuous. For a class of linear hybrid systems, singular perturbation results in the spirit of Kokotović, Khalil, and O'Reilly (1986) are also presented in Chitour, Haidar, Mason, and Sigalotti (2023) and Rejeb, Morărescu, Girard, and Daafouz (2018). Singularly perturbed switched systems are also studied in Yang, Wang, Wen, and Daafouz (2020). This diversity of results has opened the door to novel applications in the context of distributed decision-making problems (Wang, Teel, Sun, Liu, & Shao, 2023), dynamic pricing (Poveda, Brown, Marden, & Teel, 2017), network control (Heijmans, Nešić, Postoyan, & Heemels, 2018), and hybrid extremum-seeking control (Poveda & Teel, 2017a).

1.4. Contributions

Motivated by recent applications in chemotactic navigation of sperm cells (Abdelgalil, Aboelkassem, & Taha, 2022) and in extremum-seeking control and source-seeking for nonholonomic vehicles (Abdelgalil, Eldesoukey, & Taha, 2023), in this paper we introduce a second-order two-time scale averaging result for a class of singularly perturbed highly oscillatory systems for which existing techniques in the literature are either not applicable or uninformative. We also highlight a connection between higher-order averaging techniques based on near-identity transformations (Sanders et al., 2007), and averaging based on iterated Lie brackets (Dürr et al., 2013; Liu, 1997a). This connection seems to have been absent from the literature (see the discussion in Dürr et al., 2013). In particular, we illustrate through a simple example that a substantial gap between the two approaches exists, despite the equivalence in certain special cases. In addition to studying higher-order averaging and singularly perturbation methods in continuous-time systems, in this paper, we also explore multi-time scale hybrid dynamical systems, encompassing both "fast" and "slow" states in the continuous-time dynamics. In particular, we study a class of sufficient Lyapunov-based conditions to ensure global asymptotic stability in singularly perturbed hybrid inclusions. This result stands in contrast to the semi-global practical results found in existing literature (Sanfelice & Teel, 2011), which are usually established under weaker assumptions than those considered here. In particular, by considering non-smooth Lyapunov-like conditions on the reduced and boundary layer dynamics, we extend to hybrid systems the well-known *composite Lyapunov method*, a widely used technique in the nonlinear control systems literature (Khalil, 2002, Ch. 11), (Narang-Siddarth & Valasek, 2014). Subsequently, we also introduce a global stability result for hybrid systems based on averaging theory. This result mirrors the averaging theorems found in the adaptive control literature of continuous-time systems (Sastry & Bodson, 1989, Thm. 4.2.5), and, unlike existing results in the literature of hybrid systems (Wang et al., 2012a), is of global nature (naturally, under stronger assumptions). The application of this result is demonstrated through various examples, including parameter estimation problems (Chowdhary & Johnson, 2010; Le & Teel, 2022; Ochoa, Poveda, Subbaraman, Schmidt, & Pour-Safaei, 2021), switching systems (Hespanha & Morse, 1999; Liberzon, 2003), and sampled-data systems (Goebel et al., 2012). Finally, and inspired by the previous discussion, we present several new algorithms and schemes for hybrid extremum-seeking (ES) problems that have not been previously studied, including ES algorithms applied to source-seeking problems under intermittent feedback and spoofing, ES algorithms that incorporate hybrid filters for improved transient performance, and switched Newton-Gradient-like ES algorithms that adaptively switch between Newton-like ES (Ghaffari, Krstić, & Nešić, 2012) and Gradient-like ES (Ariyur & Krstić, 2003) to facilitate the semi-global implementation of the method in applications where the

Hessian estimation is highly susceptible to noise and disturbances far away from the extremum. Furthermore, we address the problem of *stabilizing* a vehicle in a predefined target area in the presence of an obstacle, and under *unknown control directions*. This problem, distinct from those studied in the literature of hybrid control (where the control direction is known a priori) (Sanfelice, 2021; Sanfelice, Messina, Tuna, & Teel, 2006) and hybrid extremum-seeking (Poveda et al., 2021), can be resolved through *hybrid vibrational* control and averaging theory for hybrid systems, paralleling the results of Scheinker and Krstić (2012) for ODEs.

The paper aims to introduce the reader to a diverse set of recent tools for analyzing dynamical systems with multiple time scales in both continuous and hybrid time domains, while also providing guidelines for future research in this field.

1.5. Organization

The rest of this paper is organized as follows: In Section 2, we introduce our notation, and some preliminaries on hybrid dynamical systems, which subsume continuous-time dynamical systems (e.g., ODEs) as a special case. In Sections 3 and 4 we discuss higher-order averaging, and singular perturbations, respectively, in ODEs. We also study higher-order singularly perturbed averaging methods that incorporate tradeoffs. Next, Section 5 studies singular perturbation techniques in hybrid dynamical systems, Section 6 covers averaging methods in such systems, Section 7 presents the proofs, and finally Section 8 ends with the conclusions.

2. Preliminaries

In this section, we introduce the notation that will be used throughout the paper, as well as some preliminaries on continuous-time and hybrid dynamical systems.

2.1. Notation

The set of (nonnegative) real numbers is denoted as $(\mathbb{R}_{\geq 0})\ \mathbb{R}$. The set of (nonnegative) integers is denoted as $(\mathbb{Z}_{\geq 0})\ \mathbb{Z}$. Given a closed set $\mathcal{A} \subset \mathbb{R}^n$, and a column vector $x \in \mathbb{R}^n$, we define $|x|_{\mathcal{A}} := \inf_{y\in\mathcal{A}} |x-y|$. If $x_i \in \mathbb{R}^{n_i}$, for $i \in \{1,\dots,k\}$, are vectors, we use $(x_1,\dots,x_k) \in \mathbb{R}^{n_1+\cdots+n_k}$ to denote their concatenation. We use $\mathbb{S}^1 \subset \mathbb{R}^2$ to denote the unit circle centered at the origin, $\mathbb{B}$ to denote a closed unit ball of appropriate dimension, $\rho\mathbb{B}$ to denote a closed ball of radius $\rho > 0$, and $\mathcal{X} + \rho\mathbb{B}$ to denote the union of all sets obtained by taking a closed ball of radius ρ around each point in the set $\mathcal{X}$. A map $f : \mathbb{R}^m \to \mathbb{R}^n$ is said to be C^k, for $k \in \mathbb{N}_{\geq 0}$, if it is k-times continuously differentiable with locally Lipschitz derivatives. We use Jf to denote the Jacobian of a continuously differentiable function $f : \mathbb{R}^m \to \mathbb{R}^n$. When $n = 1$, we use ∇f to denote the gradient of f. When $\mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k} \ni (x_1,\dots,x_k) \to f(x_1,\dots,x_k) \in \mathbb{R}^n$ is a continuously differentiable map, we use $J_i f$ to denote the Jacobian of f with respect to the ith argument, and $J_{x_i} f$ to denote the Jacobian with respect to the argument x_i, for $i \in \{1,\dots,k\}$. If f is locally Lipschitz, then the set of points where the gradient ∇f is not defined, denoted $\mathcal{Z}$, is of measure zero (Rockafellar & Wets, 1998, pp. 403). In this case, the Clarke generalized gradient of f at $x \in \mathrm{dom}(f)$ is given by $\partial f(x) := \mathrm{co}\{v \in \mathbb{R}^m : \exists\ x^k \to x, x^k \notin \mathcal{Z}, \lim_{k\to\infty} \nabla f(x^k) = v\}$. For $i \in \{1,2,\dots,m\}$, we use $\partial_{x_i} f(x)$ to denote the partial Clarke gradient of f with respect to the component x_i. The function f is said to be regular at x if, for every $u \in \mathbb{R}^n$, the directional derivative $f'(x;u) := \lim_{s\to 0^+} \frac{f(x+su)-f(x)}{s}$ exists, and $f'(x;u) = \max\{v^\top u : v \in \partial f(x)\}$, for all $u \in \mathbb{R}^n$. Typical examples of locally Lipschitz regular functions include continuously differentiable and convex functions. The Lie bracket between two continuously differentiable maps f_1 and f_2 is defined as $[f_1,f_2] := Jf_2 \cdot f_1 - Jf_1 \cdot f_2$. A set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is outer semi-continuous (OSC) at $x \in \mathbb{R}^m$ if for all sequences $x_i \to x$ and $y_i \in M(x_i)$ such that $y_i \to y$ we have that $y \in M(x)$. A set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is locally bounded (LB) at $x \in \mathbb{R}^m$ if there exists a neighborhood U_x of x such that $M(U_x) \subset \mathbb{R}^n$ is bounded. Given a set $\mathcal{X} \subset \mathbb{R}^m$ the mapping M is said to be OSC and LB relative to $\mathcal{X}$ if the set-valued mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$ defined by $M(x)$ for $x \in \mathcal{X}$ and $\emptyset$ for $x \notin \mathcal{X}$ is OSC and LB at each $x \in \mathcal{X}$. We use $\overline{\mathrm{co}}\ \mathcal{X}$ to denote the closed convex hull of $\mathcal{X}$, $\overline{\mathcal{X}}$ to denote the closure of $\mathcal{X}$, and $\mathrm{int}(\mathcal{X})$ to denote its interior. A function $\sigma_L : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{L}$, i.e., $\sigma_L \in \mathcal{L}$, if: (i) it is continuous, (ii) decreasing, and (iii) converging to zero as its argument grows unbounded. A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}$, i.e., $\alpha \in \mathcal{K}$, if: (i) it is continuous, (ii) zero at zero, and (iii) strictly increasing. A function $\tilde{\alpha} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}_\infty$, i.e., $\tilde{\alpha} \in \mathcal{K}_\infty$, if $\tilde{\alpha} \in \mathcal{K}$ and $\tilde{\alpha}$ grows unbounded as its argument grows unbounded. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to be of class $\mathcal{KL}$, i.e., $\beta \in \mathcal{KL}$ if: (i) it is of class $\mathcal{K}$ in its first argument; (ii) it is of class $\mathcal{L}$ in its second argument. Given a compact set $\mathcal{A} \subset \mathbb{R}^n$, a function $\gamma : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is said to be positive semi-definite with respect to $\mathcal{A}$ if $\gamma(\mathcal{A}) = 0$ and $\gamma(x) \geq 0$ for all $x \in \mathbb{R}^n \setminus \mathcal{A}$, and we write $\gamma \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$. If, additionally, $\gamma(x) > 0$ for all $x \in \mathbb{R}^n \setminus \mathcal{A}$, then we say that γ is positive definite with respect to $\mathcal{A}$, and we write $\gamma \in \mathcal{PD}(\mathcal{A})$. For the case when $\mathcal{A} = 0$, we simply write $\gamma \in \mathcal{P}_s\mathcal{D}$ and $\gamma \in \mathcal{PD}$. A function $\delta : (0,\infty) \to (0,\infty)$ is called and order function if there exists $\varepsilon_0 \in (0,\infty)$ such that δ is continuous and positive in $(0,\varepsilon_0]$ and the limit $\lim_{\varepsilon\to 0^+} \delta(\varepsilon)$ exists. We say that a map $\varphi : \mathbb{R}^n \times (0,\infty) \to \mathbb{R}^m$ is $O(\delta(\varepsilon))$ on $\mathcal{D} \subset \mathbb{R}^n$ for some order function $\delta(\varepsilon)$ if there exists positive constants c and ε_0 such that $|\varphi(x,\varepsilon)| \leq c\,\delta(\varepsilon)$, for all $(x,\varepsilon) \in \mathcal{D} \times (0,\varepsilon_0]$.

2.2. Continuous-time and hybrid dynamical systems

We consider finite-dimensional dynamical systems modeled as ordinary differential equations (ODE) or, more generally, as hybrid dynamical systems (HDS) (Goebel et al., 2012). Such systems can be written as

$$x \in C, \qquad \dot{x} \in F(x), \tag{2.1a}$$

$$x \in D, \qquad x^+ \in G(x), \tag{2.1b}$$

where $x \in \mathbb{R}^n$ is the main state of the system. In (2.1), the set-valued mappings $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are called the flow map and the jump map, respectively, and they describe the evolution of the system when x belongs to the flow set C, or the jump set D, respectively. System (2.1) is represented by the notation $\mathcal{H} := \{C,F,D,G\}$, where C, F, D and G comprise the *data* of $\mathcal{H}$. Solutions x to (2.1) are parameterized by a continuous-time index $t \in \mathbb{R}_{\geq 0}$, which increases continuously during the flows (2.1a), and a discrete-time index j, which increases by one during the jumps (2.1b). As such, the notation $\dot{x}$ in (2.1a) stands for $\dot{x} = \frac{\mathrm{d}x(t,j)}{\mathrm{d}t}$, and the notation x^+ in (2.1b) stands for $x^+ = x(t,j+1)$. When $D = \emptyset$, system (2.1) recovers a continuous-time system, and in this case the index j can be omitted from the solutions. Additionally, when F is singled-valued and continuous, system (2.1) reduces to a standard ODE. In this way, system (2.1) provides a unifying formalism to study both continuous-time systems and hybrid dynamical systems. For the case when (2.1) depends on exogenous time-varying signals, we will also consider systems of the form

$$(x,\tau) \in C \times \mathbb{R}_{\geq 0}, \qquad \dot{x} \in F(x,\tau), \qquad \dot{\tau} = \rho, \tag{2.2a}$$

$$(x,\tau) \in D \times \mathbb{R}_{\geq 0}, \qquad x^+ \in G(x), \qquad \tau^+ = \tau, \tag{2.2b}$$

which are common in the averaging literature of hybrid systems (Teel & Nešić, 2010), and where $\rho > 0$ dictates the rate of evolution of the auxiliary state τ. In some cases, we will also use a different continuous-time scale (usually, denoted by s) to study the behaviors of systems (2.1) and (2.2) under fast-varying signals or dynamics.

2.3. Solutions and stability concepts for HDS

Since the solutions to system (2.1) are parameterized by both continuous-time and discrete-time indexes, they are defined on *hybrid time domains*. A set $E \subset \mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ is called a *compact* hybrid time domain if $E = \cup_{j=0}^{J-1}([t_j, t_{j+1}], j)$ for some finite sequence of times $0 = t_0 \leq t_1 \cdots \leq t_J$. The set E is a hybrid time domain if for all $(T, J) \in E$, $E \cap ([0, T] \times \{0, \ldots, J\})$ is a compact hybrid time domain. Again, if $D = \emptyset$, then the index j can be omitted and in this case the solutions to system (2.1) are defined on intervals of $\mathbb{R}_{\geq 0}$.

Definition 2.1. A function $x : \mathrm{dom}(x) \mapsto \mathbb{R}^n$ is a hybrid arc if $\mathrm{dom}(x)$ is a hybrid time domain and $t \mapsto x(t, j)$ is locally absolutely continuous for each j such that the interval $I_j := \{t : (t, j) \in \mathrm{dom}(x)\}$ has nonempty interior. A hybrid arc x is a *solution* to (2.1) if $x(0, 0) \in \overline{C} \cup D$, and the following two conditions hold: (a) For each $j \in \mathbb{Z}_{\geq 0}$ such that I_j has nonempty interior: $x(t, j) \in C$ for all $t \in \mathrm{int}(I_j)$, and $\dot{x}(t, j) \in F(x(t, j))$ for almost all $t \in I_j$; (b) For each $(t, j) \in \mathrm{dom}(x)$ such that $(t, j + 1) \in \mathrm{dom}(x)$: $x(t, j) \in D$, and $x(t, j + 1) \in G(x(t, j))$. □

A solution x to system (2.1) is said to be forward pre-complete if its domain is compact or unbounded. It is said to be forward complete if its domain is unbounded, and it is said to be maximal if there does not exist another solution ψ to system (2.1) such that $\mathrm{dom}(x)$ is a proper subset of $\mathrm{dom}(\psi)$, and $x(t, j) = \psi(t, j)$ for all $(t, j) \in \mathrm{dom}(x)$.

In general, given an initial condition $x(0, 0) \in \mathbb{R}^n$, solutions to system (2.1) might not be unique. This will be the case if, for example, $C \cap D$ is not empty, or if the flow or jump maps in (2.1a)–(2.1b) admit non-unique solutions. Therefore, in the stability analysis of systems of the form (2.1) (including non-Lipschitz ODEs), we will generally insist that *every* solution satisfies suitable boundedness and convergence bounds expressed in terms of $\mathcal{KL}$ functions, as formalized in the following definition.

Definition 2.2. Let $\mathcal{A} \subset \mathbb{R}^n$ be compact. The set $\mathcal{A}$ is said to be *uniformly globally asymptotically stable* (UGAS) for system (2.1) (or (2.2)) if there exists a $\mathcal{KL}$ function β such that any maximal solution to $\mathcal{H}$ satisfies

$$|x(t, j)|_{\mathcal{A}} \leq \beta(|x(0, 0)|_{\mathcal{A}}, t + j), \tag{2.3}$$

for all $(t, j) \in \mathrm{dom}(x)$. If there exist $c_1, c_2 > 0$ such that $\beta(r, s) = c_1 r e^{-c_2 s}$, then we say that $\mathcal{A}$ is *uniformly globally exponentially stable* (UGES). □

In this paper, we will make the following standing assumption on the data of the systems.

Assumption 2.1 (*The Hybrid Basic Conditions*)**.** The sets C and D are closed. The set-valued mapping F is OSC and LB relative to C, and $C \subset \mathrm{dom}(F)$. Moreover, for every $x \in C$ the set $F(x)$ is convex. The set-valued mapping G is OSC and LB relative to D, and $D \subset \mathrm{dom}(G)$. □

In the next two sections, we will first consider systems of the form (2.1) and (2.2) with $D = \emptyset$ and F being a continuous function. Afterwards, in Sections 5–6 we will consider general hybrid dynamical systems where $D \neq \emptyset$.

3. Higher-order averaging in ODEs

First-order averaging techniques, also referred to as standard averaging, have been thoroughly studied in the control literature (Khalil, 2002, Chapter 10) and have found extensive application in various domains, such as adaptive control (Sastry & Bodson, 1989), pulse-width modulation-based control (Wang, Teel, & Nešić, 2012b), vibrational control (Khalil, 2002), and extremum-seeking control (Krstić & Wang, 2000). On the other hand, *higher-order* averaging techniques are less widely used, but they emerge in the context of geometric control (Bullo & Lewis, 2004; Liu, 1997a; Vela & Burdick, 2003) and controllability analysis (Sussmann & Liu, 1991). Various methodologies and approaches to higher-order averaging have been studied in the literature, including: (i) averaging based on *near-identity* (or *Lie-*) *transforms* (Sanders et al., 2007; Volosov, 1962); (ii) averaging based on the *chronological calculus* (Bullo, 2001, 2002; Sarychev, 2001); and averaging based on iterated *Lie brackets* (Liu, 1997a, 1997b). In a recent article (Maggia, Eisa, & Taha, 2020), it was argued that the higher-order averaging methods based on chronological calculus and on near-identity transforms are equivalent up to the fourth-order. Other than the discussion in Dürr et al. (2013), it seems that there are limited results on the connections between the averaging method based on iterated Lie brackets (Liu, 1997a, 1997b) and the other two methods. However, averaging based on Iterated Lie Brackets turns out to be connected to a specific form of higher-order averaging known as *averaging with trade-off* (Murdock, 1983).

In this section, we review the concept of higher-order averaging and illustrate the aforementioned connection by focusing our discussion on the second-order and time-periodic cases.

3.1. Second-order averaging based on near-identity transforms

Consider the following dynamical system evolving on the s-time scale:

$$\frac{\mathrm{d}x}{\mathrm{d}s} = \sum_{k=1}^{2} \varepsilon^k f_k(x, \tau), \qquad \frac{\mathrm{d}\tau}{\mathrm{d}s} = 1, \tag{3.1}$$

where $(x, \tau) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}$, $\varepsilon \in (0, \varepsilon_0]$ is a parameter, and where the functions f_k are C^{3-k} in the first argument, and continuous and T-periodic in the second argument, i.e., there exists $T > 0$ such that:

$$f_k(x, \tau + T) = f_k(x, \tau), \quad \forall\, x \in \mathbb{R}^n. \tag{3.2}$$

Due to periodicity, we can always take the initial condition for τ to be $\tau(0) = 0$, without any loss of generality. In the s-time scale, system (3.1) has the form of (2.2) with $D = \emptyset$, $C = \mathbb{R}^n$ and $\rho = 1$. For each $(\xi, \tau, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} \times [0, \varepsilon_0]$, we define the function

$$W(\xi, \tau, \varepsilon) := \xi + \varepsilon\, w(\xi, \tau), \tag{3.3}$$

where w is given by:

$$w(\xi, \tau) = \int_0^{\tau} \left(f_1(\xi, \nu) - \bar{f}_1(\xi)\right) d\nu, \tag{3.4}$$

$$\bar{f}_1(\xi) = \frac{1}{T} \int_0^{T} f_1(\xi, \nu)\, d\nu. \tag{3.5}$$

Restricted to any bounded domain $D \subset \mathbb{R}^n$, and for ε sufficiently small, the map $W(\cdot, \tau, \varepsilon)$ is a well-defined diffeomorphism, for all $\tau \in \mathbb{R}_{\geq 0}$ (Sanders et al., 2007, Lemma 2.8.3). For $\varepsilon = 0$, the map W reduces to the identity map. Hence, we refer to W as a near-identity transformation.

To study system (3.1) using the near-identity transformation W, we introduce the *second-order averaged system*, in the s-time scale, given by:

$$\frac{\mathrm{d}\xi}{\mathrm{d}s} = \sum_{k=1}^{2} \varepsilon^k \bar{f}_k(\xi), \tag{3.6}$$

where the function $\bar{f}_2 : \mathbb{R}^n \to \mathbb{R}^n$ is defined as follows:

$$\bar{f}_2(\xi) = \frac{1}{T} \int_0^{T} \left(f_2(\xi, \nu) + \frac{1}{2}\left[w, f_1 + \bar{f}_1\right](\xi, \nu)\right) d\nu.$$

In this way, system (3.6) is defined in terms of $\bar{f}_1$, and, on average, also in terms of the function f_2 and the Lie brackets between w and $f_1 + \bar{f}_1$. The relationship between the trajectories of system (3.1) and its average system (3.6) is characterized by the following theorem, adapted from Sanders et al. (2007):

Theorem 3.1 (*Sanders et al., 2007*). *Let $K_0 \subset \mathbb{R}^n$ be a compact set. Suppose that for each $\xi_0 \in K_0$ and $\varepsilon \in (0, \varepsilon_0]$ system (3.6) has a unique and uniformly bounded solution ξ starting at the initial condition $\xi(0) = \xi_0$. Then, for each $T_f \in \mathbb{R}_{>0}$ there exists $\varepsilon^* \in (0, \varepsilon_0]$ such that for all $\varepsilon \in (0, \varepsilon^*)$ and all $\xi_0 \in K_0$ system (3.1) has a unique solution (x, τ) satisfying*

$$|x(s) - \bar{x}(s;\varepsilon)| = O(\varepsilon^2), \qquad \forall\ s \in \left[0, \frac{T_f}{\varepsilon}\right], \tag{3.7}$$

for $(x(0), \tau(0)) = (\xi_0, 0)$, where $\bar{x}(s;\varepsilon) := W(\xi(s), \tau(s), \varepsilon)$. □

A crucial difference between higher-order periodic averaging and first-order averaging is that in the latter case we may directly compare the trajectories of the averaged system (3.6) to those of the original system (3.1) and obtain an $O(\varepsilon)$-estimate on the error (Sanders et al., 2007, Theorem 2.8.1). However, in the second-order case the near-identity transform W must be used to obtain the $O(\varepsilon^2)$-estimate on the error. Nevertheless, a special situation arises when $\bar{f}_1 = 0$. In this case, it is possible to sacrifice part of the accuracy of the error estimate, which is $O(\varepsilon^2)$, in exchange for a longer interval of validity. This situation, known as *averaging with trade-off*, is captured by the following theorem, which is adapted from Sanders et al. (2007, Section 2.9) and Murdock (1983):

Theorem 3.2 (*Murdock, 1983; Sanders et al., 2007*). *Let $K_0 \subset \mathbb{R}^n$ be a compact set, and suppose that $\bar{f}_1 = 0$. Moreover, suppose that for all $\varepsilon \in (0, \varepsilon_0]$ and all $\xi_0 \in K_0$ system (3.6) has a unique and uniformly bounded solution ξ starting at the initial condition $\xi(0) = \xi_0$. Then for each $T_f \in \mathbb{R}_{>0}$ there exists $\varepsilon^* \in (0, \varepsilon_0]$ such that for all $\varepsilon \in (0, \varepsilon^*)$ and all $\xi_0 \in K_0$ there exists a unique solution (x, τ) to (3.1) satisfying*

$$|x(s) - \xi(s)| = O(\varepsilon), \qquad \forall\ s \in \left[0, \frac{T_f}{\varepsilon^2}\right], \tag{3.8}$$

for $(x(0), \tau(0)) = (\xi_0, 0)$. □

In Theorem 3.2, the near-identity transform W does not appear in (3.8). The reason is that, in this special case, the deviation of the near-identity transform W from the identity has the same asymptotic order as the error term, i.e., $O(\varepsilon)$. Therefore, the near-identity transform does not contribute to the asymptotic order of the estimate and it can be replaced by the identity map. Moreover, since $\bar{f}_1 = 0$, in this case the averaged system in the s-time scale becomes

$$\frac{\mathrm{d}\xi}{\mathrm{d}s} = \varepsilon^2 \bar{f}_2(\xi), \tag{3.9}$$

which is equivalent, after the time-scale change $t = \varepsilon^{-2} s$, to the following system in the t-time scale:

$$\dot{\xi} = \bar{f}_2(\xi), \tag{3.10}$$

where $\bar{f}_2$ is now given by:

$$\bar{f}_2(\xi) = \frac{1}{T}\int_0^T \left(f_2(\xi, \tau) + \frac{1}{2}\left[w, f_1\right](\xi, \tau)\right) d\nu. \tag{3.11}$$

with

$$w(\xi, \tau) = \int_0^\tau f_1(\xi, \nu)\, d\tau. \tag{3.12}$$

Later, we will show that averaging with trade-off encompasses the approach of averaging based on iterated Lie brackets (Dürr et al., 2013; Liu, 1997a) under a specific assumption about the structure of the vector fields f_1 and f_2.

3.2. Second-order averaging based on iterated Lie brackets

Let $r \in \mathbb{N}_{\geq 2}$ and consider the time-varying control-affine system:

$$\dot{x} = b_0(x) + \sum_{i=1}^{r} b_i(x)\, \omega^{\frac{1}{2}} u_i(\tau), \qquad \dot{\tau} = \omega, \tag{3.13}$$

where $(x, \tau) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}$, and $\omega \in \mathbb{R}_{>0}$ is a parameter. This system has the form of (2.2) with $C = \mathbb{R}^n$, $D = \emptyset$, and $\rho = \omega$. We assume that $b_0 \in C^1$, $b_i \in C^2$, and the functions $u_i(\cdot)$ are continuous and T-periodic. Moreover, the functions $u_i(\cdot)$ are such that:

$$0 = \int_0^T u_i(\tau)\, d\tau, \quad v_{ji} = \frac{1}{T}\int_0^T u_j(\tau) \int_0^\tau u_i(\nu) d\nu\, d\tau.$$

Remark 3.3. Systems of the form (3.13), as discussed in Dürr et al. (2013) in the context of extremum-seeking, represent the second-order version of the class of systems extensively treated in Liu (1997a). In Dürr et al. (2013), the functions u_i can also depend on the slow-time t, but for the sake of clarity, we omit this generalization as it is not crucial for our purposes. Additionally, we note that the regularity assumptions adopted here are stronger than those in Dürr et al. (2013), but they can be easily relaxed. □

We now introduce the autonomous *Lie Bracket Approximation* system for (3.13), with state $\xi \in \mathbb{R}^n$ and dynamics:

$$\dot{\xi} = b_0(\xi) + \sum_{i=1,\, j>i}^{r} [b_i, b_j](\xi)\, v_{ji}. \tag{3.14}$$

The following theorem, adapted from Dürr et al. (2013), characterizes the relationship between the trajectories of the two systems when ω is large enough to induce a time-scale separation between the time-varying periodic vector field (3.13) and the average dynamics.

Theorem 3.4. *Let $K_0 \subset \mathbb{R}^n$ be a compact set. Suppose that for all $\xi_0 \in K_0$ system (3.14) has a unique and uniformly bounded solution ξ starting at the initial condition $\xi(0) = \xi_0$. Then, for all $\epsilon, T_f \in \mathbb{R}_{>0}$ there exists $\omega^* \in \mathbb{R}_{>0}$ such that for all $\omega \geq \omega^*$ and all $\xi_0 \in K_0$, system (3.13) has a unique solution (x, τ) satisfying*

$$|x(t) - \xi(t)| \leq \epsilon, \tag{3.15}$$

for $(x(0), \tau(0)) = (\xi_0, 0)$, and for all $t \in [0, T_f]$. □

We note that the asymptotic order of the error is not explicitly stated in Theorem 3.4, though careful inspection of the proofs in Dürr et al. (2013) reveals that the asymptotic order is in fact $O(1/\sqrt{\omega})$. This subtlety is already a strong hint for the connection to averaging with trade-off.

3.3. A connection between the two approaches

The statement of Theorem 3.4 is nearly identical to the statement of Theorem 3.2. This similarity is not coincidental. In fact, the two theorems are equivalent, provided that an additional assumption is made regarding the structure of the function f_1 in Theorem 3.2. We summarize this observation in the following proposition. The proof is provided in Section 7.1.

Proposition 3.5. *Consider the system (3.1) and suppose that f_1 and f_2 are of the form:*

$$f_1(x, \tau) = \sum_{i=1}^{r} b_i(x)\, u_i(\tau), \quad f_2(x) = b_0(x), \tag{3.16}$$

where the functions $u_i(\cdot)$ are continuous and T-periodic. Then, the averaged map $\bar{f}_2$ given by (3.11) reduces to:

$$\bar{f}_2(\xi) = b_0(\xi) + \sum_{i=1,j>i}^{r} [b_i, b_j](\xi) v_{ji}, \tag{3.17}$$

$$v_{ji} = \frac{1}{T}\int_0^T u_j(\tau) \int_0^\tau u_i(\nu)\, d\nu\, d\tau. \tag{3.18}$$

for all $\xi \in \mathbb{R}^n$. □

We note that the two systems (3.1) and (3.13) are written in different time-scales. That is, the time-scale change $s = \varepsilon^{-2} t$ and the identification $\omega = \varepsilon^{-2}$ transform either system to the other, provided that the assumptions of Proposition 3.5 hold.

Remark 3.6. Higher-order periodic averaging also applies to quasi-periodic vector fields, with the same error estimates, provided that the quasi-periodicity arises from a finite summation of periodic functions. In the general case, the error estimates are necessarily more conservative, see e.g., the discussion in Sanders et al. (2007, Sections 4.5–4.6). Finally, we note that in the context of Proposition 3.5 it is possible to relax the continuity assumptions on u to mere measurability. □

The following example illustrates a gap between second-order averaging based on near-identity transforms with trade-off (c.f. Theorem 3.2) and second-order averaging based on iterated Lie brackets (c.f. Theorem 3.4)

Example 3.7. Consider the system

$$\dot{x} = \varepsilon^{-1} h(\cos(c(x) + \tau)), \qquad \dot{\tau} = \varepsilon^{-2}, \tag{3.19}$$

where $x \in \mathbb{R}$, $c(\cdot)$ is strictly convex with a unique minimizer x^*, and $h(\cdot)$ is a non-zero and odd C^2 function that is not necessarily linear, thus violating the structure required by Theorem 3.2. Moreover, since h is smooth and odd, we have:

$$\int_0^{2\pi} h(\cos(a + \tau))\, d\tau = 0, \qquad \forall\ a \in \mathbb{R}, \tag{3.20}$$

which shows that the standard (first-order) average of (3.19) vanishes. However, we can apply second-order averaging with trade-off to analyze system (3.19). Indeed, using (3.11) we obtain the following average dynamics:

$$\dot{\xi} = -\alpha(c(\xi))\, \nabla c(\xi), \tag{3.21}$$

where the function $\alpha(\cdot)$ is given by

$$\alpha(a) = \frac{1}{2\pi} \int_0^{2\pi} h(\cos(a + \tau))^2\, d\tau. \tag{3.22}$$

Since (3.22) is the integral of a non-negative function that is strictly positive in a subset of the domain of integration, it follows that $\alpha(a) > 0$ for all a. Moreover, α is 2π-periodic in a, and it can be written as

$$\alpha(a) = \frac{1}{2\pi} \int_a^{2\pi+a} h(\cos(\sigma))^2 d\sigma.$$

Differentiating α using Leibniz's rule:

$$\alpha'(a) = \frac{1}{2\pi} \left(h(\cos(2\pi + a))^2 - h(\cos(a))^2 \right) = 0$$

which shows that α remains constant and $\alpha(a) = \alpha(0)$, for all $a \in \mathbb{R}$. Therefore, the averaged dynamics (3.21) describes a gradient flow, for which the estimate of Theorem 3.2 holds. Fig. 1 shows different trajectories of system (3.19) for the function $c(x) = \frac{1}{2}x^2$ and various choices of the function h, including highly nonlinear ones. For the specific case where $h(u) = u$, and by utilizing special properties of trigonometric functions, it is possible to transform the system into a control-affine form, allowing the application of the results in Dürr et al. (2013). □

Remark 3.8. We remark that the function h in Example 3.7 can be thought of as a control non-linearity. The Lie bracket approximation approach (Dürr et al., 2013; Grushkovskaya et al., 2018; Suttner & Dashkovskiy, 2017) assumes a control-affine structure for the dynamical system under consideration. However, higher-order averaging based on near-identity transforms does not assume a control-affine structure, illustrating its generality. In fact, the function h may reverse sign at several points in its domain as long as it remains an odd function. □

It is worth mentioning that the full potential of higher-order averaging based on near-identity transforms in control theory, including its extension to averaging over angles where the frequency of oscillation may depend on the state (Sanders et al., 2007, Chapter 7), remains largely unexplored.

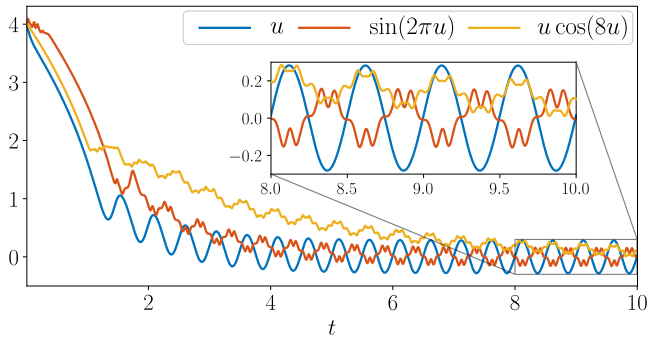

Fig. 1. Trajectories of system (3.19) for $\varepsilon = 1/\sqrt{6\pi}$, $c(x) = x^2$, $x(0) = 5$, and several choices for the nonlinearity $h(u)$, shown in the legend.

Remark 3.9. Theorem 3.2, respectively Theorem 3.4, establishes the *closeness-of-solutions property* in compact sets and compact time domains between system (3.1), respectively (3.13), and system (3.10), respectively (3.14). A consequence of this fact is that the original dynamics (3.1), respectively (3.13), can inherit uniform stability properties (characterized by, e.g., $\mathcal{KL}$ bounds) from the average dynamics (3.10), respectively (3.14), in a practical sense (Moreau & Aeyels, 2000; Teel et al., 2003). We refer the reader to Moreau and Aeyels (2000) and Teel et al. (2003) for more information on practical stability in ODES. □

We conclude this section with a note on the upper bound ε^* in Theorems 3.1 and 3.2 (or, equivalently, the frequency lower bound ω^* in Theorem 3.4). The most general results, i.e. for arbitrary order, in both approaches to averaging are qualitative in nature (Liu, 1997a; Sanders et al., 2007). However, there exist results (Dürr et al., 2013; Grushkovskaya et al., 2018) that provide constructive proofs for the second-order case, and more recently the works (Zhang & Fridman, 2023; Zhu & Fridman, 2022) studied LMIs to obtain quantitative estimates on the frequency lower bound ω^* in the context of Lie Bracket averaging.

4. Higher-order two-time scale averaging in ODEs

In the previous section, we studied a class of systems in which the time-scale separation exists between the full state of the system and fast time-periodic variations. In that case, the method of averaging allowed us to derive a suitably averaged system that captures the average effects of the fast periodic variations. However, when the "fast" dynamics involve not only periodic time-variations but also a portion of the states of the system, a different set of analytical tools needs to be used to study the system. Such tools are referred to in the literature as *two-time scale averaging* (Sastry & Bodson, 1989).

In order to study two-time scale averaging, we consider continuous-time systems of the form

$$\dot{x} = f_x(x, z, \tau), \tag{4.1a}$$

$$\varepsilon\, \dot{z} = f_z(x, z, \tau) \tag{4.1b}$$

$$\varepsilon\, \dot{\tau} = 1, \tag{4.1c}$$

where $x \in \mathbb{R}^{n_1}$ is the "slow" state, $z \in \mathbb{R}^{n_2}$ is the "fast" state, and $\varepsilon \in (0, \infty)$ is a small parameter. To simplify the presentation, we will assume that the vector fields f_x, f_z are sufficiently smooth in (x, z), and T-periodic and continuous with respect to τ. Note that system (4.1) has the form of (2.2) with $C = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, $D = \emptyset$, and $\rho = \frac{1}{\varepsilon}$.

To study system (4.1), we assume the existence of a time-invariant quasi-steady state mapping $\varphi_0 : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$, such that for all $x \in \mathbb{R}^{n_1}$ and all $\tau \in \mathbb{R}_{\geq 0}$:

$$f_z(x, z, \tau) = 0 \iff z = \varphi_0(x). \tag{4.2}$$

By using φ_0, we can define a *reduced system*, which has states $(\tilde{x}, \tau) \in \mathbb{R}^{n_1} \times \mathbb{R}_{\geq 0}$ and dynamics

$$\dot{\tilde{x}} = \tilde{f}_x(\tilde{x}, \tau) := f_x(\tilde{x}, \varphi_0(\tilde{x}), \tau), \quad \varepsilon \dot{\tau} = 1. \tag{4.3}$$

The study of system (4.3) via standard (i.e., first-order) averaging leads to the following averaged reduced-order dynamics with state $\bar{x} \in \mathbb{R}^{n_1}$:

$$\dot{\bar{x}} = \bar{f}_x(\bar{x}), \tag{4.4}$$

where the mapping $\bar{f}_x$ is given by

$$\bar{f}_x(\bar{x}) = \frac{1}{T} \int_0^T \tilde{f}_x(\bar{x}, \tau)\, d\nu,$$

for all $\bar{x} \in \mathbb{R}^{n_1}$. Then, the relationship between the slow part of the trajectories of the original system (4.1) and the reduced-order averaged system (4.4) can be characterized via standard results in two-time scale averaging, as described in Sastry and Bodson (1989, Thm. 4.4.2). Fundamentally, two-time scale averaging combines (first-order) averaging and (first-order) *singular perturbation* techniques.

Substantial generalizations of this classical result have been studied in the literature. For example, Teel (2000) presents a unified framework for averaging, singular perturbation, and two-time scale averaging using Input-to-State Stability (ISS) properties (Sontag, 1989) and relaxing various classical requirements on the dynamics of the fast states. However, most results in the literature on two-time scale averaging are limited to the first-order case, which restricts their applicability to systems in which the first-order approximation does not vanish. Nevertheless, as illustrated in Section 3, there are systems for which the first-order average does vanish, and it is the second-order average that provides valuable information for stability analysis. It turns out that a similar situation arises in two-time scale averaging. We illustrate this fact through the following stylized example:

Example 4.1. Consider the following system:

$$\dot{x} = \beta x + \varepsilon^{-1} \sin(Y + \tau), \tag{4.5a}$$

$$\varepsilon^2 \dot{\tau} = 1 \tag{4.5b}$$

$$\sigma_1 \mu\, \dot{z}_1 = \phi(x) - z_1, \tag{4.5c}$$

$$\sigma_2 \mu\, \dot{z}_2 = z_1 - z_2 \tag{4.5d}$$

$$Y = 2k(z_1 - z_2), \tag{4.5e}$$

where $(x, \tau, z_1, z_2) \in \mathbb{R}^4$, $\phi(x) = x^2$, $\beta \in (0, 1)$, $k, \sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, and μ and ε are small positive parameters. Eq. (4.5a) describes a one-dimensional Lie bracket based extremum-seeking algorithm (Scheinker & Krstić, 2014) expressed in the form of a highly oscillatory system. On the other hand, the last three equations in (4.5) model an LTI system with input $\phi(x)$, output Y, and transfer function given by:

$$G(\hat{s}) = \frac{2k\sigma_1 \mu \hat{s}}{(\sigma_1 \mu \hat{s} + 1)(\sigma_2 \mu \hat{s} + 1)},$$

which corresponds to a band-pass filter with passband $(1/(\max\{\sigma_1, \sigma_2\}\mu), 1/(\min\{\sigma_1, \sigma_2\}\mu))$. An alternative block diagram description of this system is shown in Fig. 2.

Setting $\mu = \varepsilon^2$ leads to the following system in the new time scale $s = \varepsilon^{-1} t$:

$$\frac{dx}{ds} = \sin(2k(z_1 - z_2) + \tau) + \varepsilon \beta x, \tag{4.6a}$$

$$\varepsilon \frac{d\tau}{ds} = 1 \tag{4.6b}$$

$$\varepsilon \frac{dz_1}{ds} = \sigma_1{}^{-1} \left(\phi(x) - z_1\right), \tag{4.6c}$$

$$\varepsilon \frac{dz_2}{ds} = \sigma_2{}^{-1} \left(z_1 - z_2\right), \tag{4.6d}$$

which is on the standard form (4.1) for the application of two-time scale averaging. In this case, the quasi-steady state of the fast dynamics corresponds to $z_1 = z_2 = \phi(x)$. Therefore, by applying the standard

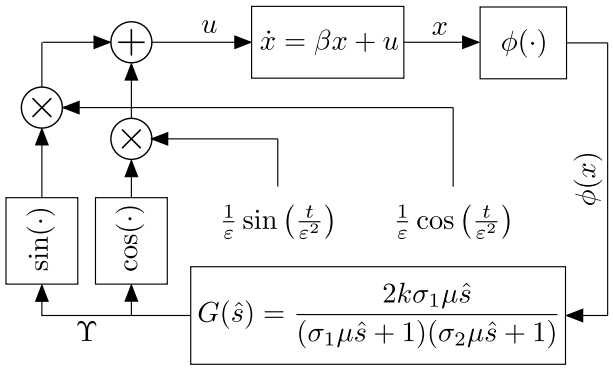

Fig. 2. Block diagram description of system (4.5).

(first-order) two-time scale averaging procedure (Sastry & Bodson, 1989), we obtain the following reduced-order dynamics:

$$\frac{d\bar{x}}{ds} = \sin(\tau), \quad \varepsilon \frac{d\tau}{ds} = 1 \tag{4.7}$$

and also the following reduced order averaged dynamics:

$$\frac{d\bar{x}}{ds} = 0, \tag{4.8}$$

which is only marginally stable. Yet, as shown in Fig. 3, the original system (4.5) exhibits a (practical) asymptotically stable behavior when $\mu = \varepsilon^2$. Indeed, since the first-order approximation vanishes, the stability properties of system (4.5) are actually dictated by higher-order approximations, which are neglected in standard (first-order) two-time scale averaging.

On the other hand, using the trigonometric identity $\sin(x + y) = \sin(x)\cos(y) + \cos(y)\sin(x)$, and setting $\omega = \varepsilon^{-2}$, system (4.5) can be expressed in the same form as the class of singularly perturbed highly oscillatory systems considered in the literature of extremum-seeking control (Dürr et al., 2015; Dürr, Krstić, Scheinker, & Ebenbauer, 2017). In this case, the quasi-steady state of the fast dynamics is still given by $z_1 = z_2 = \phi(x)$. However, using the singularly perturbed Lie bracket approximation results from Dürr et al. (2015, 2017), we obtain the following reduced order system:

$$\dot{\tilde{x}} = \beta \tilde{x} + \sqrt{\omega} \sin(\omega t), \tag{4.9}$$

and the following reduced order average Lie bracket system:

$$\dot{\bar{x}} = \beta \bar{x}, \tag{4.10}$$

which is exponentially unstable. Thus, this approximation also does not capture the stable behavior shown in Fig. 3. Indeed, since the analysis in Dürr et al. (2015, 2017) requires that $\mu = O(\omega^{-\kappa})$ for some $\kappa > 2$, the case $\mu = O(\omega^{-1}) = O(\varepsilon^2)$ strictly violates the sufficient conditions considered in Dürr et al. (2015). Intuitively, when $\mu = O(\omega^{-1})$, there is not enough time-scale separation between the frequency of the sinusoidal signals and the fast dynamics to justify replacing the fast states with a stationary quasi-steady state. Indeed, if $\mu = O(\omega^{-1})$, the frequency of the dither signal ω lies inside the passband of the filter $G(\hat{s})$ when $\omega \gg 1$, and therefore has a significant contribution to the second-order effects which is not accounted for by the framework considered in Dürr et al. (2015, 2017). □

To properly capture the behavior of systems of the form (4.5) when $\mu = \varepsilon^2$, one needs a suitable higher-order extension of two-time scale averaging. Such an extension requires an asymptotic approximation of the fast motion *around the slow manifold*. Indeed, applying second-order averaging to system (4.7) would still result in the uninformative system (4.8). In other words, a higher-order singular perturbation analysis is needed to study (4.5).

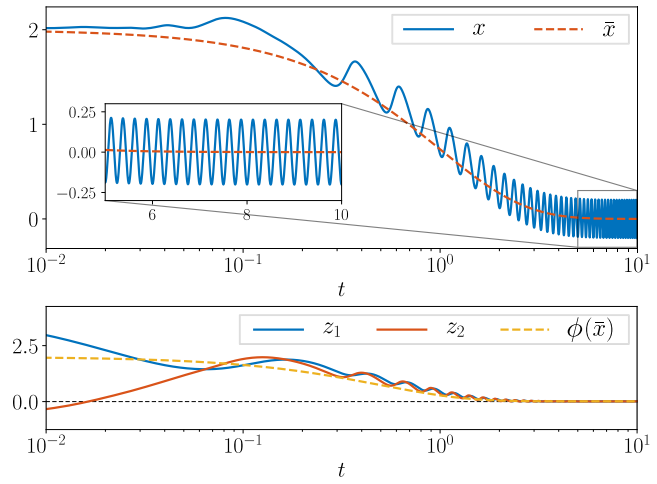

Fig. 3. Trajectories of system (4.5) for $\varepsilon = 1/\sqrt{8\pi}, \mu = \varepsilon^2$, $\beta = 1/2$, $k = 3$, and initial conditions $(x(0), z_1(0), z_2(0)) = (2, 4, -1)$.

4.1. Higher-order singular perturbation

We now provide a brief introduction to higher-order asymptotics in singular perturbations, mostly based on Vasil'Eva, Butuzov, and Kalachev (1995), focusing on the first-order correction. After this, and motivated by applications in extremum-seeking control and source seeking for nonholonomic systems (Abdelgalil et al., 2022, 2023), we state and prove a second-order two-time scale averaging result.

Consider the following sub-class of systems (4.1):

$$\dot{x} = f_x(x, z), \tag{4.11a}$$

$$\varepsilon\, \dot{z} = f_z(x, z), \tag{4.11b}$$

where $(x, z) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ and $\varepsilon > 0$ is a small parameter. As before, we assume the existence of a quasi-steady state mapping $\varphi_0 : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ satisfying

$$f_z(x, z) = 0 \quad \Longleftrightarrow \quad z = \varphi_0(x), \tag{4.12}$$

for all $x \in \mathbb{R}^{n_1}$. The reduced order system in the t-time scale is given by

$$\dot{\tilde{x}} = \tilde{f}_x(\tilde{x}) := f_x(\tilde{x}, \varphi_0(\tilde{x})), \tag{4.13}$$

for all $\tilde{x} \in \mathbb{R}^{n_1}$. Similarly, the *boundary layer system*, in the time-scale $s = \varepsilon^{-1}t$, is defined as:

$$\frac{\mathrm{d}\tilde{z}}{\mathrm{d}s} = f_z(\tilde{x}, \tilde{z}), \tag{4.14}$$

for all $(\tilde{x}, \tilde{z}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ where $\tilde{x}$ is treated as a parameter. The relationship between the trajectories of system (4.11) and systems (4.13)–(4.14), on compact time intervals and compact sets of initial conditions, is given by the classical theorem of Tikhonov (Khalil, 2002, Section 11.2), and can be extended to the infinite interval under appropriate smoothness and stability assumptions (Khalil, 2002, Section 11.2).

On the other hand, the construction of higher-order asymptotics in singular perturbation problems is, in general, substantially more difficult than regular perturbations and typically involves matched asymptotic expansions (Kuehn et al., 2015, Section 9.1) with asymptotic series that are not clear a priori. As a consequence, with rare exceptions (Bernard, Jebai, & Martin, 2020), higher-order singular perturbation has rarely appeared in the controls literature. Nevertheless, when the linearization of the fast vector field f_2 around the quasi-steady state φ_0 is uniformly Hurwitz, the construction of higher-order asymptotics reduces to a regular perturbation problem. Specifically, the uniform exponential attractivity property of the boundary layer dynamics clarifies the appropriate fast-time scale and the structure of the composite expansion to be sought in the perturbation analysis. By exploiting this additional structure, the so-called Boundary Function method (Vasil'Eva et al., 1995) provides an algorithmic procedure for obtaining higher-order asymptotics.

The starting point for the Boundary Function method is a composite expansion of the solutions to (4.11), in the form:

$$x(t, s; \varepsilon) = \tilde{x}(t; \varepsilon) + \hat{x}(s; \varepsilon), \tag{4.15a}$$

$$z(t, s; \varepsilon) = \tilde{z}(t; \varepsilon) + \hat{z}(s; \varepsilon), \tag{4.15b}$$

where $s = \varepsilon^{-1}t$ denotes the "fast" time-scale. Similarly, the vector fields in (4.11) are decomposed as

$$f_x = \tilde{f}_x + \hat{f}_x, \qquad f_z = \tilde{f}_z + \hat{f}_z, \tag{4.16}$$

where $\tilde{f}_x$, $\tilde{f}_z$, $\hat{f}_x$, and $\hat{f}_z$ are defined as:

$$\begin{aligned}
\tilde{f}_x(\tilde{x}, \tilde{z}) &:= f_x(\tilde{x}, \tilde{z}),\\
\tilde{f}_z(\tilde{x}, \tilde{z}) &:= f_z(\tilde{x}, \tilde{z}),\\
\hat{f}_x(\hat{x}, \hat{z}; \tilde{x}, \tilde{z}) &:= f_x(\tilde{x} + \hat{x}, \tilde{z} + \hat{z}) - f_x(\tilde{x}, \tilde{z}),\\
\hat{f}_z(\hat{x}, \hat{z}; \tilde{x}, \tilde{z}) &:= f_z(\tilde{x} + \hat{x}, \tilde{z} + \hat{z}) - f_z(\tilde{x}, \tilde{z}).
\end{aligned}$$

Using (4.15) and (4.16), and computing the t-time derivatives of (x, z), we obtain:

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} + \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\hat{x}}{\mathrm{d}s} = \frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} + \frac{1}{\varepsilon}\frac{\mathrm{d}\hat{x}}{\mathrm{d}s} = \tilde{f}_x + \hat{f}_x, \tag{4.17a}$$

$$\varepsilon\frac{\mathrm{d}\tilde{z}}{\mathrm{d}t} + \varepsilon\frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\hat{z}}{\mathrm{d}s} = \varepsilon\frac{\mathrm{d}\tilde{z}}{\mathrm{d}t} + \frac{\mathrm{d}\hat{z}}{\mathrm{d}s} = \tilde{f}_z + \hat{f}_z. \tag{4.17b}$$

Considering a composite expansion for the solutions $\tilde{x}, \hat{x}, \tilde{z}, \hat{z}$, of the form

$$\tilde{x}(t; \varepsilon) = \tilde{x}_0(t) + \varepsilon\, \tilde{x}_1(t) + O(\varepsilon^2), \tag{4.18a}$$

$$\hat{x}(s; \varepsilon) = \hat{x}_0(s) + \varepsilon\, \hat{x}_1(s) + O(\varepsilon^2), \tag{4.18b}$$

$$\tilde{z}(t; \varepsilon) = \tilde{z}_0(t) + \varepsilon\, \tilde{z}_1(t) + O(\varepsilon^2), \tag{4.18c}$$

$$\hat{z}(s; \varepsilon) = \hat{z}_0(s) + \varepsilon\, \hat{z}_1(s) + O(\varepsilon^2), \tag{4.18d}$$

substituting into the Eqs. (4.17), expanding in a Taylor series, and collecting the terms of like-powers in ε (see Kuehn et al. (2015, Section 9.3) and Vasil'Eva et al. (1995)) yields a hierarchy of equations of which the leading order is:

$$\frac{\mathrm{d}\tilde{x}_0}{\mathrm{d}t} = \tilde{f}_x(\tilde{x}_0, \tilde{z}_0), \tag{4.19a}$$

$$0 = \tilde{f}_z(\tilde{x}_0, \tilde{z}_0), \tag{4.19b}$$

$$\frac{\mathrm{d}\hat{x}_0}{\mathrm{d}s} = 0, \tag{4.19c}$$

$$\frac{\mathrm{d}\hat{z}_0}{\mathrm{d}s} = \hat{f}_z(\hat{x}_0, \hat{z}_0; \tilde{x}_0(0), \tilde{z}_0(0)). \tag{4.19d}$$

The asymptotic properties of the solutions to the system of Eqs. (4.19) with respect to the original system (4.11) are those asserted in the classical Tikhonov's theorem (Khalil, 2002, Theorem 11.1). Subsequent orders, however, provide incremental corrections to the classical error estimates. In particular, the *first-order correction* is given by the system of linear time-varying equations:

$$\frac{\mathrm{d}\tilde{x}_1}{\mathrm{d}t} = J_1\tilde{f}_x(t)\tilde{x}_1 + J_2\tilde{f}_x(t)\tilde{z}_1, \tag{4.20a}$$

$$\frac{\mathrm{d}\tilde{z}_0}{\mathrm{d}t} = J_1\tilde{f}_z(t)\tilde{x}_1 + J_2\tilde{f}_z(t)\tilde{z}_1, \tag{4.20b}$$

$$\frac{\mathrm{d}\hat{x}_1}{\mathrm{d}s} = \hat{f}_x(\hat{x}_0(s), \hat{z}_0(s); \tilde{x}(0), \tilde{z}(0)), \tag{4.20c}$$

$$\frac{\mathrm{d}\hat{z}_1}{\mathrm{d}s} = J_1\hat{f}_z(s)\hat{x}_1 + J_2\hat{f}_z(s)\hat{z}_1 + g_1(s), \tag{4.20d}$$

where we used the condensed notation

$$\begin{aligned}
J_i\tilde{f}_j(t) &:= J_i\tilde{f}_j(\tilde{x}_0(t), \tilde{z}_0(t)), \quad i, j \in \{x, z\}\\
J_i\hat{f}_z(s) &:= J_i\hat{f}_z(\hat{x}_0(s), \hat{z}_0(s); \tilde{x}_0(0), \tilde{z}_0(0)), \quad i \in \{x, z\},
\end{aligned}$$

and where the vector-valued map g_1 is defined by

$$g_1(s) := \left(J_1\hat{f}_z(s) - J_1\tilde{f}_z(0)\right)\left(\tilde{x}_1(0) + \tilde{x}_0'(0)s\right) + \left(J_2\hat{f}_z(s) - J_2\tilde{f}_z(0)\right)\left(\tilde{z}_1(0) + \tilde{z}_0'(0)s\right),$$

where $(\tilde{x}_0, \tilde{z}_0)$ and $(\hat{x}_0, \hat{z}_0)$ are the trajectories of system (4.20), and $\tilde{x}_0'(0) = \mathrm{d}\tilde{x}_0(t)/\mathrm{d}t\big|_{t=0}$. We remark that (4.20b) is an algebraic equation rather than an ODE, since $\mathrm{d}\tilde{z}_0/\mathrm{d}t$ is fully determined by the trajectories of the system (4.19).

In order to be able to use the trajectories of the systems (4.19)–(4.20) to approximate the trajectories of the original system (4.11), we need to determine how the initial conditions of these systems relate to each other. The decomposition (4.15), along with the expansion (4.18), provides the relationships:

$$\tilde{x}_0(0) + \hat{x}_0(0) = x(0), \qquad \tilde{x}_1(0) + \hat{x}_1(0) = 0, \tag{4.21a}$$

$$\tilde{z}_0(0) + \hat{z}_0(0) = z(0), \qquad \tilde{z}_1(0) + \hat{z}_1(0) = 0, \tag{4.21b}$$

However, these relationships are not sufficient to completely define the initial conditions for the systems (4.19)–(4.20). To obtain a complete characterization of the appropriate initial conditions, the Boundary Function method imposes the following condition

$$\lim_{s\to\infty} \hat{x}_j(s) = 0, \qquad j \in \{1, 2\}. \tag{4.22}$$

Indeed, it is shown in Vasil'Eva et al. (1995) that condition (4.22), along with the standing assumption in the Boundary Function method, which is that the linearization of the map f_2 around the quasi-steady manifold φ_0 is uniformly Hurwitz, provides a complete characterization of the initial conditions. In particular, as a result of imposing the condition (4.22), the appropriate initial conditions of the systems (4.19)–(4.20) are expressed in terms of the initial conditions of the original system (4.11) by the relationships:

$$\tilde{x}_0(0) = x(0), \tag{4.23a}$$

$$\hat{x}_0(0) = 0, \tag{4.23b}$$

$$\hat{z}_0(0) = z(0) - \varphi_0(x(0)) \tag{4.23c}$$

$$\tilde{x}_1(0) = -\hat{x}_1(0) = \int_0^\infty g_2(s)\,ds \tag{4.23d}$$

$$\hat{z}_1(0) = -\tilde{z}_1(0) \tag{4.23e}$$

where g_2 is defined as follows:

$$g_2(s) = \hat{f}_x(\hat{x}_0(s), \hat{z}_0(s); \tilde{x}_0(0), \tilde{z}_0(0)).$$

It can be shown that the improper integral in (4.23d) is well-defined due to the uniform Hurwitz condition on the linearization of the vector field f_2 around the quasi-steady manifold φ_0. We refer the reader to the monograph (Vasil'Eva et al., 1995) for the details.

The trajectories of the systems (4.19)–(4.20), with the initial conditions (4.23), provide a second-order approximation to the solutions of the original dynamics (4.11). In particular, it is shown in Vasil'Eva et al. (1995) that carrying out the procedure, outlined above for the first-order correction, leads to asymptotic approximations with arbitrary order to the solutions of the singularly perturbed system (4.11) under sufficient smoothness properties. The following theorem, specialized from Vasil'Eva et al. (1995), characterizes the asymptotic properties of the second-order approximations produced by the Boundary Function method:

Theorem 4.2 (*Vasil'Eva et al., 1995*). *Let $K_0 \subset \mathbb{R}^{n_1+n_2}$ be compact. Consider system (4.11) with initial conditions $(x(0), z(0)) \in K_0$, and let f_x, f_z be sufficiently smooth. Suppose that the linearization of f_z around the quasi-steady state φ_0 is uniformly Hurwitz, and there exists $T_f > 0$ such that the functions*

$$(\tilde{x}(t;\varepsilon), \tilde{z}(t;\varepsilon)) = \sum_{k=0}^{1} \left(\varepsilon^k \tilde{x}_k(t), \varepsilon^k \tilde{z}_k(t)\right),$$

are well-defined for all $t \in [0, T_f]$, and correspond to the unique solutions to the first two equations in (4.19) and (4.20) with the initial conditions defined by (4.23). Then, there exists $\varepsilon^* > 0$ such that for all $\varepsilon \in (0, \varepsilon^*)$, system (4.11) has unique solutions (x, z) that satisfy:

$$|x(t) - \tilde{x}(t;\varepsilon) - \hat{x}(t/\varepsilon;\varepsilon)| = O(\varepsilon^2), \qquad \forall\ t \in [0, T_f],$$

$$|z(t) - \tilde{z}(t;\varepsilon) - \hat{z}(t/\varepsilon;\varepsilon)| = O(\varepsilon^2), \qquad \forall\ t \in [0, T_f].$$

for all $(x(0), z(0)) \in K_0$. □

Having introduced the Boundary Function method for the construction of higher-order asymptotics in singular perturbation problems, and motivated by a class of highly oscillatory systems that emerge in the context of chemotactic navigation of sperm cells (Abdelgalil et al., 2022) and source seeking for nonholonomic vehicles (Abdelgalil et al., 2023), we show in the next subsection that higher-order averaging can be combined with the key ideas of the Boundary Function method to provide a higher-order extension of *two-time scale averaging with trade-off*. As before, we restrict our presentation to the second-order case.

4.2. Second-order two-time scale averaging

Consider the following system:

$$\dot{x} = \sum_{k=1}^{2} \varepsilon^{k-2} f_{x,k}(x, z, \tau), \tag{4.24a}$$

$$\dot{z} = \sum_{k=0}^{2} \varepsilon^{k-2} f_{z,k}(x, z, \tau), \tag{4.24b}$$

$$\dot{\tau} = \varepsilon^{-2}, \tag{4.24c}$$

where $x \in \mathbb{R}^{n_1}$, $z \in \mathbb{R}^{n_2}$, $\tau \geq 0$, and $\varepsilon > 0$. Note that this system is of the form (2.2) with $C = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, $D = \emptyset$, and $\rho = \varepsilon^{-2}$. We make the following regularity assumptions on (4.24):

Assumption 4.1. The functions $f_{x,k}$ and $f_{z,k}$ are C^{3-k} in the first two arguments, continuous and T-periodic in the last argument, and the function $f_{x,1}$ satisfies $\int_0^T f_{x,1}(x, z, \tau)\,d\tau = 0$. Moreover, there exists a unique C^3 mapping $\varphi_0 : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ such that $f_{z,0}(x, \varphi_0(x), \tau) = 0$, for all $x \in \mathbb{R}^{n_1}$ and all $\tau \in \mathbb{R}_{\geq 0}$. □

To study system (4.24), we will leverage the exponential stability properties of the boundary layer dynamics, in the same spirit of the Boundary Function method. These properties are expressed in terms of Lyapunov functions, which will also play a prominent role in Sections 5–6.

Assumption 4.2. There exists a C^1 function $V : \mathbb{R}^{n_2} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, and constants $\kappa_i > 0$, $i \in \{1, \ldots, 4\}$, such that for all $y \in \mathbb{R}^{n_2}$, all $x \in \mathbb{R}^{n_1}$, and all $\tau \in \mathbb{R}_{\geq 0}$, the following inequalities holds:

$$\kappa_1|y|^2 \leq V(y, \tau) \leq \kappa_2|y|^2, \tag{4.25a}$$

$$\nabla_1 V^\top \tilde{f}_{z,0}(y, \tau) + \nabla_2 V(y, \tau) \leq -\kappa_3|y|^2, \tag{4.25b}$$

$$|\nabla_1 V(y, \tau)| \leq \kappa_4|y|, \tag{4.25c}$$

where $\tilde{f}_{z,0}(x, y, \tau) := f_{z,0}(x, y + \varphi_0(x), \tau)$. □

Next, we introduce the reduced order system with state $(\tilde{x}, \tau) \in \mathbb{R}^{n_1} \times \mathbb{R}_{\geq 0}$, and dynamics:

$$\dot{\tilde{x}} = \sum_{k=1}^{2} \varepsilon^{k-2} \tilde{f}_k(\tilde{x}, \tau), \qquad \dot{\tau} = \varepsilon^{-2}, \tag{4.26}$$

where the functions $\tilde{f}_k$ are defined as:

$$\tilde{f}_1(\tilde{x}, \tau) = f_{x,1}(\tilde{x}, \varphi_0(\tilde{x}), \tau), \tag{4.27a}$$

$$\tilde{f}_2(\tilde{x}, \tau) \quad = f_{x,2}(\tilde{x}, \varphi_0(\tilde{x}), \tau) + J_2 f_{x,1}(\tilde{x}, \varphi_0(\tilde{x}), \tau)\varphi_1(\tilde{x}, \tau), \tag{4.27b}$$

$$\varphi_1(\bar{x},\tau) = \int_{\tau}^{\tau+T} E_{\Phi}(\tau,\nu) b_{\varphi}(\bar{x},\nu)\, d\nu, \tag{4.27c}$$

$$b_{\varphi}(\bar{x},\tau) \quad = f_{z,1}(\bar{x},\varphi_0(\bar{x}),\tau) - J\varphi_0(\bar{x})\, f_{x,1}(\bar{x},\varphi_0(\bar{x}),\tau), \tag{4.27d}$$

with a matrix-valued mapping E_{Φ} given by:

$$E_{\Phi}(\tau,\nu) = (I - \Phi(\tau+T,\tau))^{-1}\, \Phi(\tau+T,\nu),$$

where Φ is the state-transition matrix associated with the following linear time-periodic system:

$$\frac{dy}{d\tau} = A(\tau;\bar{x})y, \quad A(\tau;\bar{x}) := J_2 f_{z,0}(\bar{x},\varphi_0(\bar{x}),\tau).$$

The reduced system (4.26) can be studied via second-order averaging with trade-off. In particular, its reduced order averaged dynamics are given by

$$\dot{\bar{x}} = \frac{1}{T}\int_0^T \left(\tilde{f}_2(\bar{x},\tau) + \frac{1}{2}[\tilde{w}_1,\tilde{f}_1](\bar{x},\tau)\right) d\tau, \tag{4.28}$$

where $\tilde{w}_1$ is

$$\tilde{w}_1(\bar{x},\tau) = \int_0^{\tau} \tilde{f}_1(\bar{x},\nu)\, d\nu. \tag{4.29}$$

The following theorem provides an approximation result for the trajectories of the original system (4.24) and the reduced-order averaged system (4.28). The proof is presented in Section 7.

Theorem 4.3. *Suppose that Assumptions 4.1 and 4.2 hold, and let $K_1 \subset \mathbb{R}^{n_1}$ be a compact set. Suppose there exists $T_f \in \mathbb{R}_{>0}$ such that for each $x_0 \in K_1$, system (4.28) has a unique solution $t \mapsto \bar{x}(t)$ with $\bar{x}(0) = x_0$, defined for all $t \in [0, T_f]$. Then, for each compact set $K_2 \subset \mathbb{R}^{n_2}$ and each $\epsilon \in \mathbb{R}_{>0}$ there exists $\varepsilon^* \in (0,\varepsilon_0)$ and $\lambda,\gamma \in \mathbb{R}_{>0}$ such that for all $\varepsilon \in (0,\varepsilon^*)$ and all $(x_0, z_0 - \varphi_0(x_0)) \in K_1 \times K_2$, system (4.24) has a unique solution (x,z) with $(x(0),z(0)) = (x_0,z_0)$, satisfying*

$$|x(t) - \bar{x}(t)| \le \epsilon,$$

$$|z(t) - \bar{z}(t)| \le \gamma |z_0 - \varphi_0(x_0)|\, e^{-\lambda\varepsilon^{-2}t} + \epsilon,$$

for all $t \in [0, T_f]$, where $\bar{z}(t) = \varphi_0(\bar{x}(t))$. □

Remark 4.4. Similar to Theorems 3.2 and 3.4, Theorem 4.3 is an approximation result that establishes (a variation of) the closeness-of-solutions property (on compact time domains and compact sets) between system (4.24) and system (4.28). However, unlike Theorem 3.2 and Theorem 3.4, the approximation provided by Theorem 4.3 is not uniform on the interval $[0, T_f]$ which is to be expected since the system (4.24) is singularly perturbed. Nevertheless, it can be shown that uniform stability properties (e.g., expressed in terms of $\mathcal{KL}$ bounds) of system (4.28) can be inherited by the system (4.24) in a practical sense. We do not further pursue this idea in paper. □

Going back to Example 4.1, we observe that for the case when $\sigma_1 = \sigma_2 = 1$, $\mu = \varepsilon^2$, and using (4.28), the reduced order averaged system is given by:

$$\dot{\bar{x}} = \left(\beta - \frac{k}{2}\right)\bar{x}, \tag{4.30}$$

whose trajectories converge to the origin exponentially fast whenever $0 < \beta < k/2$, thus explaining the behavior observed in Fig. 3 for the case when $\mu = O(\varepsilon^2)$. Finally, we remark that system (4.5) also highlights the intricacies of multiple time-scale phenomena when more than one parameter is involved. Indeed, depending on the relative asymptotic order of μ with respect to ε, the system can exhibit strikingly different behaviors.

One of the motivations for the study of systems of the form (4.24) arises from the phenomenon of sperm chemotaxis (Abdelgalil et al., 2022; Alvarez, Friedrich, Gompper, & Kaupp, 2014; Friedrich & Jülicher, 2007). In the next section, we give a brief exposition of this phenomenon

4.3. Application to 3D chemotactic navigation in sperm cells

Certain marine invertebrates, such as sea urchins, are broadcast spawners, i.e., male and female gametes are released into open water, where fertilization takes place (Abdelgalil et al., 2022; Alvarez et al., 2014; Friedrich & Jülicher, 2007; Jikeli et al., 2015). In this process, an egg cell secretes chemical cues that act as chemoattractants for the sperm cells. Diffusion in the surrounding water establishes an approximately radial concentration field centered around the egg. Sperm cells then propel themselves through the undulating motion of their flagella. The asymmetric beating pattern of the flagella leads to a periodic trajectory pattern, which forms a helical trajectory in 3D. The presence of chemoattractants, coupled with the periodic trajectory pattern, exposes the sperm cell to a periodic stimulus. Through a simple modulation of the mean curvature of the beating flagella with the periodic variations in the perceived stimulus (Jikeli et al., 2015), sperm cells can dynamically align the centerline of their helical trajectory with the direction of the increasing stimulus. This direction corresponds to higher chemoattractant concentration, allowing the sperm cells to swim towards an egg cell.

To model the above phenomenon, we assume that the sperm cells flow in low Reynolds number fluid regimes, where their motion can be well-approximated by kinematic models (Friedrich, Riedel-Kruse, Howard, & Jülicher, 2010). In this case, the kinematics of rigid body motion in 3D is given by the system:

$$\dot{q} = R\,v, \qquad \dot{R} = R\,\hat{\Omega}, \tag{4.31}$$

where $R \in SO(3)$ is a rotation matrix, $q \in \mathbb{R}^3$ is the position of the cell, $v \in \mathbb{R}^3$ is the velocity in body coordinates, and $\Omega \in \mathbb{R}^3$ is the angular velocity in body coordinates. The mean curvature of the beating flagellum affects the angular velocities of the cell in body coordinates. A simplified model that captures the essential features of this behavior assumes that the linear and angular velocities in body coordinates take the form:

$$v = \begin{pmatrix} v_{p,0} \\ 0 \\ 0 \end{pmatrix}, \quad \Omega(u) = \begin{pmatrix} \Omega_{p,0} \\ 0 \\ \Omega_{\perp} \end{pmatrix} + \begin{pmatrix} \Omega_{p,1} \\ 0 \\ 0 \end{pmatrix} u, \tag{4.32}$$

where u is the mean curvature of the flagellum, and $v_{p,0}, \Omega_{p,0}, \Omega_{p,1}$, and $\Omega_{\perp}$ are constant coefficients. Let the local concentration of the chemoattractant at the position q be denoted by $\phi(q)$. Exposure to the chemoattractant molecules stimulates the signaling pathway of the sperm cell. The simplest model of such stimulation dynamics is the first-order lag

$$\sigma_1 \mu \dot{z}_1 = \phi(q) - z_1, \tag{4.33}$$

where z_1 is the excitation due to the stimulus $\phi(q)$. In addition, the signaling pathway of the sperm cell is known for its ability to adapt to the mean stimulus level. The simplest model of such adaptation is the first-order lag

$$\sigma_2 \mu \dot{z}_2 = z_1 - z_2, \tag{4.34}$$

where z_2 is the amount of relaxation in response to an excitation z_1. Therefore, the adaptive response of the pathway due to the local concentration $\phi(q)$ is given by the difference between the excitation and relaxation, i.e. $z_1 - z_2$. Finally, the simplest transfer function between the adaptive response of the signaling pathway $z_1 - z_2$ and the mean curvature of the flagellum u is a simple gain k, that is

$$u = 2k(z_1 - z_2). \tag{4.35}$$

Given this simple model of the signaling pathway, we observe that the dynamic relation between the local stimulus $\phi(q)$ and the mean curvature of the flagellum u is given by an LTI system whose transfer function coincides with the transfer function of the bandpass filter $G(\hat{s})$ in (4.5). Indeed, the step response of the signaling pathway has been

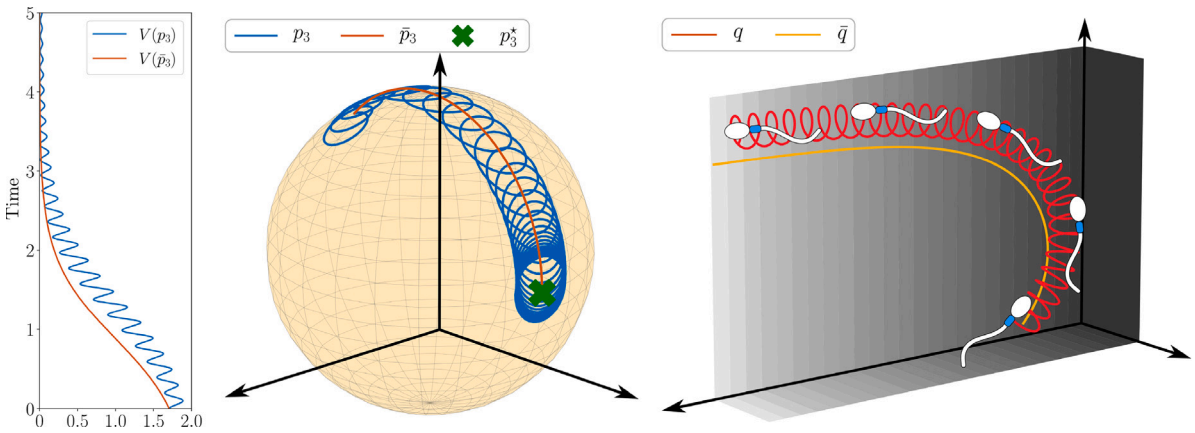

Fig. 4. Simulation results for system (4.36) for the case $\nabla\phi(q) \neq 0$ but $\nabla^2\phi(q) = 0$. Left plot: the misalignment between the gradient $\nabla\phi$ and the unit vectors p_3 and $\bar{p}_3$ defined by $V(p_3) = 1 - \langle p_3, \nabla\phi/|\nabla\phi|\rangle$. Center plot: the trajectory of p_3 and $\bar{p}_3$ on the unit sphere $\mathbb{S}^2$ where $p_3^\star = \nabla\phi/|\nabla\phi|$ is the normalized gradient. Right plot: evolution of the trajectory of q and $\bar{q}$ in $\mathbb{R}^3$. The simulation parameters are $\tilde{v}_p = \Omega_{p,0} = \Omega_{p,1} = \sigma_1 = \sigma_2 = 1$, $k = 2$, and $\epsilon = 1/\sqrt{10\pi}$.

shown to closely resemble that of a band-pass filter (Abdelgalil et al., 2022; Alvarez et al., 2012; Kaupp et al., 2003).

To study the dynamic properties of the sperm cells, let $\epsilon \in (0, 1)$, $v_{p,0} = \tilde{v}_p/\epsilon$, $\Omega_{p,0} = \tilde{\Omega}_{p,0}/\epsilon$, $\Omega_{p,1} = \tilde{\Omega}_{p,1}/\epsilon$, $\Omega_\perp = 1/\epsilon^2$, and $\mu = \epsilon^2$. In addition, let $P = R\exp(-\Omega_o t\hat{e}_3)$, where e_3 is the unit vector $e_3 = (0, 0, 1)$ and $\hat{e}_3$ is the skew-symmetric matrix associated with e_3. Then, partition the matrix $P \in SO(3)$ into three columns with $P = [p_1, p_2, p_3]$, where $p_i \in \mathbb{S}^2$. Since $P \in SO(3)$, the vectors p_i form an orthonormal basis of $\mathbb{R}^3$. Using this change of coordinates, system (4.31)–(4.35) takes the form

$$\dot{q} = \tilde{v}_p\epsilon^{-1}\cos\left(\epsilon^{-2}t\right)p_1 + \tilde{v}_p\epsilon^{-1}\sin\left(\epsilon^{-2}t\right)p_2 \tag{4.36a}$$

$$\dot{p}_1 = -(\tilde{\Omega}_{p,0} + \tilde{\Omega}_{p,1}u)\epsilon^{-1}\sin\left(\epsilon^{-2}t\right)p_3 \tag{4.36b}$$

$$\dot{p}_2 = (\tilde{\Omega}_{p,0} + \tilde{\Omega}_{p,1}u)\epsilon^{-1}\cos\left(\epsilon^{-2}t\right)p_3 \tag{4.36c}$$

$$\dot{p}_3 = (\tilde{\Omega}_{p,0} + \tilde{\Omega}_{p,1}u)\epsilon^{-1}\left(\sin\left(\epsilon^{-2}t\right)p_1 - \cos\left(\epsilon^{-2}t\right)p_2\right) \tag{4.36d}$$

where the input u is given by $u = 2k(z_1 - z_2)$, and

$$\dot{z}_1 = \epsilon^{-2}{\sigma_1}^{-1}\left(\phi(q) - z_1\right), \tag{4.36e}$$

$$\dot{z}_2 = \epsilon^{-2}{\sigma_2}^{-1}\left(z_1 - z_2\right). \tag{4.36f}$$

It follows that system (4.36) takes the same form as the class of systems (4.24) with $x = (q, p_1, p_2, p_3)$ and $z = (z_1, z_2)$. Therefore, we may apply the formulas (4.27)–(4.29) to obtain the following reduced-order averaged system:

$$\begin{aligned}
\dot{\bar{q}} &= \tilde{v}_p\tilde{\Omega}_{p,0}\bar{p}_3\\
\dot{\bar{p}}_1 &= k\tilde{v}_p\tilde{\Omega}_{p,1}\left(\gamma_1\bar{p}_3\bar{p}_1^\top + \gamma_2\bar{p}_3\bar{p}_2^\top\right)\nabla\phi(\bar{q}) + \frac{1}{2}\tilde{\Omega}_{p,0}^2\bar{p}_2\\
\dot{\bar{p}}_2 &= k\tilde{v}_p\tilde{\Omega}_{p,1}\left(\gamma_1\bar{p}_3\bar{p}_2^\top - \gamma_2\bar{p}_3\bar{p}_1^\top\right)\nabla\phi(\bar{q}) - \frac{1}{2}\tilde{\Omega}_{p,0}^2\bar{p}_1\\
\dot{\bar{p}}_3 &= k\tilde{v}_p\tilde{\Omega}_{p,1}\gamma_1\left(I - \bar{p}_3\bar{p}_3^\top\right)\nabla\phi(\bar{q}),
\end{aligned}$$

where the constants γ_1 and γ_2 are given by $\gamma_1 = \frac{\sigma_2(\sigma_1+\sigma_2)}{(1+{\sigma_1}^2)(1+{\sigma_2}^2)}$ and $\gamma_2 = \frac{\sigma_2(1-\sigma_1\sigma_2)}{(1+{\sigma_1}^2)(1+{\sigma_2}^2)}$.

Since the dynamics of $\bar{q}$ and $\bar{p}_3$ do not depend on $\bar{p}_1$ and $\bar{p}_2$, we can restrict our attention to the dynamics of $\bar{q}$ and $\bar{p}_3$, which evolves on $\mathbb{R}^3 \times \mathbb{S}^2$. The vector $\bar{p}_3 \in \mathbb{S}^2$ is a unit vector that points along the centerline of the helical trajectory. The chemotactic behavior emerges through the dynamics of $\bar{p}_3$. For illustration, let us assume that the concentration field is linear, i.e. $\nabla^2\phi(q) = 0$, which is a valid assumption for a local analysis when the concentration field is weakly non-linear, i.e. $|\nabla^2\phi(q)| \ll |\nabla\phi(q)|^2$. In this case, the dynamics of $\bar{p}_3$ decouples further from the dynamics of $\bar{q}$, leading to the system:

$$\dot{\bar{q}} = \tilde{v}_p\tilde{\Omega}_{p,0}\bar{p}_3,\ \dot{\bar{p}}_3 = k\tilde{v}_p\tilde{\Omega}_{p,1}\gamma_1\left(I - \bar{p}_3\bar{p}_3^\top\right)\nabla\phi, \tag{4.37}$$

where $\nabla\phi$ does not depend on q. The operator $\left(I - \bar{p}_3\bar{p}_3^\top\right)$ is the projection onto the orthogonal complement of the linear space spanned by $\bar{p}_3$. Therefore, the helical centerline will dynamically bend until it is aligned with the gradient. We illustrate this chemotactic behavior through the simulation results in Fig. 4. As the figure illustrates, the axis of the helical trajectory dynamically bends in order to align with the (constant) direction of the gradient. In particular, the reduced-order average system provides a suitable approximation of the expected behavior of the trajectories of the cell. The deviation between the trajectories of the original and the reduced-order averaged system on a compact time interval can be made arbitrarily small by choosing a sufficiently small ϵ.

We conclude this section by highlighting recent research on chemotactic navigation (Kromer, Märcker, Lange, Baier, & Friedrich, 2018; Li, Chakrabarti, Castilla, Mahajan, & Saintillan, 2022), which suggests that sperm cells may demonstrate diverse dynamics during the seeking process, switching between different operating modes depending on their current state. The study of such systems will require suitable extensions of second-order two-time scale averaging theory for switching and hybrid dynamical systems.

5. Singularly perturbed hybrid dynamical systems

In the previous sections, we considered second-order averaging and singular perturbation tools for continuous-time dynamical systems modeled as ODEs. In this section, we now turn our attention to multi-time scale dynamical systems that also incorporate discrete-time dynamics. Such systems, called hybrid dynamical systems (HDS), can be modeled as (2.1) or (2.2), with $D \neq \emptyset$, and their solutions are parameterized by both continuous-time and discrete-time indices evolving on hybrid time domains (c.f., Definition 2.1). Since HDS usually exhibit non-smooth solutions with complex behaviors, second-order perturbation-based analyses of their trajectories, similar to those presented in Sections 3 and 4, are more challenging and the subject of ongoing research. Nevertheless, HDS that exhibit multiple time scales in their continuous-time dynamics can be studied via *first-order* singular perturbation techniques using Lyapunov-based methods.

The use of Lyapunov-based conditions to analyze multi-time scale dynamical systems has a rich history in the field of adaptive control (Kosut et al., 1987; Sastry & Bodson, 1989) and high-gain observers (Grujić, 1979; Khalil, 2002; Kokotović et al., 1986; Saberi & Khalil, 1984). For instance, a set of sufficient Lyapunov-like conditions for certifying asymptotic stability in first-order singularly perturbed ordinary differential equations (ODEs) was studied in Saberi and Khalil (1984). These conditions are expressed in terms of quadratic-type Lyapunov functions that decrease along the trajectories of both

the reduced dynamics and the boundary layer dynamics. They also satisfy suitable interconnection inequalities, akin to small-gain conditions. In the literature of singular perturbations, this approach is commonly referred to as the "composite Lyapunov method" (Grujić, 1979; Khalil, 2002; Narang-Siddarth & Valasek, 2014; Saberi & Khalil, 1984). It has been widely applied in various domains, including consensus and distributed optimization problems (Kia, Cortés, & Martínez, 2013), feedback optimization algorithms (Colombino et al., 2020; Hauswirth et al., 2020), competitive neural networks (Meyer-Bäse, Ohl, & Scheich, 1996), multi-link flexible robots (Zhu, Commuri, & Lewis, 1994), power system transient stability analysis (Anup, Verma, & Bhatti, 2022), and network control systems (Heijmans et al., 2018). Thus, motivated by practical applications of hybrid systems in feedback optimization (Bianchin, Poveda et al., 2022), parameter estimation for adaptive control (Ochoa et al., 2021), and traffic network control systems (Bianchin, Cortes et al., 2022; Kutadinata et al., 2016), in this section, we discuss how Lipschitz Lyapunov-like functions of quadratic-type can also be used for the stability analysis of singularly perturbed HDS. Similar sufficient conditions have been discussed in Fang, Liu, Sun, and Teel (2020), Heijmans et al. (2018) and Sanfelice and Teel (2011). We present different numerical examples to illustrate the results.

5.1. Model and main assumptions

We consider the singularly perturbed (SP) HDS:

$$(x,z) \in C := C_x \times C_z, \quad \begin{cases} \dot{x} \in F_x(x,z) \\ \varepsilon \dot{z} \in F_z(x,z) \end{cases} \tag{5.1a}$$

$$(x,z) \in D := D_x \times D_z, \quad (x^+, z^+) \in G(x,z), \tag{5.1b}$$

where $\varepsilon \in \mathbb{R}_{>0}$ is a small parameter, $x \in \mathbb{R}^{n_1}$ is the "slow" state of the system, $z \in \mathbb{R}^{n_2}$ is the "fast" state, $F_x : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightrightarrows \mathbb{R}^{n_1}$ and $F_z : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightrightarrows \mathbb{R}^{n_2}$ are set-valued mappings that characterize the continuous-time dynamics of x and z, respectively, $G : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightrightarrows \mathbb{R}^{n_1+n_2}$ characterizes the jump map of the system, and the sets $C_x \subset \mathbb{R}^{n_1}, C_z \subset \mathbb{R}^{n_2}$, and $D_x \subset \mathbb{R}^{n_1}, D_z \subset \mathbb{R}^{n_2}$, characterize the flow set C and the jump set D, respectively.

Similar to the continuous-time systems studied in Section 4, when ε is sufficiently small the stability properties of system (5.1) can be studied based on the stability properties of a simpler reduced hybrid system. To characterize these reduced dynamics, we first introduce the map φ_0, which plays the role of the quasi-steady state map in classical singular perturbation problems, characterized by the following assumption:

Assumption 5.1. There exists a $\mathcal{C}^1$-map $\varphi_0 : \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ satisfying $\varphi_0(x) \in C_z$ for all $x \in C_x$. □

Remark 5.1. Assumption 5.1 is similar to the standard assumptions in ODEs, see Khalil (2002, Ch. 11) and Kokotović et al. (1986). In general, for the sake of stability analysis, it is not necessary to assume that the quasi-steady state is a single-valued map, see for instance Sanfelice and Teel (2011). For example, in Section 6.4.3, we consider a system where a multi-valued quasi-steady state map is used when discussing hybrid extremum-seeking controllers applied to self-tuning amplitude problems in oscillators. Nevertheless, we assume singled-valued maps to follow a similar approach as in Section 4 and to parallel the presentation for singularly-perturbed ODEs considered in Khalil (2002). □

By using the mapping φ_0, we introduce the shifted state $y = z - \varphi_0(x)$, and the change of coordinates from (x,z) to $\psi = (x,y)$. The coordinate change leads to the following HDS:

$$(x,y) \in \hat{C}, \quad \begin{cases} \dot{x} \in \hat{F}_x(x,y) \\ \varepsilon \dot{y} \in \hat{F}_y(x,y,\varepsilon) \end{cases} \tag{5.2a}$$

$$(x,y) \in \hat{D}, \quad (x^+, y^+) \in \hat{G}(x,y), \tag{5.2b}$$

where the set-valued flow maps $\hat{F}_x$ and $\hat{F}_y$ are defined by

$$\hat{F}_x(x,y) := \left\{ f_x \,:\, f_x \in F_x(x, y+\varphi_0(x)) \right\},$$

$$\hat{F}_y(x,y,\varepsilon) := \Big\{ f_z - \varepsilon J\varphi_0(x) f_x \,:\, f_z \in F_z(x, y+\varphi_0(x)), \; f_x \in F_x(x, y+\varphi_0(x)) \Big\}.$$

Since φ_0 is $\mathcal{C}^1$, the Jacobian $J\varphi_0(x) \in \mathbb{R}^{n_2 \times n_1}$ is a well-defined continuous single-valued map. The jump map $\hat{G}$ is given by

$$\hat{G}(x,y) := \Big\{ (v_1, v_2) : v_1 = g_1, \; v_2 = g_2 - \varphi_0(g_1), \; (g_1, g_2) \in G(x, y+\varphi_0(x)) \Big\},$$

the flow set $\hat{C}$ can be written as

$$\hat{C} := \{(x,y) \;:\; x \in C_x, \; y + \varphi_0(x) \in C_z\},$$

and the jump set $\hat{D}$ can be written as

$$\hat{D} := \{(x,y) \;:\; x \in D_x, \; y + \varphi_0(x) \in D_z\}.$$

We make the standing assumption that for any $\varepsilon > 0$ system (5.2) satisfies the hybrid basic conditions of Assumption 2.1.

Next, we introduce the *boundary layer dynamics* for the SP-HDS (5.2), which ignores the jumps (5.2b) and treats x as a constant:

$$(x,y) \in \hat{C}, \quad \dot{y} \in \hat{F}_y(x,y,0), \quad \dot{x} = 0. \tag{5.3}$$

We also introduce the *reduced* HDS associated with the SP-HDS (5.2). Unlike the reduced dynamics of Sections 3 and 4.2, this reduced system is hybrid. It has a state $\tilde{x} \in \mathbb{R}^{n_1}$ and dynamics

$$\tilde{x} \in C_x, \quad \dot{\tilde{x}} \in \tilde{F}(\tilde{x}), \tag{5.4a}$$

$$\tilde{x} \in D_x, \quad \tilde{x}^+ \in \tilde{G}(\tilde{x}), \tag{5.4b}$$

where the set-valued mappings $\tilde{F} : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{n_1}$ and $\tilde{G} : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{n_1}$ are defined as follows:

$$\tilde{F}(\tilde{x}) := \hat{F}_x(\tilde{x}, 0), \tag{5.5a}$$

$$\tilde{G}(\tilde{x}) := \left\{ s \in \mathbb{R}^{n_1} \,:\, (s,l) \in \hat{G}(\tilde{x}, y), \; y + \varphi_0(\tilde{x}) \in D_z \right\}. \tag{5.5b}$$

Note that, since D_z is not necessarily assumed to be compact, in order to have a locally bounded mapping $\tilde{G}$ one might need that $\hat{G}(\tilde{x}, \cdot)$ is uniformly bounded, or alternatively, that the jumps of x do not depend on the fast state z in Eq. (5.1b).

The following assumption captures the stability properties of the boundary-layer dynamics (5.3). For generality, and since it is common to have non-smooth Lyapunov functions in hybrid systems, we consider Lyapunov functions that are locally Lipschitz and regular, but not necessarily continuously differentiable.

Assumption 5.2. There exists a locally Lipschitz and regular function $W : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}_{\geq 0}$, functions $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$, a continuous function $\varphi_y \in \mathcal{P}_s\mathcal{D}$, and a constant $k_y \in \mathbb{R}_{>0}$ such that the following conditions hold: (1) $\alpha_1(|y|) \leq W(x,y) \leq \alpha_2(|y|)$ for all $(x,y) \in \hat{C} \cup \hat{D} \cup \hat{G}(\hat{D})$; and (2) $\max_{v \in \partial_y W(x,y)} \langle v, \hat{f}_y \rangle \leq -k_y \varphi_y^2(y)$, for all $(x,y) \in \hat{C}$ and all $\hat{f}_y \in \hat{F}_y(x,y,0)$. □

When φ_y is also positive definite, the conditions of Assumption 5.2 establish UGAS of the origin for the boundary-layer dynamics (5.3), uniformly on x. In addition, when $\hat{F}_y$ is single-valued and locally Lipschitz, and W is $\mathcal{C}^2$, Assumptions 5.1–5.2 recover the quadratic-type conditions considered in the literature of singularly perturbed ODEs (Khalil, 2002, Ch.11.4). We note that other stability and Lyapunov-based characterizations that enable the incorporation of exogenous disturbances in singularly perturbed ODEs have also been studied in the literature using the notion of input-to-state stability (Christofides & Teel, 1996; Christofides, Teel, & Daoutidis, 1996) .

The stability properties of the reduced hybrid dynamics (5.4) are studied with respect to a compact set $\mathcal{A} \subset \mathbb{R}^{n_1}$, and are characterized by the following two assumptions:

Assumption 5.3. There exists a locally Lipschitz and regular function $V : \mathbb{R}^{n_1} \to \mathbb{R}_{\geq 0}$, functions $\alpha_3, \alpha_4 \in \mathcal{K}_\infty$, a continuous function $\varphi_x \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$, and a constant $k_x \in \mathbb{R}_{>0}$, such that: (1) $\alpha_3(|\tilde{x}|_{\mathcal{A}}) \leq V(\tilde{x}) \leq \alpha_4(|\tilde{x}|_{\mathcal{A}})$, for all $\tilde{x} \in C_x \cup D_x \cup \tilde{G}(D_x)$; (2) $\max_{v \in \partial V(\tilde{x})} \langle v, \tilde{f}_x \rangle \leq -k_x \varphi_x^2(\tilde{x})$, for all $\tilde{x} \in C_x$ and all $\tilde{f}_x \in \tilde{F}(\tilde{x})$. □

When V is $\mathcal{C}^2$, the left-hand side in the second condition of Assumption 5.3 can be substituted by the inner product $\langle \nabla V, \tilde{f}_x \rangle$. The following assumption considers the same function V of Assumption 5.3.

Assumption 5.4. There exists a continuous function $\rho_x \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$ and a constant $c_x \in \mathbb{R}_{>0}$, such that $V(\tilde{g}) - V(\tilde{x}) \leq -c_x \rho_x(\tilde{x})$, for all $\tilde{x} \in D_x$ and all $\tilde{g} \in \tilde{G}(\tilde{x})$. □

As discussed in Christofides et al. (1996, Sec. 6), in general, further interconnection conditions are needed to guarantee stability in singularly perturbed systems that have asymptotically stable boundary layers and reduced dynamics. The following assumptions provide some sufficient interconnection conditions for the flows and jumps of (5.2). Below, in Assumptions 5.5–5.7 the functions V, W, φ_x, and φ_y are the same from Assumptions 5.2–5.4.

Assumption 5.5. There exist $k_1, k_2, k_3 \in \mathbb{R}_{>0}$, such that:

(a) For all $(x, y) \in \hat{C}$, and for all $f_x, \hat{f}_x \in \hat{F}_x(x, y)$, we have:

$$\max_{v \in \partial_x W(x,y)} \langle v, f_x \rangle - \max_{v \in \partial_y W(x,y)} \langle v, J\varphi_0(x)\hat{f}_x \rangle \leq k_1 \varphi_y(y)\varphi_x(x) + k_2 \varphi_y^2(y), \tag{5.6}$$

(b) For all $(x, y) \in \hat{C}$, and for all $\hat{f}_x \in \hat{F}_x(x, y)$, there exists $\tilde{f}_x \in \tilde{F}(x)$ such that:

$$\max_{v \in \partial V(x)} \langle v, \hat{f}_x - \tilde{f}_x \rangle \leq k_3 \varphi_y(y)\varphi_x(x). \tag{5.7}$$

The following two additional assumptions will be used whenever the composite hybrid Lyapunov function constructed from V and W does not necessarily decrease during jumps.

Assumption 5.6. There exist $k_4 \in \mathbb{R}_{>0}$ and a continuous function $\rho_4 \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$, such that

$$W(g_x, g_z) - W(x, y) \leq k_4 \rho_4(x), \tag{5.8}$$

for all $(g_x, g_z) \in \hat{G}(x, y)$ and all $(x, y) \in \hat{D}$. □

Assumption 5.7. There exist $k_5 \in \mathbb{R}_{>0}$, and a continuous function $\rho_5 \in \mathcal{P}_s D$, such that

$$V(\tilde{g}) - V(x) \leq k_5 \rho_5(y),$$

for all $\tilde{g} \in \tilde{G}(\tilde{x})$, and for all $(\tilde{x}, y) \in \hat{D}$. □

When $D = G = \emptyset$, $\mathcal{A} = \{0\}$, $\varphi_x, \varphi_y \in \mathcal{PD}$, and the flow map is singled-valued and locally Lipschitz, the conditions of Assumptions 5.2–5.5 essentially recover the quadratic-type characterization presented in Saberi and Khalil (1984) for locally Lipschitz ODEs. In this sense, the different conditions of Assumptions 5.2–5.7 are natural extensions to study the stability properties of HDS with respect to compact sets.

5.2. Stability analysis via composite Lyapunov functions

By using the functions V and W, it is possible to study the stability properties of the shifted SP-HDS (5.2) with respect to the set

$$\tilde{\mathcal{A}} := \{(x, y) \in \mathbb{R}^n : x \in \mathcal{A},\ y = 0\}, \tag{5.9}$$

which is compact due the compactness of $\mathcal{A}$. To do this, and following similar constructions for ODEs (Saberi & Khalil, 1984), consider the regular and locally Lipschitz function

$$E_\theta(\psi) := (1 - \theta)V(x) + \theta W(x, y), \tag{5.10}$$

where $\theta \in (0, 1)$. Let $\varepsilon^* := \frac{k_x k_y}{k_x k_2 + k_1 k_3}$, $\theta^* = \frac{k_3}{k_1 + k_3}$, where the positive constants $(k_1, k_2, k_3, k_x, k_y)$ were introduced in Assumptions 5.2, 5.3 and 5.5. The next theorem studies UGAS of $\tilde{\mathcal{A}}$ under an additional strong decrease condition of E_{θ^*} during the jumps of (5.2). For completeness, we present the proof in Section 7.3.

Theorem 5.1. *Let $\mathcal{A} \subset \mathbb{R}^{n_1}$ be compact, $\varepsilon \in (0, \varepsilon^*)$, and suppose that Assumption 5.1 holds, and:*

(a) Assumptions 5.2, 5.3 and 5.5 hold, $\varphi_x \in \mathcal{PD}(\mathcal{A})$ and $\varphi_y \in \mathcal{PD}$.

(b) There exists a function $\hat{\rho} \in \mathcal{PD}(\tilde{\mathcal{A}})$ such that

$$E_{\theta^*}(g) - E_{\theta^*}(\psi) \leq -\hat{\rho}(\psi), \tag{5.11}$$

for all $g \in \hat{G}(x, y)$, and all $(x, y) \in \hat{D}$, where E_{θ^*} is given by (5.10).

Then, the set $\tilde{\mathcal{A}}$ is UGAS for the SP-HDS (5.2). □

The following academic example, inspired by Khalil (2002, Ex. 11.12), illustrates the use of Theorem 5.1 in a simple hybrid system with jumps triggered by a timer.

Example 5.1. Consider the following SP-HDS with jumps triggered by a timer $x_2 \in [0, 1]$:

$$(x_1, x_2, z) \in C = C_x \times \mathbb{R},\ C_x := \mathbb{R} \times [0, 1], \quad \begin{cases} \dot{x}_1 = x_1 - x_1^3 + z \\ \dot{x}_2 \in [\underline{\rho}, \overline{\rho}] \\ \varepsilon \dot{z} = -x_1 - z, \end{cases} \tag{5.12a}$$

$$(x_1, x_2, z) \in D = D_x \times \mathbb{R},\ D_x := \mathbb{R} \times \{1\}, \quad \begin{cases} x_1^+ = -a x_1 \\ x_2^+ = 0 \\ z^+ = -bz + c x_1, \end{cases} \tag{5.12b}$$

where $\varepsilon > 0$ is a small parameter, $0 < \underline{\rho} \leq \overline{\rho}$, and $a, b, c \in \mathbb{R}$. Because the rate of evolution of the timer x_2 is bounded from below by $\underline{\rho}$ and from above by $\overline{\rho}$, the solutions to (5.12) experience a jump after at most $1/\underline{\rho}$ units of flow time, but not before $1/\overline{\rho}$ units of flow time. Therefore, solutions to (5.12) are not unique when $\overline{\rho} \neq \underline{\rho}$.

For this system, the quasi-steady state map related to the fast state z is given by $\varphi_0(x) = -x_1$, and the hybrid system (5.2) becomes:

$$(x, y) \in \hat{C} = C, \quad \begin{cases} \dot{x}_1 = -x_1^3 + y \\ \dot{x}_2 \in [\underline{\rho}, \overline{\rho}] \\ \varepsilon \dot{y} = -y, \end{cases} \tag{5.13a}$$

$$(x, y) \in \hat{D} = D, \quad \begin{cases} x_1^+ = -a x_1 \\ x_2^+ = 0 \\ y^+ = -by + (b + c - a)x_1. \end{cases} \tag{5.13b}$$

We study the stability properties of this system with respect to the compact set $\tilde{\mathcal{A}} = \mathcal{A} \times \{0\}$, where $\mathcal{A} = \{0\} \times [0, 1]$. The boundary layer dynamics (5.3) take the form $\dot{y} = -y$, which renders the origin UGES. Indeed, using $W(x, y) = \frac{1}{2}|y|^2$, we obtain:

$$\langle \nabla_y W(x, y), -y \rangle = -|y|^2, \quad \forall (x, y) \in \hat{C},$$

which implies that Assumption 5.2 holds with $k_y = 1$, and $\varphi_y(y) = |y|$.

The reduced dynamics (5.4) are given by the HDS:

$$\tilde{x} \in C_x, \quad \dot{\tilde{x}}_1 = -\tilde{x}_1^3, \quad \dot{\tilde{x}}_2 \in [\underline{\rho}, \overline{\rho}],$$
$$\tilde{x} \in D_x, \quad \tilde{x}_1^+ = -a\tilde{x}_1, \quad \tilde{x}_2^+ = 0.$$

Since $\tilde{x}_2$ is restricted to evolve in the set $[0, 1]$ along every solution, we just need to study the behavior of the state $\tilde{x}_1$ with respect to the origin.

Thus, using $V(\tilde{x}) = \frac{1}{4}\tilde{x}_1^4$, we obtain:

$$\left\langle \nabla V(\tilde{x}), \dot{\tilde{x}} \right\rangle = -\tilde{x}_1^6, \quad \text{and}$$
$$\Delta V(\tilde{x}) := V(\tilde{x}^+) - V(\tilde{x}) = (a^4 - 1)V(\tilde{x}).$$

Thus, Assumption 5.3 holds with $\varphi_x(\tilde{x}) = |\tilde{x}_1|^3$ and $k_x = 1$. If $a \in (0,1)$, then Assumption 5.4 also holds with $\rho_x(\tilde{x}) = V(\tilde{x})$ and $c_x = 1 - a^4$.

Similarly, from (5.13), we have that $\hat{F}_x(x,y) = \{-x_1^3 + y\} \times [\underline{\rho}, \overline{\rho}]$. Therefore, the interconnection conditions of Assumption 5.5 are verified with the inequalities:

$$\left\langle \nabla_y W(x,y), -J\varphi_0(x)\hat{f}_x \right\rangle = y(-x_1^3 + y) \le \varphi_x(x)\varphi_y(y) + \varphi_y(y)^2,$$

for all $\hat{f}_x \in \hat{F}_x(x,y)$, and $\left\langle \nabla V(x), \hat{f}_x - \tilde{f}_x \right\rangle = x_1^3 y \le \varphi_x(x)\varphi_y(y)$, for all $(x,y) \in \hat{C}$, $\hat{f}_x \in \hat{F}(x,y)$, and all $\tilde{f}_x \in \tilde{F}_x(x) = \{-x_1^3\} \times [\underline{\rho}, \overline{\rho}]$. Therefore, item (a) of Theorem 5.1 is verified. To evaluate (5.11) during the jumps of (5.13), we note that $\Delta W(x,y) := W(x^+, y^+) - W(x,y)$ satisfies

$$\Delta W(x,y) = \frac{1}{2}(-by + (b + c - a)x_1)^2 - \frac{1}{2}y^2.$$

If, for instance, $b = (a - c) \in (0, \rho)$, with $\rho < 1$, then

$$\Delta W(x,y) \le -(1 - \rho^2)\frac{y^2}{2} = -c_y W(x,y),$$

where $c_y \in (0,1)$. Since for all $\psi = (x,y) \in \hat{D}$ we have

$$E_{\theta^*}(\psi^+) - E_{\theta^*}(\psi) \le -c_x(1 - \theta^*)V(x) - c_y\theta^* W(x,y),$$

inequality (5.11) holds with $\hat{\rho}(\psi) = \min\{c_x, c_y\}E_{\theta^*}(\psi)$. Thus, by Theorem 5.1, if $\varepsilon \in (0, \frac{1}{2})$ we can conclude that the set $\tilde{\mathcal{A}}$ is UGAS for the SP-HDS (5.13). □

In some applications, it might be difficult to find functions V, W that satisfy all the conditions of Theorem 5.1. In that case, some of the assumptions can be relaxed if certain complete solutions to system (5.1) can be ruled out. In those scenarios, we shall also use the following assumption that captures a non-increasing property on W during the jumps of the system:

Assumption 5.8. There exists $c_y > 0$ and a continuous function $\rho_y \in \mathcal{P}_s\mathcal{D}$ such that the function W of Assumption 5.2 satisfies: $W(g_x, g_y) - W(x,y) \le -c_y\rho_y(y)$, for all $(g_x, g_z) \in \hat{G}(x,y)$, and all $(x,y) \in \hat{D}$. □

Theorem 5.2 below exploits the construction of the composite Lyapunov function (5.10) and the hybrid invariance principle. The proof is presented in Section 7.3

Theorem 5.2. *Let $\varepsilon \in (0, \varepsilon^*)$. Suppose that Assumption 5.1 holds, and:*

(a) Assumptions 5.2, 5.3, and 5.5 hold with $\varphi_x \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$, $\varphi_y \in \mathcal{P}_s\mathcal{D}$.

(b) At least one of the following conditions holds:

(1) Assumptions 5.4 and 5.6 hold with $\rho_x = \rho_4$ and $\frac{k_3 k_4}{k_1} < c_x$.

(2) Assumptions 5.7 and 5.8 hold with $\rho_y = \rho_5$ and $\frac{k_1 k_5}{k_3} < c_y$.

(c) There does not exist a complete solution ψ that remains in a non-zero level set of E_{θ^*}.

Then, the set $\tilde{\mathcal{A}}$ is UGAS for the SP-HDS (5.2). □

Example 5.2 (*Feedback Optimization via Heavy-Ball Systems with Resets*)**.** Consider the following SP-HDS:

$$(x_1, x_2, x_3, z) \in C = C_x \times \mathbb{R}^n \quad C_x := \mathbb{R}^n \times \mathbb{R}^n \times [0, T] \quad \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_2 - M\nabla\phi(z) \\ \dot{x}_3 \in [0,1] \\ \varepsilon\dot{z} = Az + Bx_1, \end{cases} \tag{5.14a}$$

$$(x_1, x_2, x_3, z) \in D = D_x \times \mathbb{R}^n \quad D_x := \mathbb{R}^n \times \mathbb{R}^n \times \{T\} \quad \begin{cases} x_1^+ = x_1 \\ x_2^+ \in [0, \rho]x_2 \\ x_3^+ = 0 \\ z^+ = z, \end{cases} \tag{5.14b}$$

where $M \in \mathbb{R}^{n\times n}$ is invertible, $A \in \mathbb{R}^{n\times n}$ is Hurwitz, $B \in \mathbb{R}^{n\times n}$, and $\rho \in [0,1)$ is a tunable parameter. System (5.14) models a feedback interconnection between a linear plant with state z and input x_1, and a heavy ball optimization algorithm that incorporates resets of x_2 via the jump rule $x_2^+ \in [0,\rho]x_2$ whenever the timer x_3 satisfies $x_3 = T$. Since $\dot{x}_3 \in [0,1]$, system (5.14) can generate different types of behaviors, ranging from solutions that never jump, to solutions that *periodically* reset the state x_3 after T amount of flow time.

For the purpose of analysis, we assume that $\phi : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, radially unbounded, and $\nabla\phi(Hx_1) = 0$ if and only if $x_1 = x^*$, where $H := -A^{-1}B$ is assumed to be nonsingular, and x_1^* is the unique minimizer of the function $x_1 \mapsto \phi(Hx_1)$. Additionally, $\nabla\phi$ is assumed to be globally ℓ-Lipschitz.

In (5.14a), the quasi steady-state map related to the fast state z is given by $\varphi_0(x) = Hx_1$, and the hybrid system (5.2) is given by

$$(x,y) \in \hat{C} = C \quad \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_2 - M\nabla\phi(y + Hx_1) \\ \dot{x}_3 \in [0,1] \\ \varepsilon\dot{y} = Ay + \varepsilon A^{-1}Bx_1, \end{cases} \tag{5.15a}$$

$$(x,y) \in \hat{D} = D \quad \begin{cases} x_1^+ = x_1 \\ x_2^+ \in [0,\rho]x_2 \\ x_3^+ = 0 \\ y^+ = y. \end{cases} \tag{5.15b}$$

We study the stability properties of system (5.15) with respect to the compact set $\tilde{\mathcal{A}} = \mathcal{A} \times \{0\}$ where $\mathcal{A} = \{x_1^*\} \times \{0\} \times [0,T]$. In this case, the boundary layer system (5.3) is the exponentially stable system $\dot{y} = Ay$. Indeed, there exists $P \succ 0$ such that $A^\top P + PA = -I$, and using $W(x,y) = y^\top Py$, we obtain

$$\left\langle \nabla_y W(x,y), Ay \right\rangle \le -|y|^2, \quad \forall\ (x,y) \in \mathbb{R}^{2n} \times \mathbb{R}^n,$$

which implies that Assumption 5.2 holds with $k_y = 1$ and $\varphi_y(y) = |y|$. The reduced HDS (5.4) takes the form

$$\tilde{x} \in C_x, \quad \begin{cases} \dot{\tilde{x}}_1 = \tilde{x}_2 \\ \dot{\tilde{x}}_2 = -\tilde{x}_2 - M\nabla\phi(H\tilde{x}_1) \\ \dot{\tilde{x}}_3 \in [0,1], \end{cases} \tag{5.16a}$$

$$\tilde{x} \in D_x, \quad \begin{cases} \tilde{x}_1^+ = \tilde{x}_1 \\ \tilde{x}_2^+ \in [0,\rho]\tilde{x}_2 \\ \tilde{x}_3^+ = 0. \end{cases} \tag{5.16b}$$

We can analyze (5.16) using the function $V(\tilde{x}) = \phi(H\tilde{x}_1) - \phi(Hx_1^*) + \frac{1}{2}|\tilde{x}_2|^2$, which satisfies Assumption 5.3-(1) since, ϕ is radially unbounded, H is non-singular, and $\phi(H\tilde{x}_1)$ attains its minimum at x_1^*. Moreover, setting $M = H^\top$ leads to:

$$\begin{aligned} \left\langle \nabla V(\tilde{x}), \tilde{f}_x \right\rangle &= \left\langle \left(H^\top\nabla\phi(H\tilde{x}_1), \tilde{x}_2\right), \left(\tilde{x}_2, -\tilde{x}_2 - M\nabla\phi(H\tilde{x}_1)\right) \right\rangle \\ &= -|\tilde{x}_2|^2, \end{aligned}$$

for all $\tilde{f}_x \in \tilde{F}_x(\tilde{x}) = \left\{\left(\tilde{x}_2, -\tilde{x}_2 - M\nabla\phi(H\tilde{x}_1)\right)\right\} \times [0,1]$, thus verifying Assumption 5.3-(2) with $k_x = 1$ and $\varphi_x(\tilde{x}) := |\tilde{x}_2|$. Additionally, the inequalities

$$\left\langle Py, A^{-1}Bx_2 \right\rangle \le \frac{|P \parallel B|}{|A|}\varphi_y(y)\varphi_x(x)$$
$$\left\langle x_2, M\left(\nabla\phi(Hx_1 + y) - \nabla\phi(Hx_1)\right) \right\rangle \le \ell|M|\varphi_x(x)\varphi_y(y),$$

imply that the interconnection conditions of Assumption 5.5 are verified with $k_1 = \overline{\sigma}_P\overline{\sigma}_B/\underline{\sigma}_A$, $k_2 > 0$, and $k_3 = \ell\overline{\sigma}_M$, where $\overline{\sigma}_\Psi$ and $\underline{\sigma}_\Psi$ denote the maximum singular value, and the minimum singular value, respectively, of a matrix Ψ. Since $\varphi_x \in \mathcal{P}_s\mathcal{D}(\mathcal{A})$ and $\varphi_y \in \mathcal{P}_s\mathcal{D}$, item (a) of Theorem 5.2 is satisfied.

Next, we show that condition (b)-(1) of Theorem 5.2 also holds. To verify Assumption 5.4, we evaluate the change of V during the jumps (5.16b): $\Delta V(x) = V(x^+) - V(x) = -c_x\varphi_x(x)^2$, with $c_x = \frac{1}{2}(1 - \rho^2)$. To

verify Assumption 5.6, we evaluate the change of W during the jumps of the original hybrid dynamics (5.14), leading to $\Delta W = W(x_1^+, y^+) - W(x_1, y) = 0 \leq k_4 \varphi_x(x)^2$, where k_4 is any positive number that satisfies $k_4 < (1-\rho^2)\overline{\sigma}_P \overline{\sigma}_B / 2\ell \overline{\sigma}_M \underline{\sigma}_A$.

Finally, we verify item (c) of Theorem 5.2. Since item (a) holds and $\varepsilon \in (0, \varepsilon^*)$, there exists $\lambda > 0$ such that $\dot{E}_{\theta^*} \leq -\lambda(\varphi_x(x)^2 + \varphi_y(y)^2) \leq 0$, for all $(x, z) \in \mathbb{R}^{2n} \times \mathbb{R}^n$. Since item (b) holds, the change of E_{θ^*} during jumps satisfies $\Delta E_{\theta^*} = E_{\theta^*}(\psi^+) - E_{\theta^*}(\psi) \leq c\varphi_x(x)^2 \leq 0$, $c > 0$, for all $\psi = (x, y) \in \mathbb{R}^{2n} \times \mathbb{R}^n$. We proceed to show that no complete solution can keep E_{θ^*} equal to a non-zero constant. In particular, since E_{θ^*} does not increase during jumps, and since every jump is separated by intervals of flow with length lower bounded by T, it suffices to study the behavior of $\dot{E}_{\theta^*}$. Since $\dot{E}_{\theta^*} = 0$ whenever $\varphi_x(x) = 0$ and $\varphi_y(y) = 0$, we study the behavior of the trajectories of (5.14) in the Kernel of these functions:

$$U = \left\{(x, y) \in \mathbb{R}^{2n} \times [0, T] \times \mathbb{R}^n : x_2 = 0,\ y = 0\right\},$$

which is invariant under the jump map. Indeed, for any trajectory that remains in U during flows, we must have $\dot{x}_2 = 0$ and $\nabla\phi(y + Hx_1) = \nabla\phi(Hx_1) = 0$. But, by assumption, the latter condition can only occur if $x_1 = x_1^*$. It follows that the only non-empty set where $E_{\theta^*}(x, y)$ can remain constant along the trajectories of the system is precisely the set $\tilde{\mathcal{A}}$, which satisfies $E_{\theta^*}(\tilde{\mathcal{A}}) = 0$. By Theorem 5.2, we can conclude UGAS of $\tilde{\mathcal{A}}$ for the SP-HDS (5.15). □

The procedure followed in Example 5.2 is similar to the one used in the literature of continuous-time feedback optimization (Bianchin, Poveda et al., 2022; Colombino et al., 2020; Cothren, Bianchin, & Dall'Anese, 2022; Hauswirth et al., 2020), where the fast dynamics of z model the dynamics of the plant, and the slow dynamics of x capture the dynamics of an optimization algorithm using real-time output feedback to steer the plant towards an optimal steady-state point. Such techniques have found fruitful applications in power systems (Colombino et al., 2020; Hauswirth et al., 2020), transportation systems (Bianchin, Cortes et al., 2022), and supply chain management (Belgioioso et al., 2022). Example 5.2 shows that similar hybrid multi-time scale feedback architectures could be considered. We finish this section by highlighting that global stability results for continuous-time systems have also been established via contraction theory (Del Vecchio & Slotine, 2012). However, for SP-HDS, such results remain unexplored.

5.2.1. Relaxations for semi-global practical stability

The assumptions required in Theorems 5.1–5.2 can be significantly relaxed in order to establish local or semi-global stability results in well-posed HDS (Sanfelice & Teel, 2011). In particular, while Tikhonov's-type of results (Khalil, 2002, Thm. 11.1), Deghat, Ahmadizadeh, Nešić, and Manzie (2021), are difficult to establish in HDS (for which solutions might not even be unique), the following closeness of solutions property (specialized to singled-valued dynamics) was established in Wang et al. (2012a, Ex. 1 & Thm. 1) for a general class of (first-order) singularly perturbed HDS of the form (5.1).

Proposition 5.3. *Consider the SP-HDS (5.1), and suppose that F_x, F_z, G, and φ_0 are continuous, C_x, D_x are closed, $\Psi = C_z = D_z$ is compact, and that for each compact set $K_0 \subset \mathbb{R}^{n_1}$ the reduced system has no solutions with finite escape times. Then, for each $\epsilon > 0$ and any pair $T, J > 0$ there exists ε^* such that for all $\varepsilon \in (0, \varepsilon^*)$ and all solutions x to (5.1) with $x(0,0) \in K_0$, there exists some solution $\tilde{x}$ of the reduced system with $\tilde{x}(0,0) \in K_0$ such that:*

- *for each $(t, j) \in dom(x)$ with $t \leq T$ and $j \leq J$ there exists t' such that $(t', j) \in dom(\tilde{x})$, with $|t - t'| \leq \epsilon$ and $|x(t, j) - \tilde{x}(t', j)| \leq \epsilon$.*
- *for each $(t, j) \in dom(\tilde{x})$ with $t \leq T$ and $j \leq J$ there exists t' such that $(t', j) \in dom(x)$, with $|t - t'| \leq \epsilon$ and $|\tilde{x}(t, j) - x(t', j)| \leq \epsilon$.*

Proposition 5.3 says that every solution x of the slow dynamics of the SP-HDS (5.1) can be made arbitrarily close, on compact time domains and in a graphical sense, to *some* solution of its reduced hybrid system (5.4), provided the reduced system is forward complete; see Fig. 5 for an illustration. The fact that for SP-HDS the closeness of solutions property is studied in a graphical sense instead of using the standard 2-norm is a key difference with respect to the Lipschitz continuous ODE case studied in the literature (Khalil, 2002) and in Sections 3–4. To the best knowledge of the authors, results on closeness of solutions for *second-order* SP-HDS are still absent in the literature.

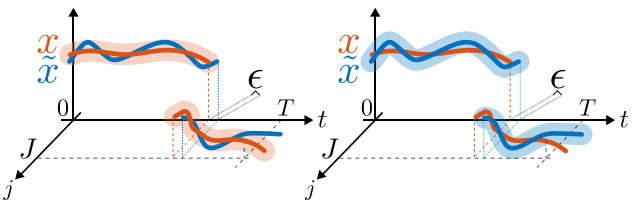

Fig. 5. (T, J)−closeness of a solution x and a solution $\tilde{x}$ of the reduced hybrid system.

If the Assumptions of Proposition 5.3 hold, and the reduced HDS has a compact set $\mathcal{A} \subset \mathbb{R}^{n_1}$ that is UGAS (in the sense of Definition 2.2), then it can be established that the original SP-HDS (5.1) renders the set $\mathcal{A} \times \Psi$ semi-globally practically asymptotically stable (SGpAS) as $\varepsilon \to 0^+$ (Wang et al., 2012a, Thm. 2), (Sanfelice & Teel, 2011, Thm. 1), namely, there exists $\beta \in \mathcal{KL}$ such that for each compact set $K_0 \subset \mathbb{R}^{n_1}$ and each $\nu > 0$, there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$, every solution to (5.1) with $x(0,0) \in K_0$ satisfies the bound

$$|x(t, j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t + j) + \nu, \tag{5.17}$$

for all $(t, j) \in \text{dom}(x, z)$. When ε^* is independent of K_0 and (5.17) holds for all initial conditions, the set $\mathcal{A}$ is said to be uniformly globally practically asymptotically stable (UGpAS). If $\beta(r, s) = c_1 r e^{-c_2 s}$, we use the acronyms SGpES and UGpES.

The above property actually applies to a more general class of HDS (including ODEs Teel et al., 2003) for which the reduced system is defined via averaging theory Wang et al. (2012a), which is the main subject of the next section.

6. Averaging theory in hybrid dynamical systems

In this section, we focus on a specific subset of HDS (2.1), whose stability properties can be effectively analyzed using averaging theory. The dynamics considered in this context usually involve oscillating or time-varying vector fields that exhibit well-defined average mappings as the frequency of the time variation increases, similar to those considered in Sections 3–4. However, for HDS we restrict our attention to *first-order* averaging. In particular, we study HDS for which we can establish global stability results by imposing suitable smoothness properties and ensuring the (exponential) stability of their (first-order) average dynamics. Such results are natural extensions of existing stability results for ODEs, Sastry and Bodson (1989, Ch.4). We illustrate these findings through various examples, including time-varying switched systems, sampled-data systems, and parameter estimation algorithms with momentum and resets. Subsequently, we consider cases where semi-global practical asymptotic stability results can be obtained under weaker assumptions. Such cases are illustrated via novel applications in the context of hybrid source-seeking problems, dealing with issues like spoofing and intermittent feedback. We also explore new hybrid switched extremum-seeking algorithms and hybrid model-free controllers for the stabilization of systems with unknown control directions.

6.1. Model and main assumptions

Consider a HDS with states $\psi = (x, q) \in \mathbb{R}^n \times \mathbb{R}^m$, $\tau \in \mathbb{R}_{\geq 0}$, and dynamics:

$$(x, q, \tau) \in C := \mathbb{R}^n \times C_q \times \mathbb{R}_{\geq 0}, \quad \begin{cases} \dot{x} = f(x, q, \tau, \varepsilon) \\ \dot{q} \in F_q(q) \\ \dot{\tau} = \dfrac{1}{\varepsilon}, \end{cases} \tag{6.1a}$$

$$(x, q, \tau) \in D := \mathbb{R}^n \times D_q \times \mathbb{R}_{\geq 0}, \quad \begin{cases} x^+ = g(x, q) \\ q^+ \in G_q(q) \\ \tau^+ = \tau \end{cases} \tag{6.1b}$$

where $\varepsilon > 0$ is a small parameter, $C_q, D_q \subset \mathbb{R}^m$, x is the main state, q is an auxiliary state that can be used to model logic modes, timers, oscillators, etc, and τ is used to model the fast variations of f. We make the following regularity assumptions on (6.1a)–(6.1b). These assumptions are natural extensions of those considered in the literature of averaging for ODEs whenever *global* results (as opposed to local, or semi-global) are sought-after, e.g., Sastry and Bodson (1989, Ch. 4).

Assumption 6.1. There exists $\varepsilon_0 > 0$ such that:

(a) The function f is C^1, and g is continuous. The mapping $F_q : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is OSC, LB, and convex-valued relative to C_q; The mapping $G_q : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is OSC and LB relative to D_q; the sets C_q and D_q are compact, and satisfy $C_q \subset \text{dom } F_q$, $D_q \subset \text{dom } G_q$.

(b) The function f satisfies $f(0, q, \tau, \varepsilon) = 0$ for all $(q, \tau, \varepsilon) \in (C_q \cup D_q) \times \mathbb{R}_{\geq 0} \times [0, \varepsilon_0)$. The function g satisfies $g(0, q) = 0$ for all $q \in D_q$.

(c) There exists $L_x > 0$, such that

$$|f(x_1, q, \tau, \varepsilon) - f(x_2, q, \tau, \varepsilon)| \leq L_x |x_1 - x_2|, \tag{6.2}$$

for all $q \in C_q$, all $x_1, x_2 \in \mathbb{R}^n$, and all $\varepsilon \in [0, \varepsilon_0)$.

(d) There exists $L_{\varepsilon^*} > 0$, such that

$$|f(x, q, \tau, \varepsilon_1) - f(x, q, \tau, \varepsilon_2)| \leq L_{\varepsilon_0} |x| |\varepsilon_1 - \varepsilon_2|, \tag{6.3}$$

for all $q \in C_q$, all $x \in \mathbb{R}^n$, and all $\varepsilon_1, \varepsilon_2 \in (0, \varepsilon_0)$.

(e) There exists $L_g > 0$ such that

$$|g(x_1, q) - g(x_2, q)| \leq L_g |x_1 - x_2|, \tag{6.4}$$

for all $q \in D_q$, all $x_1, x_2 \in \mathbb{R}^n$.

(f) The mapping $G(x, q) := g(x, q) \times G_q(q)$ satisfies $G(\mathbb{R}^n \times D_q) \subset \mathbb{R}^n \times (C_q \cup D_q)$. □

Conditions (6.2), (6.3), and (6.4) can be relaxed to local Lipschitz continuity (uniform over q) whenever local or semi-global stability results are of interest, or when x in (6.1a)–(6.1b) is restricted to evolve within a compact set.

To study the stability properties of the HDS (6.1a)–(6.1b), we introduce the (first-order) average map of f.

Definition 6.1. f is said to have an *average map* $\bar{f}$ if there exists a class-$\mathcal{L}$ function γ such that

$$\left| \frac{1}{T} \int_\tau^{\tau+T} \Big(f(x, q, s, 0) - \bar{f}(x, q) \Big) \mathrm{d}s \right| \leq \gamma(T) |x|, \tag{6.5}$$

for all $x \in \mathbb{R}^n$, all $q \in C_q \cup D_q$, and all $\tau, T \in \mathbb{R}_{\geq 0}$. The function $\gamma(\cdot)$ is called the *convergence function*. □

As in Section 3, the average map is defined with respect to the flow map of (6.1a). If f is periodic in τ, one can take $\gamma(T) = \frac{1}{1+T}$ (Khalil, 2002, Ex. 10.12). Other choices of convergence functions are discussed in Anderson et al. (1986), Khalil (2002, Ch. 10) and Sastry and Bodson (1989, Ch.4).

Let $d(x, q, \tau) := f(x, q, \tau, 0) - \bar{f}(x, q)$. The following smoothness assumption on d and $\bar{f}$, uniformly on $C_q \cup D_q$, will be instrumental in our analysis.

Assumption 6.2. The function f admits a C^1 average map $\bar{f}$, with convergence function γ, and there exists $L_{\text{ave}} > 0$ such that:

$$|\bar{f}(x_1, q) - \bar{f}(x_2, q)| \leq L_{\text{ave}} |x_1 - x_2|,$$

for all $x_1, x_2 \in \mathbb{R}^n$, and all $q \in C_q$. Moreover,

$$\left| \frac{1}{T} \int_\tau^{\tau+T} J_\psi d(x, q, \tau) \mathrm{d}\tau \right| \leq \gamma(T), \tag{6.6}$$

for all $\psi = (x, q) \in \mathbb{R}^n \times (C_q \cup D_q)$ and all $\tau, T \in \mathbb{R}_{\geq 0}$. □

Next, we introduce the average HDS associated to system (6.1a)–(6.1b).

Definition 6.2. The average HDS of system (6.1a)–(6.1b) has state $\bar{\psi} = (\bar{x}, \bar{q}) \in \mathbb{R}^n \times C_q \cup D_q$, and dynamics

$$\bar{\psi} \in \mathbb{R}^n \times C_q, \quad \dot{\bar{\psi}} \in \bar{F}(\bar{\psi}) := \begin{pmatrix} \bar{f}(\bar{x}, \bar{q}) \\ F_q(\bar{q}) \end{pmatrix} \tag{6.7a}$$

$$\bar{\psi} \in \mathbb{R}^n \times D_q, \quad \bar{\psi}^+ \in \bar{G}(\bar{\psi}), \tag{6.7b}$$

where the average jump map is defined as $\bar{G}(\bar{\psi}) := G(\bar{x}, \bar{q})$, for all $\bar{\psi} \in (\mathbb{R}^n \times D_q)$. □

By Assumptions 6.1 and 6.2, the average hybrid dynamics (6.7) satisfy the hybrid basic conditions of Assumption 2.1. To characterize the stability properties of the average dynamics, we use the following assumption, which is an extension to hybrid systems of the assumptions made in Sastry and Bodson (1989, Thm. 4.2.5) for ODEs. For simplicity, we will study the stability properties of the set $\mathcal{A} := \{0\} \times (C_q \cup D_q)$.

Assumption 6.3. There exists a C^1 function $V : \mathbb{R}^n \times (C_q \cup D_q) \to \mathbb{R}_{\geq 0}$, $p > 1$, and $c_i > 0$, for $i \in \{1, 2, \dots, 5\}$, such that:

(a) $c_1 |\bar{\psi}|_{\mathcal{A}}^p \leq V(\bar{\psi}) \leq c_2 |\bar{\psi}|_{\mathcal{A}}^p$, for all $\bar{\psi} \in \mathbb{R}^n \times (C_q \cup D_q)$.

(b) $\langle \nabla V(\bar{\psi}), \bar{f} \rangle \leq -c_4 V(\bar{\psi})$, and $|\nabla V(\bar{\psi})| \leq c_3 |\bar{\psi}|_{\mathcal{A}}^{p-1}$, for all $\bar{\psi} \in \mathbb{R}^n \times C_q$, and all $\bar{f} \in \bar{F}(\bar{\psi})$.

(c) $V(\bar{g}) \leq c_5 V(\bar{\psi})$, for all $\bar{\psi} \in \mathbb{R}^n \times D_q$, and all $\bar{g} \in \bar{G}(\bar{\psi})$, where $c_5 \in (0, 1)$ satisfies $2^{1-p} > \frac{c_2}{c_1} c_5 =: \lambda$. □

By Teel, Forni, and Zaccarian (2013, Thm. 1), Assumption 6.3 guarantees that the set $\mathcal{A}$ is UGES for the averaged HDS (6.7).

6.2. Stability properties and examples

The following theorem establishes that the original hybrid dynamics (6.1) preserve the stability properties of the average hybrid dynamics (6.7) whenever ε is sufficiently small. For completeness, the proof is presented in Section 7.4.

Theorem 6.1. *Suppose that Assumptions 6.1–6.3 hold. Then:*

(a) For each $\nu > 0$ there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$, all $\psi(0, 0), \tau(0, 0) \in \mathbb{R}^n \times (C_q \cup D_q) \times \mathbb{R}_{\geq 0}$, all solutions (ψ, τ) to system (6.1) satisfy the bound:

$$|\psi(t, j)|_{\mathcal{A}} \leq \kappa_1 |\psi(0, 0)|_{\mathcal{A}} e^{-\kappa_2 (t+j)} + \nu, \tag{6.8}$$

for all $(t, j) \in \text{dom}(\psi, \tau)$, where $\kappa_1, \kappa_2 > 0$.

(b) If, additionally, $J_q f(x, q, \tau, \varepsilon) \dot{q} = 0$ for all $(x, q, \tau, \varepsilon) \in \mathbb{R}^n \times C_q \times \mathbb{R}_{\geq 0} \times [0, \varepsilon_0]$, then there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ all solutions (ψ, τ) to system (6.1) satisfy the bound (6.8) with $\nu = 0$. □

The following example illustrates the application of Theorem 6.1 in a class of switched systems with linear time-varying mappings.

Example 6.1 (*Switched Systems with Resets and Time-Varying Vector Fields*)**.** Consider a switching system with resets, of the form $\dot{x} = A_\sigma(\tau) x$,

$x^+ = \rho x$, where $\sigma : \mathbb{R}_{\geq 0} \to Q = \{1,2,3\}$ is a switching signal, $\dot{\tau} = \frac{1}{\varepsilon}$, $\varepsilon > 0$ is a small parameter, $\rho \in (0, \frac{1}{2})$ and

$$A_1(\tau) = \begin{bmatrix} -0.1 + \sin(\tau) & -1 \\ 4\sin(\tau)^2 & -0.2\cos(\tau)^2 \end{bmatrix},$$
$$A_2(\tau) = \begin{bmatrix} -0.1 & 2 + 5\sin(\tau) \\ -1 + \cos(\tau) & -0.1 \end{bmatrix},$$
$$A_3(\tau) = \begin{bmatrix} 0.2\sin^2(\tau) & 0.2 \\ -0.1 & 0.3 + \sin(\tau) \end{bmatrix}.$$

To analyze this system via averaging theory, we model the switching system as a HDS with continuous-time dynamics:

$$\dot{x} = A_{q_1}(\tau)x, \quad \dot{\tau} = \frac{1}{\varepsilon}, \tag{6.9}$$

where $q_1 \in Q$ is now a logic state generated by the following hybrid automaton:

$$\dot{q}_1 = 0, \quad \dot{q}_2 \in \left[0, \frac{1}{\tau_d}\right], \quad \dot{q}_3 \in \left[0, \frac{1}{\tau_a}\right] - \mathbb{I}_{Q_u}(q_1). \tag{6.10}$$

where $Q_u := \{3\}$. The dynamics (6.9)–(6.10) evolve in the flow set $\left(x, (q_1, q_2, q_3), \tau\right) \in \mathbb{R}^2 \times C_q \times \mathbb{R}_{\geq 0}$, where $C_q = Q \times [0, N] \times [0, T]$, $N \geq 1$, $T \geq 0$. The discrete-time dynamics of the HDS are given by:

$$x^+ = \rho x, \quad q_1^+ \in Q \backslash \{q_1\}, \quad q_2^+ = q_2 - 1, \quad q_3^+ = q_3, \tag{6.11}$$

which evolve in the jump set $(x, (q_1, q_2, q_3), \tau) \in \mathbb{R}^2 \times D_q \times \mathbb{R}_{\geq 0}$, where $D_q = Q \times [1, N] \times [0, T]$. Solutions to system (6.9)–(6.11) have the property that for any $t_2 \geq t_1$ the number of switches of the signal q_1 in the interval $[t_1, t_2)$ is bounded above as $N_{t_2,t_1} \leq \frac{1}{\tau_d}(t_2 - t_1) + N_0$ (Cai, Teel, & Goebel, 2008, Prop. 1.1). Additionally, every solution to (6.10) also satisfies the average activation-time constraint $A_{t_2,t_1} \leq \frac{1}{\tau_a}(t_2 - t_1) + T_0$ (Poveda & Teel, 2017a, Lemma 7), where A_{t_2,t_1} corresponds to the total amount of time that the signal q_1 satisfies $q_1 = 3$ during the interval $[t_1, t_2)$, i.e., $\int_{t_1}^{t_2} \mathbb{I}_{Q_u}(q_1(s, j(s)))ds$.

Computing the average of (6.9) leads to the average matrices $\bar{A}_1 = [-0.1, -1; 2, -0.1] = \bar{A}_2^\top$, $\bar{A}_3 = [0.1, 0.2; -0.1, 0.3]$, and since $\bar{A}_1$ and $\bar{A}_2$ are Hurwitz, following the procedure of Galarza, Poveda, Bianchi, and Dallenese (2021, Lemma 6) there exist constants $\tau_d, \tau_a > 0$ such that when ρ is sufficiently small the average HDS

$$\dot{\bar{x}} = \bar{A}_{q_1}\bar{x}, \ \dot{\bar{q}}_1 = 0, \ \dot{\bar{q}}_2 \in \left[0, \frac{1}{\tau_d}\right], \ \dot{\bar{q}}_3 \in \left[0, \frac{1}{\tau_a}\right] - \mathbb{I}_{Q_u}(\bar{q}_1),$$
$$\bar{x}^+ = \rho\bar{x}, \ \ \bar{q}_1^+ \in Q\backslash\{\bar{q}_1\}, \ \ \bar{q}_2^+ = \bar{q}_2 - 1, \quad \bar{q}_3^+ = \bar{q}_3,$$

renders UGES the set $\mathcal{A} = \{0\} \times Q \times [0, N] \times [0, T] = \{0\} \times (C_q \cup D_q)$ with a Lyapunov function V satisfying Assumption 6.3. Since the original and the average HDS satisfy Assumptions 6.1–6.3, by Theorem 6.1 we can conclude there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ the solutions of the HDS (6.9)–(6.11) satisfy a bound of the form (6.8). □

The stability of switched systems via averaging theory has been considered in Liberzon (2003), Mostacciuolo, Trenn, and Vasca (2017), and Wang and Nesic (2010), although most of the results in the literature usually consider average systems obtained by averaging along the switching signal q_1. Instead, in (6.9), both the state x and the switching signal evolve in a slower time scale compared to the fast variations induced by τ.

In the following example, we study the application of averaging theory to sampled-data systems with oscillating plants.

Example 6.2 (*Sampled-Data Systems with Oscillatory Plant Dynamics*)**.** Consider a linear time-varying plant with dynamics $\dot{x}_1 = A(\tau)x_1 + Bx_2$, where $B \in \mathbb{R}^{n_1 \times n_2}$, $A : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_1 \times n_1}$ is a continuous time-varying matrix, and $x_2 \in \mathbb{R}^{n_2}$ is the control input. To stabilize the plant to $x_1 = 0$, we consider a sampled-data structure with control dynamics $x_2^+ = Kx_1$, $K \in \mathbb{R}^{n_2 \times n_1}$, interconnected with the plant via a zeroth-order hold device with a periodic resetting timer. The resulting closed-loop system can be modeled as a HDS of the form (6.1a)–(6.1b), given by

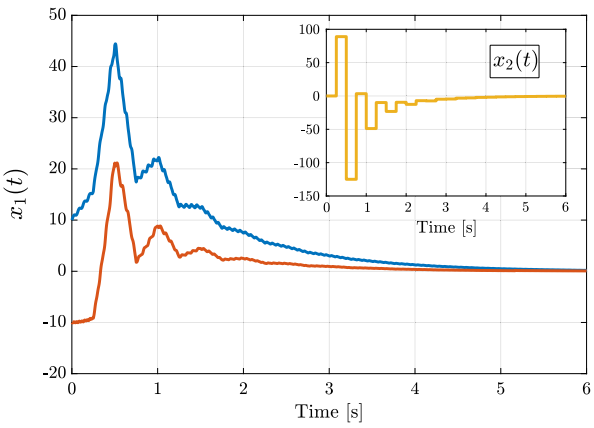

Fig. 6. Trajectories of a sampled-data system with oscillatory plant dynamics.

$$\dot{x}_1 = A(\tau)x_1 + Bx_2, \quad \dot{x}_2 = 0, \quad \dot{q} = 1, \quad \dot{\tau} = \frac{1}{\varepsilon} \tag{6.12a}$$
$$x_1^+ = x_1, \qquad x_2^+ = Kx_1, \qquad q^+ = 0, \qquad \tau^+ = \tau, \tag{6.12b}$$

with flow set $C = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times C_q \times \mathbb{R}_{\geq 0}$, jump set $D = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times D_q \times \mathbb{R}_{\geq 0}$, $C_q = [0, T]$, $D_q = \{T\}$. The average hybrid dynamics of system (6.12) are given by:

$$\dot{\bar{x}}_1 = \bar{A}\bar{x}_1 + B\bar{x}_2, \quad \dot{\bar{x}}_2 = 0, \qquad \dot{\bar{q}} = 1,$$
$$\bar{x}_1^+ = \bar{x}_1, \qquad \bar{x}_2^+ = K\bar{x}_1, \qquad \bar{q}^+ = 0,$$

with flow set $\mathbb{R}^n \times \mathbb{R}^m \times C_q$ and jump set $\mathbb{R}^n \times \mathbb{R}^m \times D_q$. Assuming that: (a) the matrix $\bar{A} = \lim_{T' \to \infty} \frac{1}{T'} \int_t^{t+T'} A(s)ds$ is well-defined; and (b) there exists $T > 0$ and $P \succ 0$ such that $M^\top PM - P \prec -\rho I$, where $\rho > 0$, $M = \exp(FT)J$, $F = [\bar{A}, B; 0, 0]$ and $J = [I, 0; K, 0]$, it follows that Assumptions 6.1 and 6.2 hold, and by Goebel et al. (2012, pp. 57) the average hybrid dynamics satisfy Assumption 6.3. Consequently, by Theorem 6.1, there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ the solutions of the HDS (6.12) satisfy a bound of the form (6.8).

To numerically illustrate this result, consider the matrix $A(\tau) = 2[\cos(\omega\tau)^2 + \cos(\omega\tau), -e^{-\tau}\sin(\omega\tau) - \sin(\omega\tau)^2; e^{-\tau}\sin(\omega\tau) + \sin(\omega\tau)^2, \cos(\omega\tau)^2 + \sin(\omega t\tau)]$, $\omega = 100$, and $B = [1; 1]$. This matrix-valued function admits a well-defined average $\bar{A} = [1, -1; 1, 1]$. By choosing $P = [25.25, -51.97, 0; -51.97, 162.54, 0; 0, 0, 1] \succ 0$, $K = [1; -8]$, and $T = 0.25$, it follows that $M^\top PM - P = -I \prec 0$, ensuring that the closed-loop system satisfies the above assumptions. Simulated trajectories of the closed-loop system are shown in Fig. 6, showing the convergence of the plant's state x_1 to the origin under the action of the input x_2. □

Averaging theory is commonly used in adaptive systems to study stability under suitable excitation conditions on signals of interest (Anderson et al., 1986; Sastry & Bodson, 1989). Adaptive dynamics with resets (e.g., covariance resetting) have also been studied to prevent instabilities in least-squares parameter estimation problems (Ioannou & Sun, 2012). Motivated by some of these results, as well as by recent accelerated optimization and estimation dynamics studied in the literature of machine learning (Wilson, Recht, & Jordan, 2021), the following example studies a higher-order gradient-based parameter estimation algorithm with momentum, similar to those studied in Gaudio, Annaswamy, Bolender, Lavretsky, and Gibson (2020), Le and Teel (2022) and Moreu and Annaswamy (2021), but implementing resets to improve transient performance. In contrast to existing results in the literature of concurrent learning (Chowdhary & Johnson, 2010; Ochoa et al., 2021), we do not assume the existence of "sufficiently rich" past recorded data.

Example 6.3 (*Hybrid Parameter Identification with Momentum and Resets*)**.** Consider a standard linear regression problem where the goal is

to estimate a parameter $\theta^* \in \mathbb{R}^n$ using real-time measurements of the noisy signal

$$y(\tau) = \xi(\tau)^\top \theta^* + \eta(\tau),$$

where $\xi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is a known regressor that is assumed to be continuous, uniformly bounded, and persistently exciting (PE) (Sastry & Bodson, 1989, pp. 72), with time-variation due to the state $\dot{\tau} = \frac{1}{\varepsilon}$, and where the noise is assumed to satisfy $|\eta(\tau)| \leq \bar{\eta}$ for all $\tau \geq 0$. A typical approach to estimating θ^* is to use the estimation error $e := \xi(\tau)^\top \theta - y$ and the gradient algorithm $\dot{\theta} = -\frac{k}{2}\nabla_\theta e^2$, $k > 0$, which, in the error coordinates $x_1 = \theta - \theta^*$, can be written as $\dot{x}_1 = -k\xi(\tau)\xi(\tau)^\top x_1 + k\xi(\tau)\eta(\tau)$. When $\eta = 0$, the gradient algorithm guarantees exponential convergence to the true parameter θ^* (Praly, 2016). When $\eta \neq 0$, the algorithm achieves convergence to a neighborhood of the true parameter. To improve transient performance, we can consider a higher-order hybrid gradient algorithm that incorporates momentum during the flows and resets during the jumps, see Ochoa et al. (2021) for similar hybrid algorithms that use past recorded values of ξ instead of PE time-varying regressors. In particular, when $\eta = 0$, we consider a hybrid system with flow set:

$$(\theta, p, q, \tau) \in C = \mathbb{R}^n \times \mathbb{R}^n \times C_q \times \mathbb{R}_{\geq 0}, \quad C_q = [0, T],$$

and continuous-time dynamics given by:

$$\dot{\theta} = k\frac{(p-\theta)}{\gamma(q)}, \quad \dot{p} = -k\gamma(q)\xi(\tau)e(\tau), \quad \dot{q} = \rho, \quad \dot{\tau} = 1,$$

with $\rho \in (0, 1)$, and $\gamma(q) = q_0 + q$, with $q_0 > 0$. The jump set is defined as:

$$(\theta, p, q, \tau) \in D = \mathbb{R}^n \times \mathbb{R}^n \times D_q \times \mathbb{R}_{\geq 0}, \quad D_q = \{T\},$$

where $T > 0$, and the discrete-time dynamics are given by:

$$\theta^+ = \theta, \quad p^+ = \theta, \quad q^+ = 0, \quad \tau^+ = \tau.$$

Using the error coordinates $x_1 = \theta - \theta^*$ and $x_2 = p - \theta^*$, the continuous-time dynamics of the hybrid system can be written as (6.1):

$$\dot{x} = f(x, q, \tau) = \begin{pmatrix} \frac{k}{\gamma(q)}(x_2 - x_1) \\ -k\gamma(q)\xi(\tau)\xi(\tau)^\top x_1 \end{pmatrix}, \tag{6.13a}$$

$$\dot{q} = \rho, \quad \dot{\tau} = \frac{1}{\varepsilon}, \tag{6.13b}$$

with discrete-time dynamics

$$x_1^+ = x_1, \quad x_2^+ = x_1, \quad q^+ = 0, \quad \tau^+ = \tau, \tag{6.13c}$$

To compute the average map $\bar{f}$ along τ, it suffices to compute the average of the vector field $f_2(x, q, \tau) = -\gamma(q)\xi(\tau)\xi(\tau)^\top x_1$. Indeed, by Sastry and Bodson (1989, Sec. 4.3), the average of this mapping with respect to τ is $\bar{f}_2(x, q) = -\gamma(q)R_\xi(0)x_1$, where $R_\xi(0)$ is the auto-covariance matrix of $\xi(\cdot)$ evaluated at 0. Therefore, the average system has continuous-time dynamics

$$\dot{\bar{x}} = \begin{pmatrix} \frac{k}{\gamma(\bar{q})}(\bar{x}_2 - \bar{x}_1) \\ -k\gamma(\bar{q})R_\xi(0)\bar{x}_1 \end{pmatrix}, \quad \dot{\bar{q}} = \rho, \quad (\bar{x}, \bar{q}) \in \mathbb{R}^{2n} \times C_q.$$

and discrete-time dynamics:

$$\bar{x}_1^+ = \bar{x}_1, \quad \bar{x}_2^+ = \bar{x}_1, \quad \bar{q}^+ = 0, \quad (\bar{x}, \bar{q}) \in \mathbb{R}^{2n} \times D_q.$$

By Sastry and Bodson (1989, Prop. 2.7.1), $R_\xi(0)$ is positive definite if and only if ξ is PE. Therefore, under the PE assumption on ξ, the Lyapunov function $V(\bar{x}, \bar{q}) = \frac{1}{4}|\bar{x}_2 - \bar{x}_1|^2 + \frac{1}{4}|\bar{x}_2|^2 + \bar{q}^2\bar{x}_1^\top R_\xi(0)\bar{x}_1$ can be used to study the stability of the set $\mathcal{A} = \{0\} \times \{0\} \times [0, T]$ for the average hybrid dynamics. Indeed, by Poveda and Li (2021, Lemma 7.4), and using $k = 2$, $\rho = \frac{1}{2}$, all the conditions of Assumption 6.3 are satisfied, provided $T^2 - q_0^2 \geq \frac{1}{2\lambda_{\min}(R_\xi(0))}$, where $\lambda_{\min}(R_\xi(0))$ is the smallest eigenvalue of the matrix $R_\xi(0)$. By Theorem 6.1, there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ the solutions of the HDS (6.13) satisfy a bound of the form (6.8).

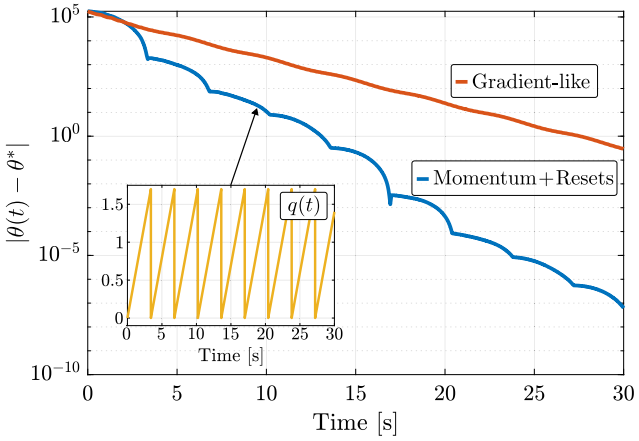

Fig. 7. Error-trajectories generated by the standard gradient estimation dynamics and the hybrid dynamics with momentum and resets.

Fig. 7 compares error trajectories of two parameter estimation dynamics: the standard gradient algorithm and the proposed hybrid dynamics with momentum and resets. Using the same PE regressor vector ξ, the hybrid approach can improve the transient performance over the standard gradient algorithm. □

6.3. Relaxations for semi-global practical stability

Similar to the ODE case (Teel et al., 2003), if *local* or *semi-global* stability results are sought-after, it is possible to significantly relax Assumptions 6.1–6.3, thus enabling the synthesis and analysis of more complex multi-time scale algorithms. For example, in Wang et al. (2012a), singular perturbation and averaging theory were studied using a unifying framework, where the states of the boundary layer dynamics do not necessarily converge to a quasi-steady state manifold, but rather to a periodic orbit or a limit cycle that induces a reduced system via averaging. To illustrate this idea in the context of model-free hybrid control and optimization, we consider the framework of Wang et al. (2012a), which studies HDS of the form:

$$(x, z) \in C \times \Psi, \quad \begin{cases} \dot{x} = f_x^\delta(x, z) \\ \dot{z} = \frac{1}{\varepsilon} f_z(x, z) \end{cases} \tag{6.14a}$$

$$(x, z) \in D \times \Psi, \quad \begin{cases} x^+ \in G_x(x) \\ z^+ = z \end{cases} \tag{6.14b}$$

where $x \in \mathbb{R}^{n_1}$ models the main state of the system, $z \in \mathbb{R}^{n_2}$ models the fast states (e.g., a dynamic oscillator), the sets $C, D \subset \mathbb{R}^{n_1}$ characterize the flow and jump sets, $\Psi \subset \mathbb{R}^{n_2}$ is the set where z evolves, the mapping $f_x^\delta : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_1}$ is allowed to be parameterized by a constant $\delta > 0$, $G_x : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightrightarrows \mathbb{R}^{n_1}$ is a set-valued mapping describing the jumps of the system, and $\varepsilon > 0$ is a small parameter inducing multiple time scales in the flows of (6.14). This system is studied under the following regularity assumptions:

Assumption 6.4. The sets C, D are closed, and Ψ is compact. For all $\delta > 0$ the functions f_x^δ and f_z are continuous, G_x is OSC and LB relative to D, and for each $(x, z) \in D \times \Psi$ the set $G_x(x)$ is not empty. □

The average map of (6.14a) is defined similarly to Definition 6.1, but restricting the slow states to compact sets, which removes the linear dependence on $|x|$ in the right-hand side of (6.5). Also, the average is now computed along the trajectories of the boundary-layer dynamics, which are obtained by keeping x constant in the flows of

(6.14a), i.e., with $\dot{x} = 0$. This average mapping is used in the following assumption, which is common in the literature of averaging (Sastry & Bodson, 1989; Wang et al., 2012a):

Assumption 6.5. There exists a continuous function $\bar{f}^{\delta} : \mathbb{R}^{n_1} \to \mathbb{R}^{n_1}$ such that for each compact set $K \subset \mathbb{R}^{n_1}$ there exists a class-$\mathcal{L}$ function $\sigma_{K,\delta}$ such that the following inequality holds:

$$\left| \frac{1}{T} \int_0^T \left(f_x^{\delta}(x, z_{bl}(s)) - \bar{f}^{\delta}(x) \right) ds \right| \leq \sigma_{K,\delta}(T). \tag{6.15}$$

for each $T > 0$, $x \in C \cap K$, and each $z_{bl} : [0,T] \to \Psi$ satisfying $\dot{z}_{bl}(t) = f_z(x, z_{bl}(t))$, for $t \in [0,T]$. □

Similar to Definition 6.2, the average dynamics of (6.14) are defined in terms of the average map $\bar{f}^{\delta}$.

Definition 6.3. The average HDS of (6.14) has state $\bar{x} \in \mathbb{R}^{n_1}$, and dynamics

$$\dot{\bar{x}} = \bar{f}^{\delta}(\bar{x}), \ \bar{x} \in C, \qquad \bar{x}^+ \in \bar{G}(\bar{x}), \ \bar{x} \in D, \tag{6.16}$$

where $\bar{G}(\bar{x}) = G_x(\bar{x})$. □

In (6.14) it is possible to allow the jump map G_x to also depend on the fast state z, but such dependence usually leads to an average jump map $\bar{G}$ with a more complex structure, see Wang et al. (2012a, Eq. (17) and Ex. 4). Since in most of our applications G_x does not depend on z, we restrict our attention to this setting.

Finally, we assume that the average system (6.16) satisfies the following semi-global practical stability property with respect to the compact set $\mathcal{A}_x$, and in a certain basin of attraction $\mathcal{B}_{\mathcal{A}_x}$. In this case, the uniform convergence properties in $\mathcal{B}_{\mathcal{A}_x}$ are characterized via a proper indicator[2] ϖ, which is sometimes referred to as a "distance function" (Sontag, 2022).

Assumption 6.6. There exists a non-empty compact set $\mathcal{A}_x \subset \mathbb{R}^{n_1}$, an open set $\mathcal{B}_{\mathcal{A}_x} \supset \mathcal{A}_x$, and a class $\mathcal{KL}$ function β such that for each proper indicator $\varpi(\cdot)$ for $\mathcal{A}_x$ on $\mathcal{B}_{\mathcal{A}_x}$, each compact set $K_0 \subset \mathcal{B}_{\mathcal{A}_x}$, and each $\nu > 0$, there exists a $\delta^* > 0$ such that for all $\delta \in (0, \delta^*)$, all solutions of (6.16) with $\bar{x}(0,0) \in K_0$ satisfy the bound:

$$\varpi(\bar{x}(t,j)) \leq \beta(\varpi(\bar{x}(0,0)), t+j) + \nu. \tag{6.17}$$

for all $(t,j) \in \mathrm{dom}(\bar{x})$. □

In (6.17), the residual term ν provides flexibility to study multi-time scale systems with average dynamics (6.16) being a δ-perturbed version of a nominal HDS with suitable uniform asymptotic stability and regularity properties. In that case, by Goebel et al. (2012, Lemma 7.20), the average system satisfies Assumption 6.6.

The following theorem links the stability properties of (6.14) to the stability properties of the average hybrid dynamics (6.16). It follows as a particular case of Poveda and Li (2021, Thm. 7), which is an extension of Wang et al. (2012a, Thm. 2) for HDS having a δ-perturbed average system.

Theorem 6.2. *Suppose that Assumptions 6.4, 6.5, and 6.6 hold. Then, for each proper indicator ϖ for $\mathcal{A}_x$ on $\mathcal{B}_{\mathcal{A}_x}$, each compact set $K_0 \subset \mathcal{B}_{\mathcal{A}_x}$ and each $\nu > 0$ there exists $\delta^* > 0$ such that for each $\delta \in (0, \delta^*)$ there exists $\varepsilon^\star > 0$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ all solutions of (6.14) with $x(0,0) \in K_0$ satisfy:*

$$\varpi(x(t,j)) \leq \beta(\varpi(x(0,0)), t+j) + \nu.$$

for all $(t,j) \in dom(x,z)$. □

[2] A proper indicator of $\mathcal{A}_x$ on $\mathcal{B}_{\mathcal{A}_x}$ is a continuous function $\varpi : \mathcal{B}_{\mathcal{A}_x} \to \mathbb{R}_{\geq 0}$ satisfying $\varpi(x) = 0$ if and only if $x \in \mathcal{A}_x$, and such that $\varpi(x_i) \to \infty$ when $i \to \infty$ if either $|x_i| \to \infty$ or the sequence $\{x_i\}_{i=1}^n$ approaches the boundary of $\mathcal{B}_{\mathcal{A}_x}$.

We finish this section by pointing out that averaging theory has also been recently developed for HDS with bounded exogenous inputs (Wang, Nesic and Teel, 2012) using the notion of input-to-state stability, and by distinguishing between strong and weak averages, as studied for ODEs in Nešić and Teel (1999). Other multi-time scale models studied in the literature include SP-HDS with hybrid boundary layer dynamics (Wang et al., 2012b), ODEs with average systems that lead to differential inclusions (Deghat et al., 2021), averaging on Riemannian manifolds and Lie groups (Bullo, 2002; Taringoo, 2017; Taringoo, Dower, Nesic, & Tan, 2018), and averaging methods that use Poincare maps (De, Burden, & Koditschek, 2018), which are commonly used in legged mechanical systems.

In the next Section, we illustrate the previous results via four novel applications in the context of *model-free optimization* using hybrid extremum-seeking, and *model-free stabilization* using hybrid vibrational control.

6.4. Applications in model-free control and optimization

We present four examples demonstrating the application of averaging and singular perturbation theory to solve model-free stabilization and optimization problems using multi-time scale hybrid control systems.

6.4.1. Robust model-free stabilization under obstacle avoidance and unknown control directions

Consider the model of a 2-dimensional vehicle with dynamics

$$\dot{x} = Ax + bu, \quad x \in \mathbb{R}^2, \tag{6.18}$$

where $x = (x_1, x_2)$ is the position in the plane, $u \in \mathbb{R}^2$ is the input, and $b \neq 0$ is an unknown scalar whose sign describes the effect of the control signal u on the dynamics of the vehicle. The main objective is to stabilize the vehicle at a desired target position $x^\star \in \mathbb{R}^2$, while simultaneously avoiding an obstacle $\mathcal{O} \subset \mathbb{R}^2$ in the space, and subject to the fact that the sign of b is unknown. As discussed in Sanfelice et al. (2006) and Sontag (1999), even when $b = 1$ and $A = 0$, this problem cannot be robustly solved using smooth feedback controllers due to the topological obstructions introduced by the obstacle. On the other hand, when $b = 1$, robust hybrid controllers able to achieve obstacle avoidance and robust target stabilization have been studied in Casau, Cunha, Sanfelice, and Silvestre (2020) and Sanfelice et al. (2006) for cases where the target's position is known a priori, and also in Poveda et al. (2021) for applications where the position of the target is "discovered" in real-time by maximizing a potential field. However, the case when the sign of b is unknown (or time-varying) has received little attention in the hybrid control literature.

To achieve obstacle avoidance and robust stabilization of the target $x^\star$ in systems of the form (6.18), we consider the oscillatory hybrid control law

$$u_q = a\omega \mathrm{R}\mu - k\frac{2}{a} V_q(x)\mu, \quad \dot{\mu} = \omega \mathrm{R}\mu, \ \mu \in \mathbb{S}^1. \tag{6.19}$$

where $\omega = \frac{2\pi}{\varepsilon}$, $\mathrm{R} := [0\ 1; -1\ 0] \in \mathbb{R}^{2\times 2}$, $k, a \in \mathbb{R}_{>0}$ are tunable parameters, μ is the state of a linear oscillator evolving on a faster time scale, and $q \in Q = \{1,2\}$ is a logic state that parameterizes the control law via the functions V_q, which are to be designed. This logic state will switch between the two values (1 and 2) depending on the current position of the vehicle. To introduce this switching law, we first need to partition the operational space of the vehicle. As in Poveda et al. (2021), we restrict our attention to admissible obstacles $\mathcal{O} \subset \mathbb{R}^n$ for which there exists $x_0 = (x_{0,1}, x_{0,2}) \in \mathbb{R}^2$, $\rho \in \mathbb{R}_{>0}$ and $\tilde{\delta} \in \mathbb{R}_{>0}$ such that $\mathcal{O} \subset x_0 + \rho\mathbb{B}$ and $(x_0 + 2\rho\sqrt{2}\mathbb{B}) \cap (x^\star + \tilde{\delta}\mathbb{B}) = \emptyset$. In other words, the obstacle $\mathcal{O}$ is assumed to be contained in a ball of radius ρ, centered at the point x_0, and located sufficiently far away from the target $x^\star$.

Consider the set

$$\mathcal{B}_{x_0,\rho} := \left\{ x : \ \|x - x_0\|_1 \leq 2\rho\sqrt{2} \right\}, \tag{6.20}$$

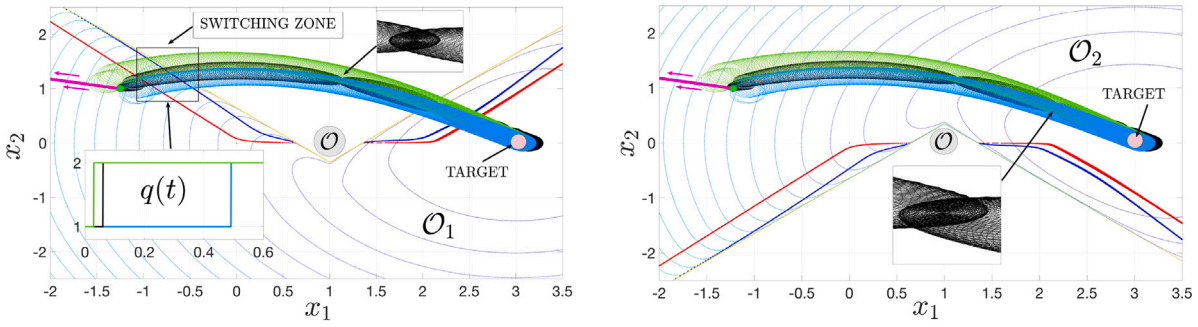

Fig. 8. The left plot shows the trajectories of the vehicle (6.18) controlled via (6.19), evolving over the operational space defined by the logic state $q = 1$. Similarly, the right plot corresponds to $q = 2$. The trajectory with the color blue corresponds to $b = 15$. The green trajectory corresponds to $b = -15$, and the black trajectory is obtained when $b(t) = 15\sin(t)$. The purple trajectory shows the instability obtained when using the model-based hybrid controller of Sanfelice et al. (2006) (which assumes $b > 0$) with $b = -15$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

which satisfies $\{x_0\}+\rho\mathbb{B} \subset \mathcal{B}_{x_0,\rho} \subset \{x_0\}+2\rho\sqrt{2}\mathbb{B}$. Additionally, consider the following subsets of $\mathbb{R}^2$:

$$L_{1a} := \left\{x : \ x_2 < -x_1 + x_{0,2} + x_{0,1} - 2\rho\sqrt{2}\right\},$$
$$L_{1b} := \left\{x : \ x_2 < x_1 + x_{0,2} + x_{0,1} + 2\rho\sqrt{2}\right\},$$
$$L_{2a} := \left\{x : \ x_2 > x_1 + x_{0,2} + x_{0,1} - 2\rho\sqrt{2}\right\},$$
$$L_{2b} := \left\{x : \ x_2 > -x_1 + x_{0,2} + x_{0,1} + 2\rho\sqrt{2}\right\},$$

and let $\mathcal{O}_1 := L_{1a} \cup L_{1b}$, $\mathcal{O}_2 := L_{2a} \cup L_{2b}$, $\mathcal{U} := \mathcal{O}_1 \cup \mathcal{O}_2$. Fig. 8 illustrates the construction of the sets $\mathcal{O}_1$ and $\mathcal{O}_2$, which satisfy $x^\star \in \mathcal{O}_1 \cap \mathcal{O}_2$, and also $\mathcal{O}_1 \cap \mathcal{O} = \emptyset$, $\mathcal{O}_2 \cap \mathcal{O} = \emptyset$. In fact, $\mathcal{U} = \mathbb{R}^2 \backslash \mathcal{B}_{x_0,\rho}$. To control the vehicle in each of the sets $\mathcal{O}_i$, $i \in Q$, we will exploit a mode-dependent control-like Lyapunov function V_q, similar to those studied in the stabilization of ODEs (Scheinker & Krstić, 2017). Specifically, we assume that the functions V_q satisfy the following properties:

(a) For each $q \in Q$, the map $V_q : \mathbb{R}^2 \to \mathbb{R}_{\geq 0} \cup \{\infty\}$ is continuously differentiable in $\mathcal{O}_q$, and as $x \to \infty$ or $x \to \text{bd}(\mathcal{O}_q)$ we have $V_q(x) \to \infty$. Moreover, for every $x \in \mathbb{R}^2 \backslash \mathcal{O}_q$, we define $V_q(x) = \infty$.

(b) For each $q \in Q$, there exist functions $\alpha_{1,q}, \alpha_{2,q} \in \mathcal{K}_\infty$, and proper indicators $\tilde{\varpi}_q$ of $x^\star$ on $\mathcal{O}_q$, such that

$$\alpha_{1,q}(\tilde{\varpi}_q(x)) \leq V_q(x) \leq \alpha_{2,q}(\tilde{\varpi}_q(x)), \quad \forall\, x \in \mathcal{O}_q.$$

(c) There exists $\rho \in \mathcal{PD}$, such that for each $q \in Q$:

$$\left\langle \nabla V_q(x), Ax - kb^2 \nabla V_q(x) \right\rangle \leq -\rho(V_q(x)),$$

for all $x \in \mathcal{O}_q$. □

When $\mathcal{O}_q = \mathbb{R}^2$, the above conditions recover the "strong stabilizability" assumption used in Scheinker and Krstić (2017, Assumption 1). This stabilizability assumption (and its extensions to nonlinear systems) is key in the context of model-free stabilization in $\mathbb{R}^n$ via oscillatory control in ODEs (Scheinker & Krstić, 2017). However, when there are obstacles in the space, this stabilizability assumption cannot be satisfied in $\mathbb{R}^2/\mathcal{O}$. To relax this requirement, conditions (a)–(c) ask for the strong stabilizability property to hold only in the subsets $\mathcal{O}_q$. In practice, this can be achieved by considering a function V_q that involves a quadratic term that is positive definite with respect to $x^\star$, and an additional q-dependent barrier function that grows to infinity as $x \to \text{bd}(\mathcal{O}_q)$. Examples of these constructions are presented in Sanfelice et al. (2006) and Poveda et al. (2021).

Using the functions V_q, we can now introduce the flow and jump set for the state (x, q) of the hybrid controller:

$$C_{x,q} := \left\{(x,q) \in \overline{\mathcal{U}} \times Q : V_q(x) \leq \chi V_{3-q}(x)\right\},$$
$$D_{x,q} := \left\{(x,q) \in \overline{\mathcal{U}} \times Q : V_q(x) \geq (\chi - \lambda) V_{3-q}(x)\right\},$$

where $\chi \in (1, \infty)$ and $\lambda \in (0, \chi - 1)$ are tunable parameters that induce suitable robustness properties. The set $C_{x,1}$ describes the points in the space where the controller u_1 is implemented by the vehicle. Similarly, the set $C_{x,2}$ describes the points where the controller u_2 is used. The sets $D_{x,1}$ and $D_{x,2}$ indicate "switching zones" for the controller, where the vehicle toggles the logic state as $q^+ = 3 - q$. By construction, this switching behavior will take place whenever the value of the current function V_q exceeds a threshold compared to the other function V_{3-q}. In particular, the construction of the flow and jump sets imposes a *hysteresis* property in the switching controller. The red and blue lines in Fig. 8 illustrate the boundaries of $D_{x,q}$ and $C_{x,q}$ respectively.

To study the closed-loop multi-time scale hybrid dynamics, we consider the change of variable $\hat{x} = x - ab\mu$, which leads to the following continuous-time dynamics:

$$\dot{\hat{x}} = A\hat{x} - \frac{2bk}{a} V_q(\hat{x} + ab\mu)\mu + Aab\mu, \quad \dot{q} = 0, \quad \dot{\mu} = \frac{2\pi}{\varepsilon}\mathrm{R}\mu,$$

flowing whenever $(\hat{x} + e_a, q, \mu) \in C_{x,q} \times \mathbb{S}^1$, where $e_a = ab\mu \in \mathcal{O}(a)$ can be seen as a small bounded disturbance because $|\mu(t, j)| \leq 1$ for all hybrid times, and a is a small constant. The discrete-time dynamics are $\hat{x}^+ = \hat{x}$, $q^+ = 3 - q$, $\mu^+ = \mu$, which are executed when $(\hat{x} + e_a, q, \mu) \in D_{x,q} \times \mathbb{S}^1$. When ε is small, the above system has the form of (6.14) with μ playing the role of the fast variable. Using the Taylor expansion $\frac{2}{a} V_q(\hat{x}+ab\mu)\mu = \frac{2}{a} V_q(\hat{x})\mu + b\mu\mu^\top \nabla V_q(\hat{x}) + \mathcal{O}(a)$ and the periodicity of μ, we can compute the average mapping:

$$\dot{\bar{x}} = A\bar{x} - kb^2 \nabla V_q(\bar{x}) + \mathcal{O}(a), \quad \dot{\bar{q}} = 0,$$

which is also an $\mathcal{O}(a)$-perturbed-version of a nominal dynamical system $\dot{\bar{x}} = A\bar{x} - kb^2 \nabla V_q(\bar{x})$, for which V_q is a suitable Lyapunov function in $C_{x,q}$. We can now use the same Lyapunov-based arguments of Sanfelice et al. (2006) to conclude that, under the conditions (a)–(c) of the functions V_q, a hybrid Lyapunov function will decrease along the solutions of this nominal average hybrid dynamics, rendering the set $\mathcal{A} = \{x^\star\} \times Q$ asymptotically stable with a basin of attraction $\mathcal{U} \times Q$. Since this nominal system satisfies the hybrid basic conditions of Assumption 2.1, the same set $\mathcal{A}$ is SGpAS as $a \to 0^+$ for the average hybrid dynamics. By using Theorem 6.2, we can conclude that the multi-time scale hybrid controller renders SGpAS as $(a, \varepsilon) \to 0^+$ the set $\mathcal{A} \times \mathbb{S}^1$.

Fig. 8 shows three different trajectories of the vehicle (6.18) under the highly oscillatory hybrid control law (6.19). The green trajectory is obtained with $b = 15$, and the blue trajectory with $b = -15$. The divergent trajectory, shown in the color magenta, is obtained when using the model-based hybrid controller of Sanfelice et al. (2006), which assumes that $b > 0$ in (6.18). We also tested the performance of the controller for $b(t) = 15\sin(t)$. The resulting trajectory is shown in black color. As observed, the hybrid controller is able to (practically) stabilize the target even when the control direction periodically vanishes, but is positive "on average". Future work will delve into this interesting case by leveraging existing results for ODEs (Scheinker & Krstić, 2017).

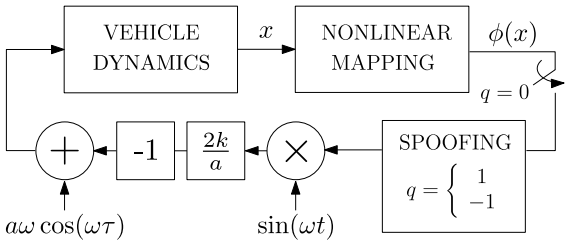

Fig. 9. Scheme of the source-seeking controller under sporadic measurements and persistent spoofing.

Extensions to non-holonomic systems (Morin & Samson, 2009), which require the use of hybrid tools (Hespanha, Liberzon, & Morse, 1999), are also possible.

6.4.2. Source seeking under spoofing and intermittent feedback

We consider a class of 2-dimensional source-seeking problems, where a mobile robot seeks the extremum of a potential field $\phi : \mathbb{R}^2 \to \mathbb{R}$, using only intensity measurements of this signal. This type of source-seeking problem has been extensively studied using traditional averaging techniques (Cochran & Krstić, 2009; Ghods & Krstić, 2010; Zhang, Siranosian, & Krstić, 2007), Lie bracket averaging (Dürr et al., 2015, 2013), and averaging in hybrid systems (Poveda et al., 2021). In these works, the mobile robot is assumed to operate in environments where it has continuous access to the intensity signal. However, in many practical applications, such measurements can be intermittently interrupted due to sensor failures (Labar, Ebenbauer, & Marconi, 2022), or corrupted due to external spoofing. To study whether or not source-seeking is still possible in these scenarios, we consider a vehicle modeled with simple point-mass dynamics of the form

$$\dot{x} = u, \quad x \in \mathbb{R}^2, \quad u \in \mathbb{R}^2, \tag{6.21}$$

where $x = (x_1, x_2)$ indicates the position of the vehicle in the plane. As shown in the scheme of Fig. 9, and inspired by the controller studied in the previous section, we consider again the control law

$$u_q = a\omega \mathrm{R}\mu - k\frac{2}{a}\boldsymbol{\Phi}_q(x)\mu, \quad \dot{\mu} = \omega \mathrm{R}\mu, \quad \mu \in \mathbb{S}^1, \tag{6.22}$$

where R is the same matrix used in (6.19) to generate the periodic dither signals, $\omega = \frac{2\pi}{\epsilon}$, $k, a \in \mathbb{R}_{>0}$ are tunable parameters, and $\boldsymbol{\Phi}_q$ is now the intensity signal *measured by the controller*, given by

$$\boldsymbol{\Phi}_q(x) = \begin{cases} \phi(x) & \text{under nominal operation } (q = 1) \\ 0 & \text{under no measurements } (q = 0) \\ -\phi(x) & \text{under spoofing } (q = -1). \end{cases}$$

In this way, the state $q \in \{-1, 0, 1\}$ now captures the current operating mode of the controller. To simplify our presentation, we assume that the potential field $\phi : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ is continuously differentiable and c_ϕ-strongly convex, with a global minimizer x^*. Moreover, $\nabla\phi$ is L_ϕ-globally Lipschitz. This assumption is standard in the literature of source seeking and it is satisfied if, for example, ϕ is a quadratic map with a positive definite Hessian matrix.

To analyze the stability properties of the source-seeking dynamics under intermittent feedback and spoofing, we consider the change of variable $\hat{x} := x - a\mu$ and the definition of $\boldsymbol{\Phi}$, and we write the resulting dynamics as

$$(\hat{x}, \mu) \in \mathbb{R}^2 \times \mathbb{S}^1, \quad \dot{\hat{x}} = -qk\frac{2}{a}\phi(\hat{x} + a\mu)\mu, \quad \dot{\mu} = \omega \mathrm{R}\mu, \tag{6.23}$$

This system has the form of (6.14) with μ acting as the fast variable. Moreover, since $q \in \{-1, 0, 1\}$ switches between three operating modes (spoofing, no-measurement, nominal), system (6.23) can be seen as a switched system with the signal q being generated by the hybrid automaton (6.10). Therefore, the resulting HDS can be written in the form of (6.14), which satisfies the hybrid basic conditions of Assumption 2.1. To compute the average dynamics, we consider the Taylor expansion $\frac{2}{a}\phi(\hat{x}+a\mu)\mu = \frac{2}{a}\phi(\hat{x})\mu + 2\mu\mu^\top\nabla\phi(\hat{x}) + \mathcal{O}(a^2)$, and the identities $\int_0^L \mu(s)ds = 0$, $\frac{1}{L}\int_0^L \mu(s)\mu(s)^\top ds = \frac{1}{2}$, where L is one period of μ. It follows that the average continuous-time dynamics of $\hat{x}$ are given by

$$\dot{\bar{x}} = \bar{f}(\bar{x}, \bar{q}) = -k\bar{q}\nabla\phi(\bar{x}) + \mathcal{O}(a), \tag{6.24}$$

which is an $\mathcal{O}(a)$-perturbation of a nominal switching dynamical system $\dot{\bar{x}} = -k\bar{q}\nabla\phi(\bar{x})$, with $\bar{q}$ being generated by the same hybrid automaton (6.10) with auxiliary states $\bar{q}_2 \in [0, N]$ and $\bar{q}_3 \in [0, T]$. As in Galarza et al. (2021, Lemma 6), the nominal average hybrid system can be studied using the Lyapunov-like function $V_1(\bar{x}) = V_0(\bar{x}) = V_{-1}(\bar{x}) = |\bar{x} - x^*|^2 =: V(\bar{x})$. Indeed, using the smoothness and strong convexity of ϕ, we have that:

$$\text{For } \bar{q} = 1: \ \langle \nabla V, \dot{\bar{x}} \rangle \leq -2k(\bar{x}-x^*)^\top\nabla\phi(\bar{x}) \leq -2kc_\phi V(\bar{x}),$$
$$\text{For } \bar{q} = 0: \ \langle \nabla V, \dot{\bar{x}} \rangle \leq 0 \leq V(\bar{x}),$$
$$\text{For } \bar{q} = -1: \ \langle \nabla V, \dot{\bar{x}} \rangle \leq 2k(\bar{x}-x^*)^\top\nabla\phi(\bar{x}) \leq 2kL_\phi V(\bar{x}).$$

It follows that, if $\tau_d > 0$ and $\tau_a > 1 + \frac{\max\{1, 2kL_\phi\}}{2kc_\phi}$, then, the average hybrid dynamics render the set $\mathcal{A} = \{x^*\} \times \bar{Q} \times [0, N] \times [0, T]$ UGAS, if $\mathcal{O}(a) = 0$, and SGpAS as $a \to 0^+$, if $\mathcal{O}(a) \neq 0$, thus verifying Assumption 6.6. By Theorem 6.2, the hybrid source-seeking dynamics render the set $\mathcal{A} \times \mathbb{S}^1$ SGPAS as $(a, \epsilon) \to 0^+$. As discussed in Example 6.1, the conditions on τ_a and τ_d translate into average dwell-time and average activation-time constraints on the switching signal q.

To numerically verify this result, we use $\phi(x) = |x - x^*|^2$, $x^* = 0$, $k = 5$, $a = 0.1$, and $\omega = 200$. The hybrid automaton (6.10) that describes the state q evolves with $T_0 = 1$, $N_0 = 1.25$, $\tau_a = 5$, and $\tau_d = 0.5$. Figs. 10 and 11 display the phase plane trajectories and the

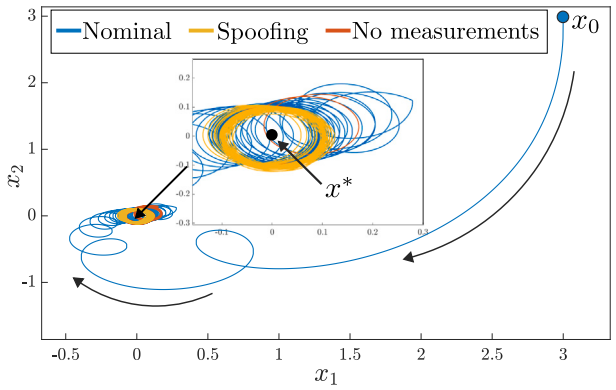

Fig. 10. Trajectories of the vehicle converging to a neighborhood of the source $x^\star$, under spoofing and intermittent feedback.

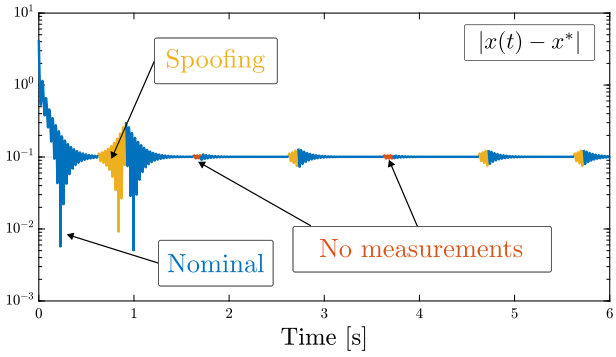

Fig. 11. Trajectory of the error between the position of the vehicle x and the source of the potential field x^*.

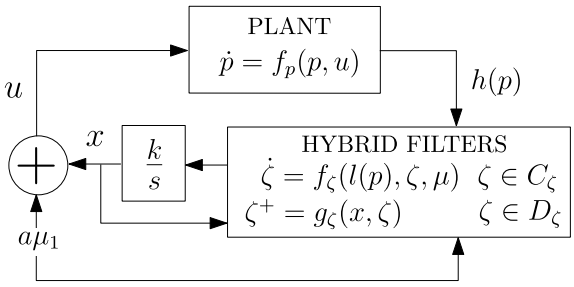

Fig. 12. Scheme of ES controller with hybrid filters.

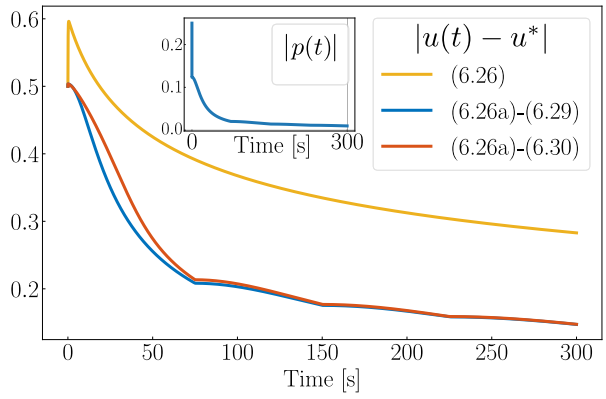

Fig. 13. Evolution in time of the error $|u-u^*|$, with $u^*=0$, using ES dynamics with hybrid filters (6.29) (orange trajectory) and (6.30) (blue trajectory), as well as the traditional low-pass filter (6.26) (black trajectory). The trajectory in the inset shows the evolution of the norm of p for the filter with the best transient performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

error-trajectories, of the vehicle, respectively. As shown, the vehicle converges to a neighborhood of the source x^*, even under persistent spoofing or intermittence on the measurements of the potential ϕ.

6.4.3. Model-free feedback optimization of dynamical systems using hybrid filters

Consider the problem of maximizing the steady-state input-to-output map of the following dynamical system

$$\varepsilon_1 \dot{p}_1 = -p_2 + p_1\left(u - \sqrt{p_1^2+p_2^2}\right) \tag{6.25a}$$

$$\varepsilon_1 \dot{p}_2 = p_1 + p_2\left(u - \sqrt{p_1^2+p_2^2}\right) \tag{6.25b}$$

$$h(p) = (p_1^2+p_2^2)^{\frac{1}{2}}, \tag{6.25c}$$

where $p = (p_1,p_2) \in C_p \subset \mathbb{R}^2\backslash\{0\}$ is the state, $u \in C_u \subset \mathbb{R}_+$ is the input, and h is the output. System (6.25) describes an oscillator with oscillating amplitude given by $h(p) = (p_1^2+p_2^2)^{\frac{1}{2}}$. Indeed, for every fixed and positive input u, every complete solution p to (6.25) rapidly converges to the set $\mathcal{M}(u) := \{p \in \mathbb{R}^2 : h(p)^2 = u^2\}$. Therefore, the steady-state input-to-output map of system (6.25) is given by $h(\mathcal{M}(u)) = u$. Based on this observation, we study the problem of tuning the amplitude u toward a desired value u^* by minimizing the steady-state cost function $\phi(u) := (h(\mathcal{M}(u)) - u^*)^2$, using real-time output feedback via the signal $\ell(p) = (h(p)-u^*)^2$.

To solve this problem, we can consider a standard extremum-seeking controller (Ariyur & Krstić, 2003; Nešić et al., 2012; Tan et al., 2006) with input $u = x + a\mu_1$ and dynamics:

$$\dot{x} = -k_0\zeta_1, \tag{6.26a}$$

$$\dot{\zeta}_1 = -k_1\left(\zeta_1 - F_G(\ell(p),\mu)\right), \tag{6.26b}$$

where $k_1 > 0$ and $k_0 \in \mathbb{R}$ are tunable scalars, μ_1 is the first component of the state $\mu \in \mathbb{S}^1$, which is generated by the same oscillator considered in (6.19) and (6.23), now with frequency $\omega = 2\pi/\varepsilon_2$, and where F_G is given by

$$F_G(\ell(p),\mu) = \frac{2}{a}\ell(p)\mu_1, \tag{6.27}$$

with $\varepsilon_1 \ll \varepsilon_2 \ll a$ to induce multiple time scales in the closed-loop dynamics. In (6.26), when $k_0 > 0$ the low-pass filter with state $\zeta_1 \in \mathbb{R}$ can be used to reduce oscillations and improve the transient (Ariyur & Krstić, 2003; Nešić et al., 2012; Tan et al., 2006) . However, in certain cases, it is also possible to improve the transient behavior by considering hybrid filters or compensators with time-triggered resets, given by

$$\zeta_2 \in [0,\overline{T}], \qquad \dot{\zeta}_1 = f_\zeta(\ell(p),\zeta,\mu), \qquad \dot{\zeta}_2 = 1, \tag{6.28a}$$

$$\zeta_2 \in [T,\overline{T}], \qquad \zeta_1^+ = g_\zeta(x,\zeta), \qquad \zeta_2^+ = 0, \tag{6.28b}$$

where ζ_2 acts as a resetting timer, $0 < T \le \bar{T} < \infty$ are tunable parameters, and $\zeta = (\zeta_1,\zeta_2) \in \mathbb{R}^2 \times [0,\overline{T}]$. Fig. 12 shows a block diagram of the resulting closed-loop system with hybrid filters, where $C_\zeta = \mathbb{R}\times[0,\bar{T}]$, and $D_\zeta = \mathbb{R}\times[T,\bar{T}]$.

The hybrid dynamics (6.28) have non-unique solutions since they enable resets of ζ_1 whenever $\zeta_2 \ge T$, but not later than when $\zeta_2 = \overline{T}$. Different choices of continuous functions (f_ζ, g_ζ) can be considered to influence the transient performance of the system. For example, the functions

$$f_\zeta(\ell(p),\zeta,\mu) = -\frac{k_1}{\gamma(\zeta_2)}\zeta_1 - k_2 F_G(\ell(p),\mu), \quad g_\zeta(x,\zeta) = 0, \quad k_2 > 0, \tag{6.29}$$

model a low-pass filter with dynamic cut-off frequency regulated by the strictly positive and increasing continuous function $\gamma : \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$, resetting the filter state to zero whenever $\zeta_2^+ = 0$. Interestingly, when $k_0 = -1$, the dynamics (6.26a), (6.28), and (6.29) can be interpreted as a momentum-based optimization algorithm with restarting. Similar dynamics have been recently studied in the context of machine learning (Su, Boyd, & Candes, 2016).

Alternatively, we could also consider the filter dynamics:

$$f'_\zeta(\ell(p),\zeta,\mu) = -\left(\frac{k_1}{\gamma(\zeta_2)} + F_H(\ell(p),\mu)\right)\zeta_1 - \left(k_2 + \frac{k_3}{\gamma(\zeta_2)}\right)F_G(\ell(p),\mu), \tag{6.30}$$

where $k_3 > 0$, which incorporates Hessian-like-driven damping via the term $F_H(\ell(p),\mu) := \frac{16}{a^2}\ell(p)(\mu_1^2 - \frac{1}{2})$, and using the same reset rule $g_\zeta(x,\mu) = 0$. In this case, the choice $k_0 = -1$ and the interconnection of (6.26a) and (6.30) also leads to a momentum-based algorithm with restarting (Poveda & Li, 2021; Poveda & Teel, 2020).

The stability properties of the closed-loop dynamics (6.25), (6.26a), (6.28) can be studied using singular perturbations and averaging theory for hybrid systems. In particular, by neglecting the plant dynamics and substituting the output feedback signal $\ell(p)$ in (6.27) and (6.30) by the steady-state input-to-output map $\phi(u)$, the resulting system can be studied via averaging theory leading to the following average maps of (6.29) and (6.30):

$$\bar{f}_\zeta(\bar{x},\bar{\zeta}) = -\frac{k_1}{\gamma(\bar{\zeta}_2)}\bar{\zeta}_1 - k_2\nabla\phi(\bar{x}) + \mathcal{O}(a),$$

$$\bar{f}'_\zeta(\bar{x},\bar{\zeta}) = -\left(\frac{k_1}{\gamma(\bar{\zeta}_2)} + \nabla^2\phi(\bar{x})\right)\bar{\zeta}_1 - \left(k_2 + \frac{k_3}{\gamma(\bar{\zeta}_2)}\right)\nabla\phi(\bar{x}) + \mathcal{O}(a).$$

Thus, if $\mathcal{O}(a) = 0$, and ϕ is continuously differentiable, c_ϕ-strongly convex with a global minimizer x^*, $\nabla\phi$ is L_ϕ-globally Lipschitz, and $k_3 = 0$, then it can be shownn that the Lyapunov-like function $V(\bar{x},\bar{\zeta}) = |\bar{\zeta}|^2 + k_2(\phi(\bar{x}) - \phi(x^*))$ does not increase during flows and jumps of the average nominal system, and asymptotic stability of the set $\mathcal{A} = \{x^*\}\times\{0\}\times[0,\overline{T}]$ can be obtained via the hybrid invariance principle (see Example 5.2). Since the average hybrid dynamics satisfy the hybrid

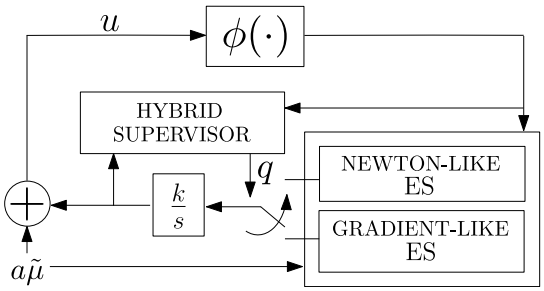

Fig. 14. Scheme of a uniting hybrid ES algorithm that implements Newton-like ES and gradient-like ES in different regions of the state space.

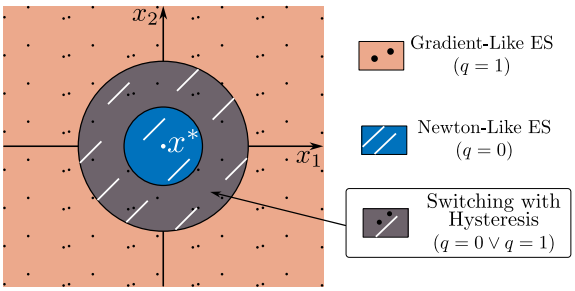

Fig. 15. Illustration of flow and jump sets associated with the uniting hybrid ES algorithm.

basic conditions, the set $\mathcal{A}$ is actually SGpAS as $a \to 0^+$ whenever $\mathcal{O}(a)$ is not neglected. Theorem 6.2 can now be invoked to establish SGpAS for the original closed-loop system interconnected with the hybrid ES controller.

The impact of the hybrid filters on the transient performance of the system is shown in Fig. 13, where we used (6.29) and (6.30) with $g_\zeta(x,\zeta) = 0$. We also show in black color the trajectories obtained when using the classic low-pass filter of (6.26). All simulations were implemented using the same frequency, amplitude, and gain parameters.

6.4.4. *Newton-gradient-like switched extremum seeking for enhanced transient performance*

Consider the following optimization problem

$$\min \ \phi(x), \quad x \in \mathbb{R}^2, \tag{6.31}$$

where ϕ is available to the optimization algorithm only via measurements or evaluations. This setting describes a zeroth-order optimization problem that can be addressed via averaging-based techniques. In particular, when ϕ is continuously differentiable, κ_ϕ-strongly convex with a global minimizer x^*, and $\nabla\phi$ is L_ϕ-globally Lipschitz, problem (6.31) can be efficiently solved using Newton-like extremum-seeking algorithms that achieve better transient performance (i.e., with user-assignable convergence rates) compared to traditional gradient-like ES methods (Galarza et al., 2021; Ghaffari et al., 2012). However, the real-time estimation and inversion of the Hessian of ϕ in Newton-like ES algorithms can be very sensitive to measurement noise and prone to instabilities when the dynamics are initialized far from the set of minimizers of (6.31). While these issues can be addressed by increasing the frequency of the dither signals, doing so might require smaller sampling intervals to mitigate aliasing. This, in turn, can complicate the implementation of the algorithms in systems with computational limitations.

An intuitive solution to address the above issue is to use the less sensitive gradient-like ES algorithm whenever the trajectories are far from the optimizer, and to switch to a Newton-like ES method to fine-tune the convergence near the optimal point; see Martens and Sutskever (2012, pp. 23) for a discussion of this approach in the context of training neural networks. To formalize this scheme, shown in Fig. 14, let $\varepsilon_2 \gg \varepsilon_1 > 0$, and consider the three-time scale HDS with continuous-time dynamics:

$$\dot{x} = -q\zeta_1 - (1-q)\zeta_2\zeta_1, \quad \dot{q} = 0, \tag{6.32a}$$

$$\varepsilon_2\dot{\zeta}_1 = -\zeta_1 + F_G(\phi,\mu) \tag{6.32b}$$

$$\varepsilon_2\dot{\zeta}_2 = \zeta_2 - \zeta_2 F_H(\phi,\mu)\zeta_2 \tag{6.32c}$$

$$\varepsilon_1\varepsilon_2\dot{\mu} = 2\pi\mathcal{R}_\omega\mu, \tag{6.32d}$$

where $q \in \{0,1\}$ is a logic state, $\mathcal{R}_\omega$ is now a 2-block diagonal matrix with diagonal blocks $\omega_i R$, where R is the same matrix of (6.19), and ω_i are positive rational numbers satisfying $\omega_1 \neq \omega_2$. The mappings F_G and F_H are the following gradient and Hessian estimators:

$$F_G(\phi,\mu) := \frac{2}{a}\phi(x+a\tilde{\mu})\tilde{\mu}, \ F_H(\phi,\mu) := \phi(x+a\tilde{\mu})N(\tilde{\mu}),$$

where $a > 0$ is small, $\tilde{\mu} = (\mu_1, \mu_3)$, and $N : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ is a matrix-valued function with entries satisfying $N_{11} = \frac{16}{a^2}(\mu_1^2 - \frac{1}{2})$, $N_{22} = \frac{16}{a^2}(\mu_3^2 - \frac{1}{2})$, $N_{12} = \frac{4}{a^2}\mu_1\mu_3$, and $N_{12} = N_{21}$.

Since the role of the logic state q in Eq. (6.32a) is to enable switching between the vector fields $-\xi_1$ and $-\xi_2\xi_1$ in (6.32a), the discrete-time dynamics of the system are given by

$$x^+ = x, \quad q^+ = 1-q, \quad \zeta^+ = \zeta, \quad \mu^+ = \mu, \tag{6.33}$$

To define the sets C and D, we follow a uniting control approach (Sanfelice, 2021), where the logic state q is switched whenever the value of the cost function ϕ is sufficiently small with respect to ϕ^*. We note that this approach might require a reasonable knowledge of the value (or range) of ϕ^*, which is usually informed by the physics of the particular problem of interest. In certain applications, we have $\phi^* = 0$, which can simplify the implementation. Based on this, let $0 < c_{10} < c_0$ be tunable parameters of the algorithm, and consider the sets

$$C_0 := \{x \in \mathbb{R}^2 : \phi(x) - \phi^* \leq c_0\}, \tag{6.34a}$$

$$C_1 := \overline{\mathbb{R}^n \backslash \{x \in \mathbb{R}^2 : \phi(x) - \phi^* < c_{10}\}}, \tag{6.34b}$$

as well as the sets

$$D_0 := \overline{\mathbb{R}^n \backslash \{x \in \mathbb{R}^2 : \phi(x) - \phi^* < c_0\}}, \tag{6.35a}$$

$$D_1 := \{x \in \mathbb{R}^2 : \phi(x) - \phi^* \leq c_{10}\}. \tag{6.35b}$$

Using (6.34) and (6.35), the set C and D are given by

$$C := \left(C_0 \times \{0\}\right) \cup \left(C_1 \times \{1\}\right), \tag{6.36a}$$

$$D := \left(D_0 \times \{0\}\right) \cup \left(D_1 \times \{1\}\right), \tag{6.36b}$$

see Fig. 15. Since in practice the ES dynamics only have access to measurements of $\phi(x + a\mu)$, the hybrid dynamics (6.32)–(6.33) are implemented under the following (perturbed) flow and jump sets:

$$\left((x+e_a, q), (\zeta_1,\zeta_2), \mu\right) \in C \times K \times \mathbb{T}$$
$$\left((x+e_a, q), (\zeta_1,\zeta_2), \mu\right) \in D \times K \times \mathbb{T},$$

where $\mathbb{T} := \mathbb{S}^1 \times \mathbb{S}^1$, $K \subset \mathbb{R}^2 \times \mathbb{R}^{2\times 2}$ is a compact set that can be taken arbitrarily large to encompass any solution of interest, and $e_a = a\mu \in \mathcal{O}(a)$ is a small measurement perturbation since $|\mu(t,j)| \leq 1$ for all hybrid times. Since $\varepsilon_1 > 0$ acts as small parameter in (6.32), we can compute the average continuous-time dynamics:

$$\dot{\bar{x}} = -\bar{q}\bar{\zeta}_1 - (1-\bar{q})\bar{\zeta}_2\bar{\zeta}_1, \quad \dot{\bar{q}} = 0, \tag{6.37a}$$

$$\varepsilon_2\dot{\bar{\zeta}}_1 = -\bar{\zeta}_1 + \nabla\phi(\bar{x}) + \mathcal{O}(a) \tag{6.37b}$$

$$\varepsilon_2\dot{\bar{\zeta}}_2 = \bar{\zeta}_2 - \bar{\zeta}_2\nabla^2\phi(\bar{x})\bar{\zeta}_2 + \mathcal{O}(a). \tag{6.37c}$$

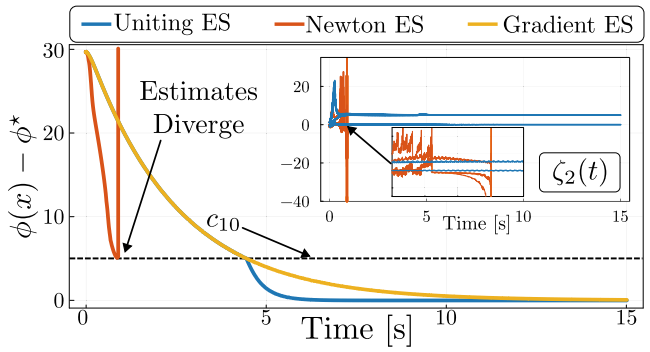

Fig. 16. Trajectories generated by Newton-like ES (diverging) and the hybrid uniting ES (converging to x^*), with $\phi(x) = \frac{1}{10}\left((x_1-1)^2 + (x_2-5)^2\right)$, $c_0 = 6$, and $c_{10} = 5$.

which evolve in the set $((\bar{x},\bar{q}),(\bar{\zeta}_1,\bar{\zeta}_2)) \in C \times K$; and the average discrete-time dynamics:

$$\bar{x}^+ = \bar{x}, \quad \bar{q}^+ = 1 - \bar{q}, \quad \bar{\zeta}^+ = \bar{\zeta}, \tag{6.38}$$

which evolve in the set $((\bar{x},\bar{q}),(\bar{\zeta}_1,\bar{\zeta}_2)) \in D \times K$. This HDS is an $\mathcal{O}(a)$-perturbed version of a nominal HDS that is singularly perturbed with small parameter ε_2. To study this nominal system, we consider its boundary layer dynamics which are given by the low-pass filter (6.37b) and the differential Ricatti Eq. (6.37c). For each fixed $\bar{x}$, these dynamics render locally exponentially stable the equilibrium points $\bar{\zeta}_1^* = \nabla\phi(\bar{x})$ and $\bar{\zeta}_2^* = \nabla^2\phi(\bar{x})^{-1}$. Thus, since by assumption $\nabla^2\phi(\bar{x}) \succ \kappa_\phi I_n$ for all $\bar{x} \in \mathbb{R}^2$, the nominal ($a = 0$) averaged reduced HDS is given by:

$$(\bar{x},\bar{q}) \in C, \quad \begin{cases} \dot{\bar{x}} = -\bar{q}\nabla\phi(\bar{x}) - (1-\bar{q})\nabla^2\phi(\bar{x})^{-1}\nabla\phi(\bar{x}), \\ \dot{\bar{q}} = 0 \end{cases} \tag{6.39a}$$

$$(\bar{x},\bar{q}) \in D, \quad \begin{cases} \bar{x}^+ = \bar{x} \\ \bar{q}^+ = 1 - \bar{q}. \end{cases} \tag{6.39b}$$

By using the smoothness and strong convexity properties of ϕ, the function $V(\bar{x}) = \phi(\bar{x}) - \phi^*$ is radially unbounded, and it satisfies:

$$\langle \nabla V(\bar{x}), -\nabla\phi(\bar{x})\rangle \leq -|\nabla\phi(\bar{x})|^2, \quad \forall\ \bar{x} \in \mathbb{R}^2,$$

and also

$$\left\langle \nabla V(\bar{x}), -\nabla^2\phi(\bar{x})^{-1}\nabla\phi(\bar{x})\right\rangle \leq -\kappa_\phi|\nabla\phi(\bar{x})|^2, \ \forall\ \bar{x} \in \mathbb{R}^2.$$

Therefore the sub-level sets of V are forward invariant for each fixed $\bar{q}$. Since each operating mode $\bar{q}$ renders the point x^* UGAS, all the assumptions of Sanfelice (2021, Thm. 4.6) are satisfied, and the hybrid controller (6.39) renders UGAS the set $\mathcal{A} = \{x^*\} \times \{0\}$. A repeated application of Theorem 6.2 (with respect to ε_2 and ε_1), and Goebel et al. (2012, Lemma 7.20), allows us to conclude SGpAS as $(\varepsilon_2, a, \varepsilon_1) \to 0^+$ of the set $\mathcal{A} \times K \times \mathbb{T}$ for the original hybrid ES dynamics (6.32)–(6.33).

We illustrate the uniting extremum-seeking algorithm (6.32) through the simulation results in Fig. 16. The figure shows diverging trajectories that emerge when using Newton-like extremum-seeking dynamics far away from the optimal point. By incorporating the hybrid uniting mechanism presented in this section, not only the instability is removed, but also the transient performance of the controller is improved, outperforming the less sensitive gradient-like extremum-seeking dynamics.

Finally, we note that while the findings in this section were presented for the case of $x \in \mathbb{R}^2$, these results can be naturally extended to the more general n-dimensional case using the uniting control methodology (Goebel et al., 2012; Sanfelice, 2021)

7. Proofs

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

7.1. Proof of Proposition 3.5

Via direct computation, we obtain that:

$$[w, f_1](\xi,\tau) = \sum_{i,j=1}^{r} [b_i, b_j](\xi)\, u_j(\tau) \int_0^\tau u_i(v)\, dv.$$

Furthermore, the properties of the Lie bracket imply

$$[b_i, b_j] = (\delta_{ij} - 1)[b_j, b_i],$$

where δ_{ij} is the Kronecker delta symbol. Therefore, we obtain that:

$$\sum_{\substack{i=1\\ j=1}}^{r} [b_i, b_j]\, u_j(\tau) \int_0^\tau u_i(v)\, dv = \sum_{\substack{i=1\\ j>i}}^{r} [b_i, b_j]\, u_{ji}(\tau),$$

where the function u_{ji} is given by:

$$u_{ji}(\tau) = u_j(\tau) \int_0^\tau u_i(v)\, dv - u_i(\tau) \int_0^\tau u_j(v)\, dv.$$

Hence, the averaged vector field $\bar{f}_2$ is given by:

$$\bar{f}_2(\xi) = b_0(\xi) + \sum_{i=1, j>i}^{r} [b_i, b_j](\xi) v_{ji},$$

where the constants v_{ji} are defined by:

$$v_{ji} = \frac{1}{2T} \int_0^T \left(u_j(\tau) \int_0^\tau u_i(v)\, dv - u_i(\tau) \int_0^\tau u_j(v)\, dv \right) d\tau.$$

Using integration by parts, we observe that:

$$\int_0^T u_i(\tau) \int_0^\tau u_j(v)\, dv\, d\tau = -\int_0^T u_j(\tau) \int_0^\tau u_i(v)\, dv\, d\tau,$$

where the boundary term coming out of the integration by parts vanishes due to the properties of the functions u_i. Hence, the constants v_{ji} simplify to:

$$v_{ji} = \frac{1}{T} \int_0^T u_j(\tau) \int_0^\tau u_i(v)\, dv\, d\tau,$$

which establishes the result. ■

7.2. Proof of Theorem 4.3

We begin by applying the time scaling $s = \varepsilon^{-2}t$. In contrast to the traditional Lyapunov-based singular perturbation analysis (e.g. Khalil (2002, Chapter 11)), we augment the standard coordinate shift with a *near-identity* part:

$$y = z - \varphi_0(x) - \sum_{i=1}^{2} \varepsilon^i \varphi_i(x,\tau), \tag{7.1}$$

where the maps $\varphi_i(x,\tau)$ for $i \in \{1,2\}$ are the solutions to the linear non-homogeneous two point boundary value problems:

$$\nabla_2\varphi_i(x,\tau) = A(\tau;x)\,\varphi_i(x,\tau) + b_i(x,\tau), \tag{7.2a}$$

$$\varphi_i(x,\tau) = \varphi_i(x,\tau+T), \tag{7.2b}$$

for $i \in \{1,2\}$, where:

$$b_1(x,\tau) = b_\varphi(x,\tau),$$
$$b_2(x,\tau) = f_{z,2}(x,\varphi_0(x),\tau) + J_2^2 f_{z,0}(x,\varphi_0(x),\tau)[\varphi_1(x,\tau)] - J_1\varphi_0(x) f_{x,2}(x,\varphi_0(x),\tau) - J_1\varphi_1(x,\tau) f_{x,1}(x,\varphi_0(x),\tau),$$

and the map $J_2^2 f_{z,0}$ is a vector-valued bi-linear form. The transformation (7.1) is akin to the composite expansion (4.15) in the Boundary Function method, whereas the boundary condition (7.2b) parallels the condition (4.22).

The following lemma is a simple consequence of Assumptions 4.1 and 4.2 and standard linear systems theory:

Lemma 7.1. *Let Assumptions 4.1 and 4.2 be satisfied. Then, the non-homogeneous BVPs (7.2a)–(7.2b) have unique periodic solutions φ_i defined by:*

$$\varphi_i(x,\tau) = \int_{\tau}^{\tau+T} E_{\Phi}(\tau,\nu) b_i(x,\nu)\, d\nu.$$

for all $\tau \in \mathbb{R}_{\geq 0}$ and for all $x \in \mathbb{R}^{n_1}$. □

We observe that under this coordinate change and time scaling, we have that:

$$\frac{dx}{ds} = \sum_{k=1}^{2} \varepsilon^k \tilde{f}_{1,k}(x,y,\tau) + O(\varepsilon^3), \tag{7.3a}$$

$$\frac{dy}{ds} = \sum_{k=0}^{2} \varepsilon^k \tilde{f}_{2,k}(x,y,\tau) + O(\varepsilon^3), \tag{7.3b}$$

$$\frac{d\tau}{ds} = 1, \tag{7.3c}$$

where the functions $\tilde{f}_{j,k}$ are given by:

$$\begin{aligned}
\tilde{f}_{2,0}(x,y,\tau) &= f_{z,0}(x,y+\varphi_0(x),\tau),\\
\tilde{f}_{1,1}(x,y,\tau) &= f_{x,1}(x,y+\varphi_0(x),\tau),\\
\tilde{f}_{1,2}(x,y,\tau) &= f_{x,2}(x,y+\varphi_0(x),\tau) + J_2 f_{x,1}(x,y+\varphi_0(x),\tau)\varphi_1(x,\tau),\\
\tilde{f}_{2,1}(x,y,\tau) &= f_{z,1}(x,y+\varphi_0(x),\tau) + J_2 f_{z,0}(x,y+\varphi_0(x),\tau)\varphi_1(x,\tau)\\
&\quad - J\varphi_0(x) f_{x,1}(x,y+\varphi_0(x),\tau) - \nabla_2\varphi_1(x,\tau),\\
\tilde{f}_{2,2}(x,y,\tau) &= f_{z,2}(x,y+\varphi_0(x),\tau) + J_2 f_{z,0}(x,y+\varphi_0(x),\tau)\varphi_2(x,\tau)\\
&\quad + J_2^2 f_{z,0}(x,y+\varphi_0(x),\tau)[\varphi_1(x,\tau)] - J\varphi_0(x) f_{x,2}(x,y+\varphi_0(x),\tau)\\
&\quad - J_1\varphi_1(x,\tau) f_{x,1}(x,y+\varphi_0(x),\tau) - \nabla_2\varphi_2(x,\tau),
\end{aligned}$$

and the remainder terms are Lipschitz continuous and bounded on every compact subset $K \subset \mathbb{R}^{n_1}\times\mathbb{R}^{n_2}$, uniformly in τ and $\varepsilon \in [0,\varepsilon_0]$ for some $\varepsilon_0 > 0$, with Lipschitz constants $L_{f_{1,k},\psi}, L_{f_{2,k},K} > 0$ and bounds $B_{f_{1,k},K}, B_{f_{2,k},K} > 0$.

Owing to the definition of the maps $\varphi_i(x,\tau)$ for $i \in \{1,2\}$, we observe that for all $x \in \mathbb{R}^{n_1}$, and for all $\tau \in \mathbb{R}$, we have that

$$\begin{aligned}
\tilde{f}_{1,1}(x,0,\tau) &= \tilde{f}_1(x,\tau),\\
\tilde{f}_{1,2}(x,0,\tau) &= \tilde{f}_2(x,\tau),\\
\tilde{f}_{2,1}(x,0,\tau) &= \tilde{f}_{2,2}(x,0,\tau) = 0.
\end{aligned}$$

That is, the origin $y = 0$ is an equilibrium point for the boundary layer model:

$$\frac{dy}{ds} = \sum_{k=0}^{2} \varepsilon^k \tilde{f}_{2,k}(x,y,s), \tag{7.5}$$

when x is treated as a parameter. From second-order averaging, we know that there exists $\varepsilon_0 \in (0,\infty)$ such that for all $x(0) \in K_1$, for all $\varepsilon \in (0,\varepsilon_0)$, unique trajectories $(\tilde{x},\tau)$ of the reduced order system (4.26) exist and $\tilde{x}(t) \in \mathcal{N}_1$ for all $t \in [0,T_f]$ for some compact subset $\mathcal{N}_1 \supset \mathcal{N}$. Equivalently, we know that $\tilde{x}(s) \in \mathcal{N}_1$ for all $s \in [0, T_f/\varepsilon^2]$.

Due to Assumption 4.1, we have that for all $(x(0),y(0)) \in K_1 \times K_2$, and for all $\varepsilon \in (0,\varepsilon_0)$, unique maximal trajectories of the system (7.3) exist. Let $[0,s_e)$ with $s_e > 0$ be the maximal interval of existence and uniqueness of a given solution (x,y,τ). For $s \in [0,s_e)$, define an open neighborhood $\mathcal{O}(s)$ around $\tilde{x}(s)$ by:

$$\mathcal{O}(s) = \{x \in \mathbb{R}^{n_1} \ : \ |x - \tilde{x}(s)| < \epsilon\},$$

and observe that the x-component of the solution to (7.3) is initially inside $\mathcal{O}(0)$. Moreover, define the compact subset $\mathcal{M}_1 = \mathcal{N}_1 + \epsilon\mathbb{B}$. From Assumption 4.2, we know that there exists a continuously differentiable function V, and positive constants κ_i, $i \in \{1,2,3,4\}$ such that:

$$\begin{aligned}
&\kappa_1|y|^2 \leq V(y,\tau) \leq \kappa_2|y|^2,\\
&\nabla_1 V^{\top}\tilde{f}_{2,0}(y,\tau) + J_2 V(y,\tau) \leq -\kappa_3|y|^2,\\
&|\nabla_1 V(y,\tau)| \leq \kappa_4|y|.
\end{aligned}$$

for all $y \in \mathbb{R}^{n_2}$ and all $\tau \in \mathbb{R}_{\geq 0}$. Let $c > 0$ be such that the compact $\mathcal{N}_2 = \{y \in \mathbb{R}^{n_2} \ : \ |y| \leq \sqrt{c/\kappa_2}\}$ contains the bounded set K_2. Define the compact set $\mathcal{M}_2 = \{y \in \mathbb{R}^{n_2} \ : \ |y| \leq \sqrt{c/\kappa_1}\}$, and let $K = \mathcal{M}_1 \times \mathcal{M}_2$. Finally, define the time s_ε as follows: $x(s) \in \mathcal{O}(s)$, for all $s \in [0,s_\varepsilon)$, and $|x(s_\varepsilon) - \tilde{x}(s_\varepsilon)| = \epsilon$, or $s_\varepsilon = s_e$ if $x(s) \in \mathcal{O}(s)$, for all $s \in [0,s_e)$. Then, we have the following lemma adapted from a portion of the proof of Lemma 1 in Dürr et al. (2015):

Lemma 7.2 (*Dürr et al., 2015*). *Let the assumptions of Theorem 4.3 be satisfied. Then, there exist constants $\lambda > 0$, $\kappa > 0$, $\gamma > 0$, and $\varepsilon_2 \in (0,\varepsilon_0)$ such that for all $\varepsilon \in (0,\varepsilon_2)$, for all $(x(0),y(0)) \in K_1 \times K_2$, every solution (x,y) to (7.3) stays inside K for all $s \in [0,s_\varepsilon]$, and:*

$$|y(s)| \leq \gamma\,|y(0)|e^{-\lambda s} + \kappa\varepsilon^{\frac{3}{2}},$$

for all $s \in [0,s_\varepsilon]$. □

The proof of this lemma is similar to that in Dürr et al. (2015), so we do not replicate it here. We proceed to define $\varepsilon_3 = \min\{\varepsilon_1,\varepsilon_2\}$, and an ε-dependent time s_ε by requiring that, for all $y(0) \in \mathcal{M}_2$, for all $s > s_\varepsilon$, the following inequality is satisfied:

$$\gamma\,|y(0)|e^{-\lambda s} \leq \kappa\varepsilon^{\frac{3}{2}}, \tag{7.7}$$

We note that this is always possible for $\varepsilon > 0$. Indeed, it can be shown that $s_\varepsilon = \max\{(3/(2\lambda))\,\log((\gamma\sqrt{c/\alpha_2})/(\alpha\varepsilon)),\,0\}$ satisfies the inequality (7.7). Now, we show that there exists $\varepsilon_4 \in (0,\varepsilon_3)$ such that $s_\varepsilon < s_e$, for all $\varepsilon \in (0,\varepsilon_4)$. To obtain a contradiction, suppose that there exists a bounded subset $K_1 \times K_2 \subset \mathbb{R}^{n_1}\times\mathbb{R}^{n_2}$, and an $\epsilon \in (0,\infty)$, such that for all $\varepsilon_4 \in (0,\varepsilon_3)$, there exists $\varepsilon \in (0,\varepsilon_4)$ such that $s_\varepsilon \geq s_e$. We estimate the difference:

$$|x(s_\varepsilon) - \tilde{x}(s_\varepsilon)| \leq \int_0^{s_\varepsilon} B_{K,f_1}\varepsilon\, d\tau \leq B_{K,f_1} s_\varepsilon\,\varepsilon,$$

where $B_{K,f_1} > 0$ is a uniform upper bound on the norm of the integrand inside the compact subset K whose existence is guaranteed by Assumption 4.1. Now, observe that $\lim_{\varepsilon\to 0} s_\varepsilon\,\varepsilon = 0$, and so there exists $\varepsilon_4 \in (0,\varepsilon_3)$ such that $B_{K,f_1} s_\varepsilon\,\varepsilon \leq \epsilon/2$, for all $\varepsilon \in (0,\varepsilon_4)$. Hence, we have that for all $\varepsilon \in (0,\varepsilon_4)$, $|x(s_\varepsilon) - \tilde{x}(s_\varepsilon)| \leq \epsilon/2$ which contradicts the definition of s_ε Next, we show that there exists $\varepsilon_5 \in (0,\varepsilon_4)$ such that $T_f/\varepsilon^2 < s_\varepsilon$, for all $\varepsilon \in (0,\varepsilon_5)$. To obtain a contradiction, suppose that for all $\varepsilon_5 \in (0,\varepsilon_4)$, there exists $\varepsilon \in (0,\varepsilon_5)$ such that $T_f/\varepsilon^2 \geq s_\varepsilon$. Once again, we estimate the difference $|x(s) - \tilde{x}(s)|$ on the interval $[0,s_\varepsilon]$. We have that:

$$|x(s) - \tilde{x}(s)| = \Big|\sum_{k=1}^{2}\varepsilon^k \int_0^{s_\varepsilon} \Delta\tilde{f}_{1,k}(\nu)d\nu + \varepsilon^k\int_{s_\varepsilon}^{s}\Delta\tilde{f}_{1,k}(\nu)d\nu + O(\varepsilon^3)\Big|,$$

where the integrands $\Delta\tilde{f}_{1,k}(s)$ are given by:

$$\Delta\tilde{f}_{1,k}(s) = \tilde{f}_{1,k}(x(s),y(s),\tau(s)) - \tilde{f}_{1,k}(\tilde{x}(s),0,\tau(s)),$$

which leads to the estimate:

$$|x(s) - \tilde{x}(s)| \leq B_{K,f_1}(s_\varepsilon + s_e\varepsilon^2)\varepsilon + |I_1|, \tag{7.8}$$

$$I_1 = \int_{s_\varepsilon}^{s}\sum_{k=1}^{2}\varepsilon^k\Delta\tilde{f}_{1,k}(\nu)d\nu, \tag{7.9}$$

on the interval $[0,s_\varepsilon]$. We proceed to estimate $|I_1|$ as follows:

$$|I_1| \leq \varepsilon(|I_2| + |I_3|) + \varepsilon^2(|I_4| + |I_5|), \tag{7.10}$$

where I_i for $i \in \{2,3,4,5\}$ are given by:

$$\begin{aligned}
I_2 &= \int_{s_\varepsilon}^{s}(\tilde{f}_{1,1}(x(\nu),y(\nu),\tau(\nu)) - \tilde{f}_{1,1}(x(\nu),0,\tau(\nu)))\,d\nu,\\
I_3 &= \int_{s_\varepsilon}^{s}(\tilde{f}_{1,1}(x(\nu),0,\tau(\nu)) - \tilde{f}_{1,1}(\tilde{x}(\nu),0,\tau(\nu)))d\nu,\\
I_4 &= \int_{s_\varepsilon}^{s}(\tilde{f}_{1,2}(x(\nu),y(\nu),\tau(\nu)) - \tilde{f}_{1,2}(x(\nu),0,\tau(\nu)))\,d\nu,\\
I_5 &= \int_{s_\varepsilon}^{s}(\tilde{f}_{1,2}(x(\nu),0,\tau(\nu)) - \tilde{f}_{1,2}(\tilde{x}(\nu),0,\tau(\nu)))d\nu.
\end{aligned}$$

We estimate each of the integrals above, starting by I_2, I_4 and I_5, which can be estimated as follows:

$$|I_2| \leq \int_{s_\varepsilon}^{s} L_{f_1,K}|y(v)|dv, \tag{7.11}$$

$$|I_4| \leq \int_{s_\varepsilon}^{s} L_{f_2,K}|y(v)|dv, \tag{7.12}$$

$$|I_5| \leq \int_{s_\varepsilon}^{s} L_{f_2,K}|x(v)-\tilde{x}(v)|dv, \tag{7.13}$$

where $L_{f_1,K}, L_{f_2,K}, L_{f_2,K} > 0$ are Lipschitz constants.

Next, we estimate $|I_3|$. We proceed by dividing the interval $\mathcal{I} = [s_\varepsilon, s]$ into sub-intervals of length T and a left over piece:

$$\mathcal{I} = \left(\bigcup_{i=1}^{N(\varepsilon)} [T_{i-1}, T_i]\right) \bigcup [N(\varepsilon)T, s],$$

where $T_i = s_\varepsilon + i\,T$, and $N(\varepsilon)$ is the unique integer such that $N(\varepsilon)T \leq s < N(\varepsilon)T + T$. Then, we divide I_3 into a sum of sub-integrals:

$$I_3 = \sum_{i=1}^{N(\varepsilon)} I_{3,i} + I_{3,N(\varepsilon)+1},$$

where:

$$I_{3,i} = \int_{T_{i-1}}^{T_i} (\tilde{f}_{1,1}(x(s),0,\tau(s)) - \tilde{f}_{1,1}(\tilde{x}(s),0,\tau(s)))\,ds,$$

$$I_{3,N(\varepsilon)+1} = \int_{N(\varepsilon)T}^{s} \tilde{f}_{1,1}(x(v),0,\tau(v)) - \tilde{f}_{1,1}(\tilde{x}(v),0,\tau(v))\,dv.$$

The term $I_{3,N(\varepsilon)+1}$ can be bounded independently from ε as follows:

$$\left|I_{3,N(\varepsilon)+1}\right| \leq 2B_{f_1,K}T.$$

Using Hadamard's lemma, we obtain:

$$I_{3,i} = \int_{T_{i-1}}^{T_i} F_1(x(s),\tilde{x}(s),s)(x(s)-\tilde{x}(s))ds,$$

where the matrix valued map F_1 is given by:

$$F_1(x,\tilde{x},s) = \int_0^1 J_1\tilde{f}_{1,1}(\tilde{x}+\lambda(x-\tilde{x}),\tau(s))d\lambda.$$

Through adding and subtracting a term, we may write:

$$I_{3,i} = \int_{T_{i-1}}^{T_i} F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),s)\,(x(s)-\tilde{x}(s))ds + \int_{T_{i-1}}^{T_i} \Delta_i\left[F_1\right](s)\,(x(s)-\tilde{x}(s))ds,$$

where the term $\Delta_i[F_1]$ is given by:

$$\Delta_i\left[F_1\right](s) = F_1(x(s),\tilde{x}(s),s) - F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),s).$$

Next, since the matrix-valued map F_1 is periodic with zero average over its third argument when the other arguments are fixed, we have that

$$\int_{T_{i-1}}^{T_i} F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),s)\,w\,ds = 0,$$

for any fixed w. Thus, we may write:

$$I_{3,i} = \int_{T_{i-1}}^{T_i} \Delta_i\left[F_1\right](x(s)-\tilde{x}(s)) + \int_{T_{i-1}}^{T_i} F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),s)\Delta_i[x-\tilde{x}]ds,$$

where $\Delta_i[x-\tilde{x}] = (x(s)-x(T_{i-1})) - (\tilde{x}(s)-\tilde{x}(T_{i-1}))$. An application of the fundamental theorem of calculus yields:

$$\begin{aligned}\Delta_i[x-\tilde{x}] =& \varepsilon\int_{T_{i-1}}^{s} (\tilde{f}_{1,1}(x(v),y(v),v) - \tilde{f}_1(\tilde{x}(v),v))dv + O(\varepsilon^2)\\ =& \varepsilon\int_{T_{i-1}}^{s} (\tilde{f}_{1,1}(x(v),y(v),v) - \tilde{f}_{1,1}(x(v),0,v))dv\\ &+ \varepsilon\int_{T_{i-1}}^{s} (\tilde{f}_1(x(v),v) - \tilde{f}_1(\tilde{x}(v),v))dv + O(\varepsilon^2).\end{aligned}$$

Through integration by parts, we obtain:

$$\begin{aligned}\int_{T_{i-1}}^{T_i} F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),s)\Delta_i[x-\tilde{x}]ds =& \left.I_{F,i}(s)\,\Delta_i[x-\tilde{x}]\right|_{s=T_{i-1}}^{s=T_i}\\ &- \varepsilon\int_{T_{i-1}}^{T_i} I_{F,i}(s)\,\Delta[\tilde{f}_1]ds\\ &- \varepsilon\int_{T_{i-1}}^{T_i} I_{F,i}(s)\,\Delta[\tilde{f}_{1,1}]ds + O(\varepsilon^2),\end{aligned}$$

where $\Delta[\tilde{f}_{1,1}]$, $\Delta[\tilde{f}_1]$, and $I_{F,i}(s)$ are defined by:

$$\begin{aligned}\Delta[\tilde{f}_{1,1}] &= \tilde{f}_{1,1}(x(s),y(s),\tau(s)) - \tilde{f}_{1,1}(x(s),0,\tau(s)),\\ \Delta[\tilde{f}_1] &= \tilde{f}_1(x(s),\tau(s)) - \tilde{f}_1(\tilde{x}(s),\tau(s)),\\ I_{F,i}(s) &= \int_{T_{i-1}}^{s} F_1(x(T_{i-1}),\tilde{x}(T_{i-1}),v)dv.\end{aligned}$$

The boundary term $I_{F,i}(s)\,\Delta_i[x-\tilde{x}]\big|_{s=T_{i-1}}^{s=T_i}$ coming out of the integration by parts vanishes because the right factor vanishes at $s = T_{i-1}$ and the left factor vanishes at $s = T_i$, leaving only the integral terms. Using Lipschitz continuity and boundedness on compact subsets, we see that:

$$\left|\int_{T_{i-1}}^{T_i} I_{F,i}(s)\,\Delta[\tilde{f}_{1,1}]ds\right| \leq \int_{T_{i-1}}^{T_i} M_{I_F,\tilde{f}_{1,1},K}\,|y(s)|ds,$$

$$\left|\int_{T_{i-1}}^{T_i} I_{F,i}(s)\,\Delta[\tilde{f}_1]ds\right| \leq \int_{T_{i-1}}^{T_i} M_{I_F,f_1,K}\,|\Delta[x]|ds,$$

$$\left|\int_{T_{i-1}}^{T_i} \Delta_i\left[F_1\right]\Delta[x]ds\right| \leq \varepsilon\int_{T_{i-1}}^{T_i} L_{F_1,K}\,|\Delta[x]|ds,$$

where $\Delta[x] = x(s)-\tilde{x}(s)$. By utilizing the above estimates, the integrals $I_{3,i}$ can be shown to satisfy the bound:

$$\left|I_{3,i}\right| \leq M_K\,\varepsilon\int_{T_{i-1}}^{T_i} (|\Delta[x]| + |y(s)|)\,ds,$$

for some constant M_K. As a consequence, the integral term I_3 satisfies the bound:

$$|I_3| \leq M_K\,\varepsilon\int_{s_\varepsilon}^{s} (|\Delta[x]| + |y(v)|)\,dv + 2B_{f_1,K}T. \tag{7.14}$$

Combining (7.8), (7.10), (7.11), (7.12), (7.13), and (7.14), in addition to the fact that $s_\varepsilon < s_\epsilon$, for all $\varepsilon \in (0,\varepsilon_4)$, we can show that the following estimate holds:

$$\begin{aligned}|x(s)-\tilde{x}(s)| \leq& (M_{K,1} + M_{K,2}s_\varepsilon + M_{K,3}s_\epsilon\varepsilon^2)\varepsilon\\ &+ M_{K,4}\varepsilon\int_{s_\varepsilon}^{s} |y(v)|dv + M_{K,5}\varepsilon^2\int_{s_\varepsilon}^{s} |x(v)-\tilde{x}(v)|dv,\end{aligned}$$

for some positive constants $M_{K,j}$, $j \in \{1,\dots,5\}$. Using the fact that $|y(s)| < 2\alpha\,\varepsilon^{\frac{3}{2}}$, for all $s \in [s_\varepsilon, s_\epsilon]$ by definition, we obtain that:

$$\int_{s_\varepsilon}^{s} |y(v)|dv \leq 2\alpha\,\tau\,\varepsilon^{\frac{3}{2}} \leq 2\alpha\,s_\epsilon\,\varepsilon^{\frac{3}{2}}.$$

Now, remember that in order to obtain a contradiction we assumed that $s_\epsilon \leq T_f/\varepsilon^2$, and so we will have:

$$|x(s)-\tilde{x}(s)| \leq \delta(\varepsilon) + \int_{s_\varepsilon}^{s} M_{K,5}\varepsilon^2|x(v)-\tilde{x}(v)|dv,$$

where the function $\delta(\varepsilon)$ is given by:

$$\delta(\varepsilon) = M_{K,1}\varepsilon + M_{K,2}s_\epsilon\varepsilon + M_{K,3}T_f\varepsilon + 2M_{K,4}T_f\varepsilon^{\frac{1}{2}}.$$

An application of Grönwall's inequality yields:

$$|x(s)-\tilde{x}(s)| \leq \delta(\varepsilon)e^{M_{K,5}\varepsilon^2 s} \leq \delta(\varepsilon)e^{M_{K,5}\varepsilon^2 s_\epsilon},$$

on the interval $s \in [s_\varepsilon, s_\epsilon]$. Once again, recall that we assumed that $s_\epsilon \leq T_f/\varepsilon^2$, and so we have:

$$|x(s)-\tilde{x}(s)| \leq \delta(\varepsilon)e^{M_{K,5}T_f}.$$

Now, observe that $\lim_{\varepsilon\to 0}\delta(\varepsilon) = 0$, and so we are guaranteed the existence of an $\varepsilon_5 \in (0,\varepsilon_4)$ such that for all $\varepsilon \in (0,\varepsilon_5)$ we have that

$|x(s_\epsilon) - \tilde{x}(s_\epsilon)| \leq \epsilon/2$, which contradicts the definition of s_ϵ. However, this contradicts the assumption that $s_\epsilon \leq T_f/\epsilon^2$, which implies the existence of $\varepsilon_5 \in (0, \varepsilon_4)$, such that for all $\varepsilon \in (0, \varepsilon_5)$, for all $(x(0), y(0)) \in K_1 \times K_2$, we have that $T_f/\varepsilon^2 < s_\epsilon$.

To summarize, we have proven that for all $\varepsilon \in (0, \varepsilon_5)$, $(x(0), y(0)) \in K_1 \times K_2$, for all $s \in [0, T_f/\varepsilon^2]$ we have that:

$$|x(s) - \tilde{x}(s)| \leq \epsilon, \quad |y(s)| \leq \gamma\, |y(0)| e^{-\lambda s} + \alpha\, \varepsilon^{\frac{3}{2}}.$$

The same statement is true if we replaced ϵ with $\epsilon/2$:

$$|x(s) - \tilde{x}(s)| \leq \epsilon/2, \quad |y(s)| \leq \gamma\, |y(0)| e^{-\lambda s} + \alpha\, \varepsilon^{\frac{3}{2}}.$$

Now recall the definition of y, which leads to the bound:

$$|x_2(s) - \varphi_0(x(s))| \leq \gamma |x_2(0) - \varphi_0(x(0))| + \bar{\delta}(\varepsilon),$$
$$\bar{\delta}(\varepsilon) = M_{\varphi,K}\varepsilon + \alpha\, \varepsilon^{\frac{3}{2}},$$

for some $M_{\varphi,K} > 0$. Since $\lim_{\varepsilon \to 0} \bar{\delta}(\varepsilon) = 0$, it follows that there exists $\varepsilon_6 \in (0, \varepsilon_5)$ such that $\bar{\delta}(\varepsilon) \leq \epsilon$. Moreover, it follows from second-order averaging with trade-off (Abdelgalil & Taha, 2022) that there exists $\varepsilon_7 \in (0, \varepsilon_6)$ such that for all $x(0) \in K_1$, for all $s \in [0, T_f/\varepsilon^2]$ we have:

$$|\bar{x}(s) - \tilde{x}(s)| \leq \epsilon/2.$$

Hence, the result follows after an application of the triangle inequality and reversing the time scaling $s = \varepsilon^{-2} t$.

7.3. Proof of Theorems 5.1 and 5.2

In this section, we present the proofs of Theorems 5.1 and 5.2.

Proof of Theorem 5.1. Let $\psi := (x, y)$. Items (1) of Assumptions 5.2 and 5.3 imply the existence of class $\mathcal{K}_\infty$ functions $\hat{\alpha}_1$ and $\hat{\alpha}_2$ such that

$$\hat{\alpha}_1(|\psi|_{\tilde{\mathcal{A}}}) \leq E_\theta(\psi) \leq \hat{\alpha}_2(|\psi|_{\tilde{\mathcal{A}}}), \tag{7.18}$$

for all $\psi \in \hat{C} \cup \hat{D} \cup \hat{G}(\hat{D})$. By Clarke (1990, Prop. 2.2.6), E_θ is regular for each $\theta > 0$, which implies that $\partial E_\theta(\psi) = (1-\theta)\partial V(x) + \theta \partial W(x, y)$ for all $(x, y) \in \mathbb{R}^n$ via (Clarke, 1990, Corollary 2 and 3, pp. 39–40).

By Clarke (1990, Prop. 2.3.15), the regularity of W implies that $\partial W(x, y) \subset \partial_x W(x, y) \times \partial_y W(x, y)$. Thus, we have that for all $\psi \in \hat{C}$:

$$\max_{e \in \partial E_\theta(\psi)} \langle e, \hat{f} \rangle \leq (1-\theta) \max_{v \in \partial_x V(x)} \langle v, \hat{f}_x \rangle + \theta \max_{v \in \partial_x W(x,y)} \langle v, \hat{f}_x \rangle + \varepsilon^{-1}\theta \max_{v \in \partial_y W(x,y)} \langle v, \hat{f}_y \rangle$$

for all $\hat{f} \in \hat{F}_\varepsilon(x, y) := \hat{F}_x(x, y) \times \frac{1}{\varepsilon}\hat{F}_y(x, y, \varepsilon)$, where $\hat{f} = (\hat{f}_x, \frac{1}{\varepsilon}\hat{f}_y)$. Adding and subtracting terms, and using item (b) in Assumption 5.5, we obtain that for all $\psi \in \hat{C}$:

$$\begin{aligned}\max_{e \in \partial E_\theta(\psi)} \langle e, \hat{f} \rangle \leq{}& (1-\theta) \max_{v \in \partial_x V(x)} \langle v, \tilde{f}_x \rangle \\ &+ \theta \max_{v \in \partial_x W(x,y)} \langle v, \hat{f}_x \rangle + \varepsilon^{-1}\theta \max_{v \in \partial_y W(x,y)} \langle v, \hat{f}_y \rangle \\ &+ (1-\theta)\left(\max_{v \in \partial_x V(x)} \langle v, \hat{f}_x \rangle - \max_{v \in \partial_x V(x)} \langle v, \tilde{f}_x \rangle \right),\end{aligned}$$

where $\tilde{f}_x \in \tilde{F}(x)$ comes from (5.7). By definition, if $\hat{f}_y \in \hat{F}_y(x, y, \varepsilon)$, then $\hat{f}_y = f_z - \varepsilon J \varphi_0(x) f_x$, where $f_x \in F_x(x, y+\varphi_0(x))$, and $f_z \in F_z(x, y+\varphi_0(x))$. Therefore, the above inequalities imply that for all $\psi \in \hat{C}$:

$$\begin{aligned}\max_{e \in \partial E_\theta(\psi)} \langle e, \hat{f} \rangle \leq{}& (1-\theta) \max_{v \in \partial_x V(x)} \langle v, \tilde{f}_x \rangle \\ &+ \theta \max_{v \in \partial_x W(x,y)} \langle v, \hat{f}_x \rangle + \varepsilon^{-1}\theta \max_{v \in \partial_y W(x,y)} \langle v, f_z \rangle \\ &- \theta \max_{v \in \partial_y W(x,y)} \langle v, J\varphi_0(x) f_x \rangle \\ &+ (1-\theta)\left(\max_{v \in \partial_x V(x)} \langle v, \hat{f}_x \rangle - \max_{v \in \partial_x V(x)} \langle v, \tilde{f}_x \rangle \right),\end{aligned}$$

Since, by construction, $\psi \in \hat{C}$ implies $x \in C_x$, we can use Assumptions 5.2, 5.3, and 5.5, to directly obtain:

$$\begin{aligned}\max_{e \in \partial E_\theta(\psi)} \langle e, \hat{f} \rangle \leq{}& -(1-\theta)k_x \varphi_x^2(x) - \theta\left(\frac{k_y}{\varepsilon} - k_2\right)\varphi_y^2(y) \\ &+ (\theta k_1 + (1-\theta)k_3)\varphi_y(y)\varphi_x(x),\end{aligned}$$

for all $\hat{f} \in \hat{F}_\varepsilon(x, y)$ and all $\psi \in \hat{C}$. We can now use similar computations as in Saberi and Khalil (1984) or Khalil (2002, pp.452) to conclude the existence of $\lambda > 0$ such that for all $\varepsilon \in (0, \varepsilon^*)$ the following holds:

$$\max_{e \in \partial E_{\theta^*}(\psi)} \langle e, \hat{f} \rangle \leq -\lambda\left(\varphi_x^2(x) + \varphi_y^2(y)\right) \leq -\tilde{\rho}(\psi), \tag{7.19}$$

for all $\psi \in \hat{C}$, and all $\hat{f} \in F_\varepsilon(x, y)$, where $\tilde{\rho} \in \mathcal{PD}(\tilde{\mathcal{A}})$. Since, by assumption, we have that:

$$E_{\theta^*}(g) - E_{\theta^*}(\psi) \leq -\hat{\rho}(\psi), \tag{7.20}$$

for all $g \in \hat{G}(x, y)$ and all $(x, y) \in \hat{D}$, it follows that the right hand sides of (7.19) and (7.20) can be upper bounded with $-\rho(\psi)$, where $\rho(\psi) := \min\{\tilde{\rho}(\psi), \hat{\rho}(\psi)\}$. These two inequalities, combined with (7.18), establish that E_{θ^*} is a hybrid Lyapunov function for the shifted SP-HDS (5.2) with respect to the compact set $\tilde{\mathcal{A}}$ via Sanfelice (2021, Thm 3.19). Therefore, $\tilde{\mathcal{A}}$ is UGAS for the hybrid system (5.2). ■

Proof of Theorem 5.2. Under the given assumptions, the function E_{θ^*} still satisfies (7.18). However, (7.19) now holds only with $\tilde{\rho} \in P_s\mathcal{D}(\tilde{\mathcal{A}})$. Similarly, and by construction, for all $\psi = (x, y) \in \hat{D}$ we must have $x \in D_x$. Moreover, also by construction of $\tilde{G}$, for all $(x, y) \in \hat{D}$ and all $\hat{g} = (g_x, g_z) \in \hat{G}(x, y)$ we have $g_x \in \tilde{G}(x)$. Thus, using Assumptions 5.4 and 5.6, we obtain:

$$\begin{aligned}E_{\theta^*}(\hat{g}) &\leq (1-\theta^*)\big(V(x) - c_x\rho_x(x)\big) + \theta^* W(x, y) + \theta^* k_4 \rho_4(x) \\ &= E_{\theta^*}(\psi) - c\rho_x(x) \\ &\leq E_{\theta^*}(\psi),\end{aligned}$$

for all $\hat{g} \in \hat{G}(x, y)$ and all $(x, y) \in \hat{D}$, where we used $c := (1-\theta^*)c_x - \theta^* k_4$, the definition of θ^*, the assumption that $\rho_x = \rho_4$, and the fact that $c > 0$ due to condition (b)-(1) of the theorem. For the case when Assumptions 5.7 and 5.8 hold with $\rho_y = \rho_5$, we have

$$\begin{aligned}E_{\theta^*}(\hat{g}) &\leq (1-\theta^*)\big(V(x) + k_5\rho_5(y)\big) + \theta^* W(x, y) - \theta^* c_y \rho_y(y) \\ &= E_{\theta^*}(\psi) - \tilde{c}\rho_5(y) \\ &\leq E_{\theta^*}(\psi),\end{aligned}$$

for all $\hat{g} \in \hat{G}(x, y)$ and all $(x, y) \in \hat{D}$, where $\tilde{c} := \theta^* c_y - (1-\theta^*)k_5$, which is positive due to condition (b)-(2) of the theorem. Thus, in both of the above cases we have

$$\max_{e \in \partial E_{\theta^*}(\psi)} \langle e, \hat{f} \rangle \leq 0, \quad \forall\, (x, y) \in \hat{C},\ \hat{f} \in \hat{F}_\varepsilon(x, y),$$
$$E_{\theta^*}(\hat{g}) - E_{\theta^*}(\psi) \leq 0, \quad \forall\, (x, y) \in \hat{D},\ \hat{g} \in \hat{G}(x, y),$$

for all $\varepsilon \in (0, \varepsilon^*)$, where as before $F_\varepsilon(x, y) := \hat{F}_x(x, y) \times \varepsilon^{-1}\hat{F}_y(x, y, \varepsilon)$. Since, by condition (c) of the theorem, there are no complete solutions ψ that remain in a non-zero level set of E_{θ^*}, UGAS of the compact set $\tilde{\mathcal{A}}$ follows by leveraging the fact that the HDS (5.2) satisfies the hybrid basic conditions, and by invoking the hybrid invariance principle Sanfelice (2021, Thm 3.23). ■

7.4. Proof of Theorem 6.1

To simplify notation, in this section we use $\Gamma := C_q \cup D_q$, and $\psi := (x, q)$. The proofs follow similar ideas as in Sastry and Bodson (1989) and Teel and Nešić (2010), and rely on the C^1 functions $\omega_\varepsilon, \omega_0 : (\mathbb{R}^n \times \Gamma) \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, defined as follows:

$$\omega_\varepsilon(\psi, \tau) := \int_0^\tau d(\psi, s) e^{-\varepsilon(\tau - s)} \mathrm{d}s, \tag{7.21a}$$

$$\omega_0(\psi, \tau) := \int_0^\tau d(\psi, s) \mathrm{d}s, \tag{7.21b}$$

where d was defined after Definition 6.1. Before presenting the proof of Theorem 6.1, we establish a key technical lemma.

Lemma 7.1. *Suppose that Assumptions 6.1–6.2 hold. Then, there exists a class $\mathcal{K}$ function $\xi(\cdot)$ such that:*

$$|\varepsilon\omega_\varepsilon(\psi,\tau)| \le \xi(\varepsilon)|x| \tag{7.22a}$$

$$\left|\nabla_\tau\omega_\varepsilon(\psi,\tau) - d(\psi,\tau)\right| \le \xi(\varepsilon)|x| \tag{7.22b}$$

$$\left|\varepsilon J_\psi\omega_\varepsilon(\psi,\tau)\right| \le \xi(\varepsilon), \tag{7.22c}$$

for all $(\psi,\tau)\in(\mathbb{R}^n\times\Gamma)\times\mathbb{R}_{\ge0}$. Moreover, $\omega_\varepsilon(\psi,0)=0$, for all $\psi\in\mathbb{R}^n\times\Gamma$. □

The proof of Lemma 7.1 parallels the steps of Sastry and Bodson (1989) and Teel and Nešić (2010). We proceed to prove each of the inequalities in (7.22) under the assumption that $\varepsilon\in(0,\varepsilon_0)$.

Proof of (7.22a). Using inequality (6.5), we have that $\forall\ (\psi,\tau)\in(\mathbb{R}^n\times\Gamma)\times\mathbb{R}_{\ge0}$ and all $\tau'\in\mathbb{R}_{\ge0}$, we have

$$\left|\int_{\tau'}^{\tau'+\tau} d(\psi,s)\mathrm{d}s\right| \le \gamma(\tau)\tau|x|. \tag{7.23}$$

Therefore, the difference $\Delta\omega_0(\tau+\tau',\tau') := \omega_0(\psi,\tau+\tau') - \omega_0(\psi,\tau')$ satisfies

$$\begin{aligned}|\Delta\omega_0(\tau+\tau',\tau')| &= \left|\int_0^{\tau+\tau'} d(\psi,s)\mathrm{d}s - \int_0^{\tau'} d(\psi,s)\mathrm{d}s\right| \\ &= \left|\int_{\tau'}^{\tau+\tau'} d(\psi,s)\mathrm{d}s\right| \le \gamma(\tau)\tau|x|,\end{aligned} \tag{7.24}$$

for all $\tau,\tau'\ge0$ and all $\psi\in\mathbb{R}^n\times\Gamma$. Using integration by parts with the choice of variables: $v(s)=\int_0^s d(\psi,l)\mathrm{d}l$, $\frac{\mathrm{d}v}{\mathrm{d}s}(s)=d(\psi,s)$, $u(s)=e^{-\varepsilon(\tau-s)}$, and $\frac{\mathrm{d}u(s)}{\mathrm{d}s}=\varepsilon e^{-\varepsilon(\tau-s)}$, we can write (7.21a) as

$$\omega_\varepsilon(\psi,\tau) = \omega_0(\psi,\tau) - \varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}\omega_0(\psi,s)\mathrm{d}s. \tag{7.25}$$

Since $1=e^{-\varepsilon\tau}+\varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}\mathrm{d}s$, it follows that $\omega_0(\psi,\tau)=\omega_0(\psi,\tau)e^{-\varepsilon\tau}+\varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}\omega_0(\psi,\tau)\mathrm{d}s$. Then, (7.25) can be rewritten as

$$\omega_\varepsilon(\psi,\tau) = \omega_0(\psi,\tau)e^{-\varepsilon\tau} + \varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}\Delta\omega_0(\tau,s)\mathrm{d}\tau.$$

Using (7.24) with the fact that $\omega_0(\psi,0)=0$, the above equality can be used to upper-bound ω_ε as follows

$$|\omega_\varepsilon(\psi,\tau)| \le \gamma(\tau)\tau|x|e^{-\varepsilon\tau} + \varepsilon\int_0^\tau e^{-\varepsilon\nu}\gamma(\nu)\nu|x|\mathrm{d}\nu, \tag{7.26}$$

where we used the change of variables $\nu=\tau-s$. Using $\sigma=\varepsilon\tau$ and $\tilde\sigma=\varepsilon\nu$, we have

$$\begin{aligned}|\omega_\varepsilon(\psi,\tau)| &\le \sup_{\sigma\ge0}\gamma\left(\frac{\sigma}{\varepsilon}\right)\frac{\sigma}{\varepsilon}e^{-\sigma}|x| + \int_0^{\varepsilon\tau}\gamma\left(\frac{\tilde\sigma}{\varepsilon}\right)\frac{\tilde\sigma}{\varepsilon}e^{-\tilde\sigma}|x|\mathrm{d}\tilde\sigma \\ &\le \sup_{\sigma\ge0}\gamma\left(\frac{\sigma}{\varepsilon}\right)\frac{\sigma}{\varepsilon}e^{-\sigma}|x| + \int_0^{\infty}\gamma\left(\frac{\tilde\sigma}{\varepsilon}\right)\frac{\tilde\sigma}{\varepsilon}e^{-\tilde\sigma}|x|\mathrm{d}\tilde\sigma.\end{aligned}$$

Multiplying both sides by $\varepsilon>0$, we obtain

$$|\varepsilon\omega_\varepsilon(\psi,\tau)| \le \sup_{\sigma\ge0}\gamma\left(\frac{\sigma}{\varepsilon}\right)\sigma e^{-\sigma}|x| + \int_0^\infty\gamma\left(\frac{\tilde\sigma}{\varepsilon}\right)\tilde\sigma e^{-\tilde\sigma}|x|\mathrm{d}\tilde\sigma.$$

We can upper-bound this expression as follows:

$$\begin{aligned}|\varepsilon\omega_\varepsilon(\psi,\tau)| \le \Bigg(&\sup_{\sigma\in[0,\sqrt{\varepsilon}]}\left(\gamma\left(\frac{\sigma}{\varepsilon}\right)qe^{-\sigma}\right) + \sup_{\sigma\ge\sqrt{\varepsilon}}\left(\gamma\left(\frac{\sigma}{\varepsilon}\right)\sigma e^{-\sigma}\right) \\ &+ \int_0^{\sqrt{\varepsilon}}\gamma\left(\frac{\tilde\sigma}{\varepsilon}\right)\tilde\sigma e^{-\tilde\sigma}\mathrm{d}\tilde\sigma + \int_{\sqrt{\varepsilon}}^\infty\gamma\left(\frac{\tilde\sigma}{\varepsilon}\right)\tilde\sigma e^{-\tilde\sigma}\mathrm{d}\tilde\sigma\Bigg)|x|.\end{aligned}$$

By the definition of a type $\mathcal{L}$ function, there exists $k>0$ such that $\gamma(\tau)\le k$ for all $\tau\ge0$. Then, using $\sigma e^{-\sigma}\le e^{-1}$ and $\sigma e^{-\sigma}\le\sigma$, we obtain:

$$\begin{aligned}|\varepsilon\omega_\varepsilon(\psi,\tau)| &\le \left(k\sqrt{\varepsilon}+\gamma\left(\frac{1}{\sqrt{\varepsilon}}\right)e^{-1} + k\int_0^{\sqrt{\varepsilon}}\tilde\sigma\mathrm{d}\tilde\sigma\right. \\ &\quad\left.+\gamma\left(\frac{1}{\sqrt{\varepsilon}}\right)\int_{\sqrt{\varepsilon}}^\infty\tilde\sigma e^{-\tilde\sigma}\mathrm{d}\tilde\sigma\right)|x| \\ &\le \left(k\sqrt{\varepsilon}+\gamma\left(\frac{1}{\sqrt{\varepsilon}}\right)(e^{-1}+1+\sqrt{\varepsilon}) + k\frac{\varepsilon}{2}\right)|x| =: \xi(\varepsilon)|x|,\end{aligned}$$

where we used $\int_{\sqrt{\varepsilon}}^\infty\tilde\sigma e^{-\tilde\sigma}\mathrm{d}\tilde\sigma = (\sqrt{\varepsilon}+1)e^{-\sqrt{\varepsilon}} \le (\sqrt{\varepsilon}+1)$. Since γ is continuous, decreasing, and $\lim_{\tau\to\infty}\gamma(\tau)=0$, it follows that $\xi(0)=0$, $\xi(\cdot)$ is continuous and increasing, and therefore that $\xi\in\mathcal{K}$. ■

Proof of (7.22b). The result follows by using the Differentiation under the Integral Sign formula, which states that for any triple of smooth functions $\{c(y,x),a(y),b(y)\}$, with $b(y)\ge a(y)$, the following holds

$$\nabla_y\int_{a(y)}^{b(y)} c(y,x)\mathrm{d}x = c(y,b(y))\nabla_y b(y) - c(y,a(y))\nabla_y a(y) + \int_{a(y)}^{b(y)}\nabla_y c(y,x)\mathrm{d}x.$$

Applying this formula to (7.21), we directly obtain:

$$\begin{aligned}\nabla_\tau\omega_\varepsilon(\psi,\tau) &= \nabla_\tau\int_0^\tau d(\psi,s)e^{-\varepsilon(\tau-s)}\mathrm{d}s \\ &= d(\psi,\tau) - \varepsilon\int_0^\tau d(\psi,s)e^{-\varepsilon(\tau-s)}\mathrm{d}s \\ &= d(\psi,\tau) - \varepsilon\omega_\varepsilon(\psi,\tau).\end{aligned} \tag{7.27}$$

Therefore, using (7.22a), we have:

$$\left|\nabla_\tau\omega_\varepsilon(\psi,\tau) - d(\psi,\tau)\right| \le |\varepsilon\omega_\varepsilon(\psi,\tau)| \le \xi(\varepsilon)|x|,$$

for all $(\psi,\tau)\in\mathbb{R}^n\times\Gamma\times\mathbb{R}_{\ge0}$. ■

Proof of (7.22c). The proof follows almost the same steps as in the proof of inequality (7.22a), but now we leverage inequality (6.6) instead of inequality (6.5). While the steps are repetitive (see also Sastry & Bodson, 1989), we present the proof for the purpose of completeness.

Consider the Jacobian matrices:

$$J_\psi\omega_\varepsilon(\psi,\tau) = \int_0^\tau e^{-\varepsilon(\tau-s)}J_\psi d(\psi,s)\mathrm{d}s, \text{ and}$$

$$J_\psi\omega_0(\psi,\tau) = \int_0^\tau J_\psi d(\psi,s)\mathrm{d}s.$$

Using (6.6), we have that $\forall\ (\psi,T)\in(\mathbb{R}^n\times\Gamma)\times\mathbb{R}_{\ge0}$ and all $\tau\in\mathbb{R}_{\ge0}$ the following holds:

$$\left|\int_\tau^{\tau+T} J_\psi d(\psi,\tau)\right| \le T\gamma(T). \tag{7.28}$$

Thus, we can bound $\Delta J_\psi\omega_0(\tau+\tau',\tau') := J_\psi\omega_0(\psi,\tau+\tau') - J_\psi\omega_0(\psi,\tau')$ as follows:

$$\begin{aligned}|\Delta J_\psi\omega_0(\tau+\tau',\tau')| &= \left|\int_0^{\tau+\tau'} J_\psi d(\psi,\tau)\mathrm{d}\tau - \int_0^{\tau'} J_\psi d(\psi,\tau)\mathrm{d}\tau\right| \\ &= \left|\int_{\tau'}^{\tau+\tau'} J_\psi d(\psi,\tau)\mathrm{d}\tau\right| \le \gamma(\tau)\tau,\end{aligned} \tag{7.29}$$

for all $(\psi,\tau)\in(\mathbb{R}^n\times\Gamma)\times\mathbb{R}_{\ge0}$ and all $\tau'\in\mathbb{R}_{\ge0}$. Using integration by parts with the variables $\frac{\mathrm{d}v}{\mathrm{d}s}(s)=J_\psi d(\psi,s)$, $v(s)=\int_0^s J_\psi d(\psi,l)\mathrm{d}l$, $u(s)=e^{-\varepsilon(\tau-s)}$, $\frac{\mathrm{d}u(s)}{\mathrm{d}s}=\varepsilon e^{-\varepsilon(\tau-s)}$, we now obtain:

$$\begin{aligned}J_\psi\omega_\varepsilon(\psi,\tau) &= \int_0^\tau J_\psi d(\psi,s)\mathrm{d}s - \int_0^\tau\varepsilon e^{-\varepsilon(\tau-s)}\int_0^s J_\psi d(\psi,l)\mathrm{d}l\mathrm{d}s \\ &= J_\psi\omega_0(\psi,\tau) - \varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}J_\psi\omega_0(\psi,s)\mathrm{d}s.\end{aligned}$$

By using $J_\psi\omega_0(\psi,\tau)=\varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}J_\psi\omega_0(\psi,\tau)\mathrm{d}s + J_\psi\omega_0(\psi,\tau)e^{-\varepsilon\tau}$, we can write the above expression as

$$J_\psi\omega_\varepsilon(\psi,\tau) = J_\psi\omega_0(\psi,\tau)e^{-\varepsilon\tau} + \varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}(J_\psi\omega_0(\psi,\tau) - J_\psi\omega_0(\psi,s))\mathrm{d}s.$$

Then, using (7.29) and $J_\psi\omega_0(\psi,0)=0$, we obtain:

$$|J_\psi\omega_\varepsilon(\psi,\tau)| \le \gamma(\tau)\tau e^{-\varepsilon\tau} + \varepsilon\int_0^\tau e^{-\varepsilon(\tau-s)}(t-\tau)\gamma(\tau-s)\mathrm{d}s.$$

From this point, the procedure follows the same steps as in the proof of inequality (7.22a), with the substitution of $J_\psi\omega_\varepsilon(\psi,\tau)$ for $\omega_\varepsilon(\psi,\tau)$ in (7.26). ■

We are now ready to present the proof of Theorem 6.1. The proof involves two main steps.

Step 1 (Near-Identity transformation): Let ω_ε and ξ come from Lemma 7.1, and consider the transformation $x \to \hat{x}$, given by $\hat{x} = x - \varepsilon\omega_\varepsilon(x,q,\tau)$. Moreover, define $\Phi(x) = \hat{x} + \varepsilon\omega_\varepsilon(x,q,\tau)$, and note that $|\Phi(x) - \Phi(x')| = |\varepsilon\omega_\varepsilon(x,q,\tau) - \varepsilon\omega_\varepsilon(x',q,\tau)| \le \xi(\varepsilon)|x-x'|$, where we used (7.22c). Let $0 < \varepsilon_1 < \varepsilon_0$ be such that $\xi(\varepsilon) < 1$ for all $\varepsilon \in (0,\varepsilon_1]$, which exists since $\xi \in \mathcal{K}$. Then, $\Phi(x)$ is a contraction (Sastry & Bodson, 1989, pp. 347). Therefore, for all $(\hat{x},q,\tau) \in \mathbb{R}^n \times \Gamma \times \mathbb{R}_{\ge 0}$, there exists a unique fixed point x satisfying $x = \Phi(x) = \hat{x} + \varepsilon\omega_\varepsilon(x,q,\tau)$. Therefore, the inverse of the map $x \to \hat{x}$ is well-defined for all $(\hat{x},q,\tau) \in \mathbb{R}^n \times \Gamma \times \mathbb{R}_{\ge 0}$, thus making the map a bijection. The smoothness properties of ω_ε imply that the bijection $x \to \hat{x}$ is a homeomorphism for all $\varepsilon \in (0,\varepsilon_1]$.

Now, using (7.22a) and the triangle inequality, we obtain $|\hat{x}| \le |x| + |\varepsilon\omega_\varepsilon(x,q,\tau)| \le |x| + \xi(\varepsilon)|x| = (1+\xi(\varepsilon))|x|$. Using the reverse triangle inequality, and (7.22a), one has $|x| - |\hat{x}| \le |x - \hat{x}| = |\varepsilon\omega_\varepsilon(x,q,\tau)| \le \xi(\varepsilon)|x|$, which implies $|x| - \xi(\varepsilon)|x| \le |\hat{x}|$. Thus, $(1-\xi(\varepsilon))|x| \le |\hat{x}|$. Combining both bounds, we obtain:

$$(1-\xi(\varepsilon))|x| \le |\hat{x}| \le (1+\xi(\varepsilon))|x|, \ \forall\, x,\hat{x} \in \mathbb{R}^n. \tag{7.30}$$

Step 2 (Lyapunov Analysis): Let (ψ,τ) be a solution to (6.1a)–(6.1b), with $\psi = (x,q)$, and consider the signal $\hat{\psi} : \mathrm{dom}(\psi) \to \mathbb{R}^n \times \Gamma$ defined by $\hat{\psi}(t,j) := (\hat{x}(t,j), \hat{q}(t,j))$, where

$$\begin{pmatrix} \hat{x}(t,j) \\ \hat{q}(t,j) \end{pmatrix} = \begin{pmatrix} x(t,j) - \varepsilon\omega_\varepsilon(x(t,j),q(t,j),\tau(t,j)) \\ q(t,j) \end{pmatrix}, \tag{7.31}$$

for all $(t,j) \in \mathrm{dom}(\hat{\psi}) = \mathrm{dom}(\psi)$, and where we assume that $\varepsilon \in (0,\varepsilon_1]$. Since $\omega_\varepsilon(\cdot)$ is continuously differentiable, it follows that $\hat{x}(\cdot,j)$ is locally absolutely continuous for every j such that $I_j := \{t : (t,j) \in \mathrm{dom}(\hat{\psi})\}$ has nonempty interior, and, therefore, that $\hat{\psi}$ is a hybrid arc.

Next, we evaluate the Lyapunov function V, as defined in Assumption 6.3, along the evolution of the hybrid arc $\hat{\psi}$, and we show that it decreases outside an $\mathcal{O}(\varepsilon)$-neighborhood of the set $\mathcal{A}$. Indeed, for all j such that I_j has nonempty interior, and all $t \in I_j$ such that $(t,j) \in \mathrm{dom}(\hat{\psi})$, we have:

$$\dot{V}(\hat{\psi}(t,j)) = \left\langle \nabla V(\hat{\psi}(t,j)), \dot{\hat{\psi}}(t,j) \right\rangle. \tag{7.32}$$

Due to (6.1a) and (7.31), it follows that

$$\dot{\hat{q}}(t,j) = \dot{q}(t,j) \in F_q(q(t,j)) = F_q(\hat{q}(t,j)), \tag{7.33}$$

which holds for almost all $t \in I_j$. Additionally, we have that

$$\begin{aligned} \dot{\hat{x}}(t,j) &= \dot{x}(t,j) - \varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j) \\ &\quad - \varepsilon\nabla_\tau\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\tau}(t,j), \end{aligned}$$

where $\psi(t,j) = (x(t,j),\ q(t,j))$. Using (6.1a) and (7.27), we obtain:

$$\begin{aligned} \dot{\hat{x}}(t,j) &= f(\psi(t,j),\tau(t,j),\varepsilon) - \varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j) \\ &\quad - \nabla_\tau\omega_\varepsilon(\psi(t,j),\tau(t,j)) \\ &= f(\psi(t,j),\tau(t,j),\varepsilon) - \varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j) \\ &\quad - f(\psi(t,j),\tau(t,j),0) + \bar{f}(\psi(t,j)) \\ &\quad + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j)). \end{aligned}$$

Adding and subtracting $\bar{f}(\hat{\psi})$ and using (7.31):

$$\begin{aligned} \dot{\hat{x}}(t,j) &= \bar{f}(\hat{\psi}(t,j)) + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j)) \\ &\quad + f\left(x(t,j),q(t,j),\tau(t,j),\varepsilon\right) - f\left(x(t,j),q(t,j),\tau(t,j),0\right) \\ &\quad + \bar{f}(\hat{x}(t,j) + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j)),\hat{q}) - \bar{f}(\hat{\psi}(t,j)) \\ &\quad - \varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j), \end{aligned}$$

for almost all $t \in I_j$. Therefore, by matching terms, $\dot{\hat{x}}(t,j)$ can be written in compact form as:

$$\dot{\hat{x}}(t,j) = \bar{f}(\hat{x}(t,j),\hat{q}(t,j)) + p_F(t,j), \tag{7.34}$$

where $p_F(t,j) := p_a(t,j) - \varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j)$ Then, using (7.33) and (7.34), it follows from (7.32) that

$$\begin{aligned} \dot{V}(\hat{\psi}(t,j)) &= \left\langle \nabla V(\hat{\psi}(t,j)), \begin{pmatrix} \bar{f}(\hat{\psi}(t,j)), & \dot{q}(t,j) \end{pmatrix} \right\rangle + \langle \nabla_{\hat{x}} V(\hat{\psi}(t,j)),\ p_a(t,j)\rangle \\ &\quad + \varepsilon \left\langle \nabla_{\hat{x}} V(\hat{\psi}(t,j)),\ -J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j) \right\rangle, \end{aligned} \tag{7.35}$$

for almost all $t \in I_j$. Note that $(\bar{f}(\hat{\psi}(t,j)),\ \dot{q}(t,j)) \in \bar{F}(\hat{\psi}(t,j))$ and that $\hat{\psi}(t,j) \in C$, since C is invariant under the transformation (7.31). Then, using Assumption 6.3-(b) and the Cauchy–Schwarz inequality, from (7.35) we obtain that

$$\begin{aligned} \dot{V}(\hat{\psi}(t,j)) &\le -c_4 V(\hat{\psi}(t,j)) + c_3|\hat{\psi}(t,j)|_{\mathcal{A}}^{p-1}|p_a(t,j)| \\ &\quad + \varepsilon c_3|\hat{\psi}(t,j)|_{\mathcal{A}}^{p-1}|J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j)|, \end{aligned} \tag{7.36}$$

for almost all $t \in I_j$. Using the Lipschitz properties in Assumptions 6.1-(d) and 6.2, for all $(x,q,\tau,\varepsilon) \in \mathbb{R}^n \times C \times \mathbb{R}_{\ge 0} \times (0,\varepsilon_1]$:

$$\begin{aligned} &|f(x,q,\tau,\varepsilon) - f(x,q,\tau,0)| \le \varepsilon L_{\varepsilon_0}|x|, \\ &|\bar{f}(\hat{x} + \varepsilon\omega_\varepsilon(x,q,\tau),q) - \bar{f}(\hat{x},q)| \le L_{\mathrm{ave}}|\varepsilon\omega_\varepsilon(x,q,\tau)| \\ &\qquad\qquad \le L_{\mathrm{ave}}\xi(\varepsilon)|x|, \end{aligned}$$

where the last inequality follows from (7.22a). Thus, using (7.30), the term p_a satisfies

$$|p_a| \le \frac{\left((L_{\mathrm{ave}}+1)\xi(\varepsilon) + \varepsilon L_{\varepsilon_0}\right)}{1-\xi(\varepsilon_1)}|\hat{x}|. \tag{7.37}$$

We proceed to bound the last term in (7.35). Using (6.1a), the fact that $f(0,q,\tau,\varepsilon) = 0$, and Assumption 6.1-(c), we obtain $|\dot{\psi}(t,j)| \le |\dot{x}(t,j)| + |\dot{q}(t,j)| \le L_x|x(t,j)| + M$, where $M > 0$ exists due to the compactness of C and the local boundedness of $F_q(\cdot)$. Therefore, using (7.22c):

$$\begin{aligned} \left|\varepsilon J_\psi\omega_\varepsilon(\psi(t,j),\tau(t,j))\dot{\psi}(t,j)\right| &\le \xi(\varepsilon)\,|\dot{\psi}(t,j)| \\ &\le \frac{\xi(\varepsilon)}{1-\xi(\varepsilon_1)}L_x|\hat{x}(t,j)| + \xi(\varepsilon)M. \end{aligned} \tag{7.38}$$

Thus, for all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that I_j has non-empty interior, we have that $p_F(t,j)$ satisfies the bound

$$\begin{aligned} |p_F(t,j)| &\le \left(\frac{(L_{\mathrm{ave}}+L_x+1)\xi(\varepsilon) + \varepsilon L_{\varepsilon_0}}{1-\xi(\varepsilon_1)}\right)|\hat{x}(t,j)| + \xi(\varepsilon)M \\ &\le \eta(\varepsilon)|\hat{x}(t,j)| + \xi(\varepsilon)M, \end{aligned} \tag{7.39}$$

for all $t \in I_j$, and all j in the domain of the solution, and where $\eta \in \mathcal{K}$ since ξ and the identity function are of class $\mathcal{K}$. Using (7.39) together with Assumption 6.3-(a) in (7.36), we obtain that

$$\begin{aligned} \dot{V}(\hat{\psi}(t,j)) &\le -c_4 V(\hat{\psi}(t,j)) + \frac{c_3}{c_1}\eta_F(\varepsilon)V(\hat{\psi}(t,j)) + \eta_F(\varepsilon)M|\hat{\psi}(t,j)|_{\mathcal{A}}^{p-1} \\ &\le -\left(\frac{c_4}{2} - \frac{c_3}{c_1}\eta_F(\varepsilon)\right)V(\hat{\psi}(t,j)) \\ &\quad - \frac{c_4c_1}{2}|\hat{\psi}(t,j)|_{\mathcal{A}}^{p} + \eta_F(\varepsilon)M|\hat{\psi}(t,j)|_{\mathcal{A}}^{p-1}, \end{aligned} \tag{7.40}$$

for almost all $t \in I_j$, and where $\eta_F(\varepsilon) := \max\{\eta(\varepsilon),\xi(\varepsilon)\}$.

Now, we study the change of the Lyapunov function V along the discrete-time evolution of $\hat{\psi}$. To this end, note that $C \cup D$ is invariant under the transformation (7.31) for all $\varepsilon \in (0,\varepsilon_1]$. Therefore, using Assumption 6.3-(a), for all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that $(t,j+1) \in \mathrm{dom}(\hat{\psi})$, it follows that

$$V(\hat{\psi}(t,j+1)) \le c_2|\hat{\psi}(t,j+1)|_{\mathcal{A}}^{p}. \tag{7.41}$$

Now, using (6.1b) and (7.31):

$$\hat{q}(t,j+1) = q(t,j+1) \in G_q(q(t,j)) = G_q(\hat{q}(t,j)). \tag{7.42}$$

Similarly, using (7.31) and (6.1b), we obtain:

$$\begin{aligned}\hat{x}(t,j+1) &= x(t,j+1) - \varepsilon\omega_\varepsilon(\psi(t,j+1),\tau(t,j+1))\\ &= g(\psi(t,j)) - \varepsilon\omega_\varepsilon(\psi(t,j+1),\tau(t,j+1))\\ &= g\Big(\hat{x}(t,j) + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j)),\hat{q}(t,j)\Big)\\ &\quad - \varepsilon\omega_\varepsilon(\psi(t,j+1),\tau(t,j+1)).\end{aligned}$$

Adding and subtracting $g(\hat{\psi}(t,j))$:

$$\begin{aligned}\hat{x}(t,j+1) &= g\Big(\hat{x}(t,j) + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j)),\hat{q}(t,j)\Big)\\ &\quad + g(\hat{\psi}(t,j)) - g(\hat{\psi}(t,j)) - \varepsilon\omega_\varepsilon(\psi(t,j+1),\tau(t,j+1))\\ &= g(\hat{\psi}(t,j)) + p_J(t,j+1),\end{aligned} \tag{7.43}$$

where

$$\begin{aligned}p_J(t,j+1) &= g\Big(\hat{x}(t,j) + \varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j),\hat{q}(t,j))\Big)\\ &\quad - g(\hat{\psi}(t,j)) - \varepsilon\omega_\varepsilon\Big(g(\psi(t,j)),q(t,j+1),\tau(t,j)\Big).\end{aligned} \tag{7.44}$$

Using Jensen's inequality we have

$$|\hat{\psi}(t,j+1)|_{\mathcal{A}}^p \le 2^{p-1}\left|\begin{bmatrix} g(\hat{\psi}(t,j))\\ \hat{q}(t,j+1)\end{bmatrix}\right|_{\mathcal{A}}^p + 2^{p-1}\left|\begin{bmatrix} p_J(t,j+1)\\ 0\end{bmatrix}\right|_{\mathcal{A}}^p.$$

Using $\hat{g}(t,j+1) := (g(\hat{\psi}(t,j)),q(t,j+1))$, and since by Assumption 6.1-(f) and the definition of the average jump map we have $\hat{g}(t,j+1) \in \bar{G}(\hat{\psi}(t,j)) \subset \mathbb{R}^n \times (C_q \cup D_q)$, by using the monomial bounds on V we obtain:

$$V(\hat{\psi}(t,j+1)) \le 2^{p-1}\frac{c_2}{c_1}V(\hat{g}(t,j+1)) + 2^{p-1}c_2|p_J(t,j+1)|^p.$$

Additionally, since $(t,j+1) \in \mathrm{dom}(\hat{\psi})$, by the definition of solutions to hybrid dynamical systems and (6.1b), it follows that $\psi(t,j) \in D$, and, thus, that $\hat{\psi}(t,j) \in D$. Then, by Assumption 6.3-(c), we have:

$$V(\hat{\psi}(t,j+1)) \le 2^{p-1}\frac{c_2}{c_1}c_5V(\hat{\psi}(t,j)) + 2^{p-1}c_2|p_J(t,j+1)|^p. \tag{7.45}$$

We proceed to bound the last term of (7.45). Using (7.44) and Assumption 6.1-(e):

$$\begin{aligned}|p_J(t,j+1)| &\le L_g|\varepsilon\omega_\varepsilon(\psi(t,j),\tau(t,j),q(t,j))|\\ &\quad + \left|\varepsilon\omega_\varepsilon\Big(g(\psi(t,j)),q(t,j+1),\tau(t,j)\Big)\right|.\end{aligned} \tag{7.46}$$

Additionally, using (7.22a) and (6.1b), the last term in (7.46) satisfies:

$$\left|\varepsilon\omega_\varepsilon\Big(g(\psi(t,j)),q(t,j+1),\tau(t,j)\Big)\right| \le \xi(\varepsilon)|g(\psi(t,j))|$$

for all $\varepsilon \in (0,\varepsilon_1]$. Since $g(0,q) = 0$ for all $q \in D_q$, using (6.4) it follows that $|g(\psi(t,j))| \le L_g|x(t,j)|$. Therefore, we obtain that

$$|p_J(t,j+1)| \le 2L_g\xi(\varepsilon)|x(t,j)|,$$

where we used (7.22a) again to upper bound the second term in (7.43). Using (7.30) and $c_J := \frac{2L_g}{1-\xi(\varepsilon_1)}$ we obtain:

$$|p_J(t,j+1)| \le \xi(\varepsilon)c_J|\hat{x}(t,j)| =: \eta_J(\varepsilon)|\hat{x}(t,j)|, \tag{7.47}$$

holds for all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that $(t,j+1) \in \mathrm{dom}(\hat{\psi})$, and where $\eta_J \in \mathcal{K}$ since $\xi \in \mathcal{K}$. Therefore, from (7.45), we obtain that

$$V(\hat{\psi}(t,j+1)) \le \left(2^{p-1}\frac{c_2}{c_1}c_5 + 2^{p-1}\frac{c_2}{c_1}\eta_J(\varepsilon)^p\right)V(\hat{\psi}(t,j)) \tag{7.48}$$

where in the last inequality we used Assumption 6.3-(a).

Let $\varepsilon^\star \in (0,\varepsilon_1)$ be such that $\frac{c_4}{2} - \eta_F(\varepsilon)\frac{c_3}{c_1} > 0$ and $2^{p-1}\lambda + 2^{p-1}\frac{c_2}{c_1}\eta_J(\varepsilon)^p < 1$ for all $\varepsilon \in (0,\varepsilon^\star]$. Such $\varepsilon^\star$ always exists because $\eta_F, \eta_J \in \mathcal{K}$ and the assumption on λ. Then, from (7.40), for almost all $(t,j) \in \mathrm{dom}(\hat{\psi})$ it follows that

$$\dot{V}(\hat{\psi}(t,j)) \le -\alpha_F V(\hat{\psi}(t,j)) - \frac{c_4c_1}{2}|\hat{\psi}(t,j)|_{\mathcal{A}}^p + \eta_F(\varepsilon^\star)c_3M|\hat{\psi}(t,j)|_{\mathcal{A}}^{p-1}, \tag{7.49}$$

where $\alpha_F := (\frac{c_4}{2} - \eta_F(\varepsilon^\star)\frac{c_3}{c_1}) > 0$.

Similarly, from (7.48), for all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that $(t,j+1) \in \mathrm{dom}(\hat{\psi})$, it follows that :

$$V(\hat{\psi}(t,j+1)) \le \alpha_J V(\hat{\psi}(t,j)), \tag{7.50}$$

where $\alpha_J := \left(2^{p-1}\lambda + 2^{p-1}\frac{c_2}{c_1}\psi_J(\varepsilon^\star)^p\right) \in (0,1)$.

Using (7.49) and (7.50), we conclude that for every hybrid arc $\hat{\psi}$, obtained through the transformation (7.31) from a solution (ψ,τ) to (6.1a)–(6.1b), the function V satisfies the following bounds:

$$\dot{V}(\hat{\psi}(t,j)) \le -\alpha_F V(\hat{\psi}(t,j)),$$

for almost all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that $|\hat{\psi}(t,j)|_{\mathcal{A}} \ge \frac{2c_3}{c_4c_1}M\eta_F(\varepsilon^\star)$, and

$$V(\hat{\psi}(t,j+1)) \le \alpha_J V(\hat{\psi}(t,j)),$$

for all $(t,j) \in \mathrm{dom}(\hat{\psi})$ such that $(t,j+1) \in \mathrm{dom}(\hat{\psi})$. Then, using the comparison principle of Cai and Teel (2009, Lemma C.1), and following the same ideas of Cai and Teel (2009, Appendix C), there exists $k > 0$ and $\gamma \in \mathcal{K}$ such that

$$V(\hat{\psi}(t,j)) \le V(\hat{\psi}(0,0))e^{-k(t+j)} + \gamma\left(M\eta_F(\varepsilon^\star)\right),$$

for all $(t,j) \in \mathrm{dom}(\hat{\psi})$. Therefore, using Assumption 6.3-(a), it follows that

$$|\hat{\psi}(t,j)|_{\mathcal{A}} \le \left(\frac{c_2}{c_1}\right)^{\frac{1}{p}} e^{-\frac{k}{p}(t+j)}|\hat{\psi}(0,0)|_{\mathcal{A}} + \gamma\left(M\eta_F(\varepsilon^\star)\right).$$

Then, using (7.30), for all $\varepsilon \in (0,\varepsilon^\star)$ and all solutions (ψ,τ) to (6.1a)–(6.1b), the following bound holds:

$$|\psi(t,j)|_{\mathcal{A}} \le \kappa_1 e^{-\kappa_2(t+j)}|\psi(0,0)|_{\mathcal{A}} + \tilde{\gamma}(M\eta_F(\varepsilon^\star)) \tag{7.51}$$

for all $(t,j) \in \mathrm{dom}((\psi,\tau))$, where $\kappa_1 = \frac{1+\xi(\varepsilon_1)}{1-\xi(\varepsilon_1)}\left(\frac{c_2}{c_1}\right)^{\frac{1}{p}}$, $\tilde{\gamma}(\cdot) = \frac{1}{1-\xi(\varepsilon_1)}\gamma(\cdot)$, and $\kappa_2 = \frac{k}{p}$. The bound (7.51) allows us to obtain Item (a) of Theorem 6.1, by setting $\nu = \tilde{\gamma}(M\eta_F(\varepsilon^\star))$.

Finally, if $J_q f(x,q,\tau,\varepsilon)\dot{q} = 0$ for all $(x,q,\tau,\varepsilon) \in \mathbb{R}^n \times C \times \mathbb{R}_{\ge 0} \times \mathbb{R}_{\ge 0}$, then the last term in (7.38) is zero, and (7.51) holds with $\tilde{\gamma} \equiv 0$. Thus, we obtain Item (b) of Theorem 6.1. This concludes the proof. ■

8. Conclusions

In this paper, we have studied different emerging techniques in the areas of averaging and singular perturbation for the control and optimization of dynamical systems. In particular, *higher-order* averaging and singular perturbation methods were introduced for dynamical systems modeled as multi-time scale ordinary differential equations. Different academic and biologically inspired examples were presented to illustrate the techniques. For dynamical systems that also incorporate discrete-time dynamics, namely, hybrid dynamical systems, we studied global stability results via first-order singular perturbation and averaging theory based on Lyapunov-like conditions. The different results were illustrated in analytical and numerical examples in the areas of optimization, adaptive systems, sampled-data systems, switching systems, hybrid extremum-seeking control, and hybrid vibrational control. The results and discussions can provide guidelines for future research in this field and for the study of open research questions that could be tackled in the future. Examples include the development of higher-order averaging and singular perturbation methods for hybrid dynamical systems, the systematic design and analysis of novel multi-time scale controllers that leverage these techniques, the incorporation of stochastic phenomena, as well as the synergistic use of hybrid control and vibrational control for the solution of model-free stabilization problems in systems that do not admit a single smooth globally stabilizing feedback law. Preliminary results in some of these directions have been pursued in Abdelgalil and Poveda (2023). Moreover, given the ubiquitous nature of multi-time scale dynamics across various practical engineering

and biological systems, the analytical techniques discussed in the paper could also find fruitful applications in diverse domains such as cyber-genetics, robotics, power systems, and learning-based decision-making algorithms, where inherent time scale separations could enable the development of reductions and decompositions that facilitate the synthesis and analysis of algorithms and control systems with provable stability guarantees.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors would like to thank A. Eldesoukey, H. Taha, and M. Krstić for fruitful discussions on multi-time scale dynamical systems, adaptive systems, and extremum-seeking control. The authors would also like to thank the Associate Editor and the anonymous reviewers for their constructive comments.

References

Abdelgalil, M., Aboelkassem, Y., & Taha, H. (2022). Sea urchin sperm exploit extremum seeking control to find the egg. *Physical Review E*, *106*(6), L062401.

Abdelgalil, M., Eldesoukey, A., & Taha, H. (2023). Singularly perturbed averaging with application to bio-inspired 3d source seeking. In *2023 American control conference (ACC)* (pp. 885–890). IEEE.

Abdelgalil, M., & Poveda, J. (2023). On Lie-bracket averaging for a class of hybrid dynamical systems with applications to model-free control and optimization. arXiv preprint arXiv:2308.15732.

Abdelgalil, M., & Taha, H. (2022). Recursive averaging with application to bio-inspired 3-d source seeking. *IEEE Control Systems Letters*, *6*, 2816–2821.

Ahmed-Ali, T., Lamnabhi-Lagarrigue, F., & Khalil, H. (2023). High-gain observer-based output feedback control with sensor dynamic governed by parabolic PDE. *Automatica*, *147*, Article 110664.

Al-Radhawi, M., Sadeghi, M., & Sontag, E. (2022). Long-term regulation of prolonged epidemic outbreaks in large populations via adaptive control: A singular perturbation approach. *IEEE Control Systems Letters*, *6*, 578–583.

Alvarez, L., Dai, L., Friedrich, B., Kashikar, N., Gregor, I., Pascal, R., et al. (2012). The rate of change in ca2+ concentration controls sperm chemotaxis. *Journal of Cell Biology*, *196*(5), 653–663.

Alvarez, L., Friedrich, B., Gompper, G., & Kaupp, U. (2014). The computational sperm cell. *Trends in Cell Biology*, *24*(3), 198–207.

Amelina, N., Granichin, O., & Fradkov, A. (2019). The method of averaged models for discrete-time adaptive systems. *Automation and Remote Control*, *80*, 1755–1782.

Anderson, B., Bitmead, R. R., Johnson, C. R., Jr., Kokotović, P., Kosut, R. L., Mareels, I., et al. (1986). *Stability of adaptive systems: passivity and averaging*. Cambridge, MA: MIT Press.

Anup, S., Verma, A., & Bhatti, T. (2022). Power system transient stability study by involving higher order corrections in improved quadratic Lyapunov function for singular perturbed model of synchronous generator. *IETE Journal of Research*, *68*(3), 1929–1942.

Ariyur, K., & Krstić, M. (2003). *Real-time optimization by extremum-seeking control*. Wiley.

Astrom, K., & Wittenmark, B. (1989). *Adaptive control*. Addison-Wesley Publishing Company.

Baradaran, M., Poveda, J., & Teel, A. (2018). Stochastic hybrid inclusions applied to global almost sure optimization on manifolds. In *2018 IEEE conference on decision and control (CDC)* (pp. 6538–6543). IEEE.

Belgioioso, G., Liao-McPherson, D., de Badyn, M., Bolognani, S., Smith, R., Lygeros, J., et al. (2022). Online feedback equilibrium seeking. arXiv preprint arXiv:2210.12088.

Bellman, R., Bentsman, J., & Meerkov, S. (1986a). Vibrational control of nonlinear systems: Vibrational controllability and transient behavior. *IEEE Transactions on Automatic Control*, *31*(8), 717–724.

Bellman, R., Bentsman, J., & Meerkov, S. (1986b). Vibrational control of nonlinear systems: Vibrational stabilizability. *IEEE Transactions on Automatic Control*, *31*(8), 710–716.

Benosman, M. (2016). *Learning-based adaptive control: an extremum seeking approach - theory and applications*. Cambridge, MA: Butterworth-Heinemann.

Bentsman, J. (1987). Vibrational control of a class of nonlinear systems by nonlinear multiplicative vibrations. *IEEE Transactions on Automatic Control*, *32*(8), 711–716.

Bernard, P., Jebai, A., & Martin, P. (2020). Higher-order singular perturbations for control design with application to the control of induction motors. In *2020 59th IEEE conference on decision and control (CDC)* (pp. 769–776).

Bianchin, G., Cortés, J., Poveda, J., & DallAnese, E. (2022). Time-varying optimization of LTI systems via projected primal–dual gradient flows. *IEEE Transactions on Control of Network Systems*.

Bianchin, G., Poveda, J., & Dall'Anese, E. (2022). Feedback optimization of switched linear systems via continuous-time and hybrid accelerated gradient flows. *Automatica*.

Borkar, V. (2009). *Stochastic approximation: a dynamical systems viewpoint, Vol. 48*. Springer.

Brockett, R., et al. (1983). Asymptotic stability and feedback stabilization. *Differential Geometric Control Theory*, *27*(1), 181–191.

Bullo, F. (2001). Series expansions for the evolution of mechanical control systems. *SIAM Journal on Control and Optimization*, *40*(1), 166–190.

Bullo, F. (2002). Averaging and vibrational control of mechanical systems. *SIAM Journal on Control and Optimization*, *41*(2), 542–562.

Bullo, F., & Lewis, A. (2004). *Geometric control of mechanical systems: modeling, analysis, and design for simple mechanical control systems, Vol. 49*. Springer.

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

Cai, C., Teel, A., & Goebel, R. (2008). Smooth Lyapunov functions for hybrid systems, part II: (pre)asymptotically stable compact sets. *IEEE Transactions on Automatic Control*, *53*(3), 734–748.

Carnevale, G., & Notarstefano, G. (2023). Nonconvex distributed optimization via Lasalle and singular perturbations. *IEEE Control Systems Letters*, *7*, 301–306.

Casau, P., Cunha, R., Sanfelice, R., & Silvestre, C. (2020). Hybrid control for robust and global tracking on smooth manifolds. *IEEE Transactions on Automatic Control*, *65*, 1870–1885.

Cassandras, C., & Lygeros, J. (2010). *Stochastic hybrid systems*. CRC Press.

Chitour, Y., Haidar, I., Mason, P., & Sigalotti, M. (2023). Upper and lower bounds for the maximal Lyapunov exponent of singularly perturbed linear switching systems. *Automatica*, *155*, Article 111151.

Chowdhary, G., & Johnson, E. (2010). Concurrent learning for convergence in adaptive control without persistency of excitation. In *49th IEEE conference on decision and control (CDC)* (pp. 3674–3679). IEEE.

Chowdhury, D., & Khalil, H. (2017). Fast consensus in multi-agent systems with star topology using high gain observers. *IEEE Control Systems Letters*, *1*(1), 188–193.

Christofides, P., & Teel, A. (1996). Singular perturbations and input-to-state stability. *IEEE Transactions on Automatic Control*, *41*(11).

Christofides, P., Teel, A., & Daoutidis, P. (1996). Robust semi-global output tracking for nonlinear singularly perturbed systems. *International Journal of Control*, *65*(4), 639–666.

Clarke, F. (1990). *Optimization and nonsmooth analysis*. Philadelphia: SIAM.

Cochran, J., & Krstić, M. (2009). Nonholonomic source seeking with tuning of angular velocity. *IEEE Transactions on Automatic Control*, *54*(4), 717–731.

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

Coron, J.-M. (1995). On the stabilization in finite time of locally controllable systems by means of continuous time-varying feedback law. *SIAM Journal on Control and Optimization*, *33*(3), 804–833.

Cothren, L., Bianchin, G., & Dall'Anese, E. (2022). Data-enabled gradient flow as feedback controller: Regulation of linear dynamical systems to minimizers of unknown functions. In *Learning for dynamics and control conference* (pp. 234–247). PMLR.

Cuba Samaniego, C., Giordano, G., & Franco, E. (2020). Periodic switching in a recombinase-based molecular circuit. *IEEE Control Systems Letters*, *4*(1), 241–246.

De, A., Burden, S., & Koditschek, D. (2018). A hybrid dynamical extension of averaging and its application to the analysis of legged gait stability. *International Journal of Robotics Research*, *37*(2–3), 266–286.

Deghat, M., Ahmadizadeh, S., Nešić, D., & Manzie, C. (2021). Practical exponential stability and closeness of solutions for singularly perturbed systems via averaging. *Automatica*, *126*, Article 109449.

Del Vecchio, D., Dy, A., & Qian, Y. (2016). Control theory meets synthetic biology. *Journal of the Royal Society Interface*, *13*(120), Article 20160380, Publisher: Royal Society.

Del Vecchio, D., & Slotine, J.-J. (2012). A contraction theory approach to singularly perturbed systems. *IEEE Transactions on Automatic Control*, *58*(3), 752–757.

Dürr, H.-B., Krstić, M., Scheinker, A., & Ebenbauer, C. (2015). Singularly perturbed Lie bracket approximation. *IEEE Transactions on Automatic Control*, *60*(12), 3287–3292.

Dürr, H.-B., Krstić, M., Scheinker, A., & Ebenbauer, C. (2017). Extremum seeking for dynamic maps using Lie brackets and singular perturbations. *Automatica*, *83*, 91–99.

Dürr, H.-B., Stanković, M., Ebenbauer, C., & Johansson, K. (2013). Lie bracket approximation of extremum seeking systems. *Automatica*, *49*(6), 1538–1552.

Dutta, A., Boker, A., & Doan, T. (2022). Convergence rates of distributed consensus over cluster networks: A two-time-scale approach. In *2022 IEEE 61st conference on decision and control (CDC)* (pp. 7035–7040). ISSN: 2576-2370.

Fang, X., Liu, K., Sun, X., & Teel, A. (2020). Lyapunov-based singular perturbation results in the framework of hybrid systems. *IFAC-Papers Online*, *53*(2), 2027–2032.

Feiling, J., Koga, S., Krstić, M., & Oliveira, T. (2018). Gradient extremum seeking for static maps with actuation dynamics governed by diffusion pdes. *Automatica*, *95*, 197–206.

Friedrich, B., & Jülicher, F. (2007). Chemotaxis of sperm cells. *Proceedings of the National Academy of Sciences*, *104*(33), 13256–13261.

Friedrich, B., Riedel-Kruse, I., Howard, J., & Jülicher, F. (2010). High-precision tracking of sperm swimming fine structure provides strong test of resistive force theory. *Journal of Experimental Biology*, *213*(8), 1226–1234.

Frihauf, P., Krstić, M., & Basar, T. (2012). Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, *57*(5), 1192–1207.

Galarza, F., Poveda, J., Bianchi, G., & Dallenese, E. (2021). Extremum seeking under persistent gradient deception: A switching systems approach. *IEEE Control Systems Letters*, 133–138.

Gaudio, J., Annaswamy, A., Bolender, M., Lavretsky, E., & Gibson, T. (2020). A class of high order tuners for adaptive systems. *IEEE Control Systems Letters*, *5*(2), 391–396.

Ghaffari, A., Krstić, M., & Nešić, D. (2012). Multivariable Newton-based extremum seeking. *Automatica*, *48*, 1759–1767.

Ghods, N., & Krstić, M. (2010). Speed regulation in steering-based source seeking. *Automatica*, *46*(2), 452–459.

Goebel, R., Sanfelice, R., & Teel, A. (2009). Hybrid dynamical systems. *IEEE Control Systems Magazine*, *29*(2), 28–93.

Goebel, R., Sanfelice, R., & Teel, A. (2012). *Hybrid dynamical systems: modeling, stability, and robustness*. Princeton University Press.

Grujić, L. (1979). Sets and singularly perturbed systems. *Systems Science*, *5*(4), 327–338.

Grunberg, T., & Del Vecchio, D. (2019). Time-scale separation based design of biomolecular feedback controllers. In *2019 IEEE 58th conference on decision and control (CDC)* (pp. 6616–6621). ISSN: 2576-2370.

Grushkovskaya, V., & Zuyev, A. (2019). Partial stability concept in extremum seeking problems. *IFAC-PapersOnLine*, *52*(16), 682–687.

Grushkovskaya, V., Zuyev, A., & Ebenbauer, C. (2018). On a class of generating vector fields for the extremum seeking problem: Lie bracket approximation and stability properties. *Automatica*, *94*, 151–160.

Guay, M., Vandermeulen, I., Dougherty, S., & McLellan, P. (2015). Distributed extremum-seeking control over networks of unstable dynamic agents. In *9th IFAC symposium on advanced control of chemical processes, Vol. 48* (pp. 693–697). (8).

Hauswirth, A., Bolognani, S., Hug, G., & Dorfler, F. (2020). Timescale separation in autonomous optimization. *IEEE Transactions on Automaton Control*, *66*(2), 611–624.

Heijmans, S., Nešić, D., Postoyan, R., & Heemels, W. (2018). Singularly perturbed networked control systems. *IFAC-PapersOnLine*, *51*(23), 106–111.

Herath, N., & Vecchio, D. (2019). Deterministic-like model reduction for a class of multiscale stochastic differential equations with application to biomolecular systems. *IEEE Transactions on Automatic Control*, *64*(1), 351–358.

Hespanha, J., Liberzon, D., & Morse, A. (1999). Logic-based switching control of a nonholonomic system with parametric modeling uncertainty. *Systems & Control Letters*, *38*(3), 26.

Hespanha, J., & Morse, A. (1999). Stability of switched systems with average dwell-time. In *38th IEEE conf. decision control, Vol. 3* (pp. 2655–2660).

Huang, X., Khalil, H., & Song, Y. (2019). Regulation of nonminimum-phase nonlinear systems using slow integrators and high-gain feedback. *IEEE Transactions on Automatic Control*, *64*(2), 640–653.

Ioannou, P., & Sun, J. (2012). *Robust adaptive control*. Mineola, NY: Dover Publications Inc.

Jaison, S., & Naidu, D. (2019). Integration of life sciences and engineering-optimal control of HIV using time scales. In *2019 IEEE 1st global conference on life sciences and technologies (LifeTech)* (pp. 1–3). IEEE.

Jikeli, J., Alvarez, L., Friedrich, B., Wilson, L., Pascal, R., Colin, R., et al. (2015). Sperm navigation along helical paths in 3D chemoattractant landscapes. *Nature Communications*, *6*(1), 1–10.

Kaupp, U., Solzin, J., Hildebrand, E., Brown, J., Helbig, A., Hagen, V., et al. (2003). The signal flow and motor response controling chemotaxis of sea urchin sperm. *Nature Cell Biology*, *5*(2), 109–117.

Khalil, H. (2002). *Nonlinear systems*. Prentice Hall.

Khalil, H. (2017). High-gain observers in feedback control: Application to permanent magnet synchronous motors. *IEEE Control Systems Magazine*, *37*(3), 25–41.

Kia, S., Cortés, J., & Martínez, S. (2013). Singularly perturbed algorithms for dynamic average consensus. In *2013 European control conference (ECC)* (pp. 1758–1763). IEEE.

Kokotović, P., Khalil, H., & O'Reilly, J. (1986). *Singular perturbation methods in control: analysis and design*. Academic Press.

Kokotović, P., O'Malley, R., Jr., & Sannuti, P. (1976). Singular perturbations and order reduction in control theory—an overview. *Automatica*, *12*(2), 123–132.

Kokotović, P., & Sannuti, P. (1968). Singular perturbation method for reducing the model order in optimal control design. *IEEE Transactions on Automatic Control*, *13*(4), 377–384.

Kong, F., Ni, H., Zhu, Q., Hu, C., & Huang, T. (2023). Fixed-time and predefined-time synchronization of discontinuous neutral-type competitive networks via non-chattering adaptive control strategy. *IEEE Transactions on Network Science and Engineering*, 1–12.

Kosut, R., Anderson, B., & Mareels, I. (1987). Stability theory for adaptive systems: Method of averaging and persistency of excitation. *IEEE Transactions on Automatic and Control*, *32*(1), 26–34.

Krilašević, S. (2023). *Derivative-free equilibrium seeking in multi-agent systems* (Ph.D. thesis), TU Delft.

Krilašević, S., & Grammatico, S. (2023). Learning generalized nash equilibria in monotone games: A hybrid adaptive extremum seeking control approach. *Automatica*, *151*, Article 110931.

Kromer, J., Märcker, S., Lange, S., Baier, C., & Friedrich, B. (2018). Decision making improves sperm chemotaxis in the presence of noise. *PLoS computational biology*, *14*(4), Article e1006109.

Krstić, M., & Wang, H.-H. (2000). Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica-Kidlington*, *36*(4), 595–602.

Kuehn, C., et al. (2015). *Multiple time scale dynamics, Vol, 191*. Springer.

Kutadinata, R., Moase, W., Manzie, C., Zhang, L., & Garoni, T. (2016). Enhancing the performance of existing urban traffic light control through extremum-seeking. *Transportation Research Part C (Emerging Technologies)*, *62*, 1–20.

Labar, C., Ebenbauer, C., & Marconi, L. (2022). Extremum seeking with intermittent measurements: A Lie-brackets approach. *IEEE Transactions on Automatic Control*, *67*(12), 6968–6974.

Le, J., & Teel, A. (2022). Concurrent learning in high-order tuners for parameter identification. In *2022 IEEE 61st conference on decision and control (CDC)* (pp. 2159–2164). IEEE.

Leblanc, M. (1922). Sur l'electrification des chemins de fer au moyen de courants alternatifs de frequence elevee. *Revue générale de l'électricité*, *12*(8), 275–277.

Li, C., Chakrabarti, B., Castilla, P., Mahajan, A., & Saintillan, D. (2022). A chemo-mechanical model of sperm locomotion reveals two modes of swimming. arXiv preprint arXiv:2210.06343.

Liang, S., Zeng, X., & Hong, Y. (2018). Distributed sub-optimal resource allocation over weight-balanced graph via singular perturbation. *Automatica*, *95*, 222–228.

Liberzon, D. (2003). *Switching in systems and control*. Boston, MA: Birkhauser.

Liu, W. (1997a). An approximation algorithm for nonholonomic systems. *SIAM Journal on Control and Optimization*, *35*(4), 1328–1365.

Liu, W. (1997b). Averaging theorems for highly oscillatory differential equations and iterated Lie brackets. *SIAM Journal on Control and Optimization*, *35*(6), 1989–2020.

Maggia, M., Eisa, S., & Taha, H. (2020). On higher-order averaging of time-periodic systems: reconciliation of two averaging techniques. *Nonlinear Dynamics*, *99*(1), 813–836.

Martens, J., & Sutskever, I. (2012). Training deep and recurrent networks with Hessian-free optimization. In *Neural networks: tricks of the trade: second edition* (pp. 479–535). Springer.

Martin, S., Morărescu, I.-C., & Nešić, D. (2016). Time scale modeling for consensus in sparse directed networks with time-varying topologies. In *2016 IEEE 55th conference on decision and control (CDC)* (pp. 7–12).

Mayhew, C. (2010). *Hybrid control for topologically constrained systems* (Ph.D Dissertation), Santa Barbara: University of California.

McCreesh, M., & Cortés, J. (2023). Selective inhibition and recruitment in Linear-Threshold thalamocortical networks. *IEEE Transactions on Control of Network Systems*, 1–12.

Meerkov, S. (1977). Vibrational control theory. *Journal of the Franklin Institute*, *303*(2), 117–128.

Meyer-Bäse, A., Ohl, F., & Scheich, H. (1996). Singular perturbation analysis of competitive neural networks with different time scales. *Neural Computation*, *8*(8), 1731–1742.

Moase, W., & Manzie, C. (2012). Fast extremum-seeking for Wiener-Hammerstein plants. *Automatica*, *48*, 2433–2443.

Moreau, L., & Aeyels, D. (2000). Practical stability and stabilization. *IEEE Transactions on Automatic Control*, *45*(8), 1554–1558.

Moreu, J., & Annaswamy, A. (2021). A stable high-order tuner for general convex functions. *IEEE Control Systems Letters*, *6*, 566–571.

Morin, P., Pomet, J.-B., & Samson, C. (1999). Design of homogeneous time-varying stabilizing control laws for driftless controllable systems via oscillatory approximation of lie brackets in closed loop. *SIAM Journal on Control and Optimization*, *38*(1), 22–49.

Morin, P., & Samson, C. (2009). Control of nonholonomic mobile robots based on the transverse function approach. *IEEE Transactions on Robotics*, *25*(5), 1058–1073.

Mostacciuolo, E., Trenn, S., & Vasca, F. (2017). Averaging for switched DAEs: Convergence, partial averaging and stability. *Automatica*, *82*, 145–157.

Mukherjee, S., Bai, H., & Chakrabortty, A. (2020). On robust model-free reduced-dimensional reinforcement learning control for singularly perturbed systems. In *2020 American control conference (ACC)* (pp. 3914–3919). IEEE.

Murdock, J. (1983). Some asymptotic estimates for higher order averaging and a comparison with iterated averaging. *SIAM Journal on Mathematical Analysis*, *14*(3), 421–424.

Nakakuki, T., & Imura, J.-i. (2020). Finite-time regulation property of DNA feedback regulator. *Automatica*, *114*, Article 108826.

Narang-Siddarth, A., & Valasek, J. (2014). *Nonlinear time scale systems in standard and nonstandard forms: analysis and control*. SIAM.

Nešić, D., & Teel, A. (1999). On averaging and the ISS property. In *Proceedings of the 38th IEEE conference on decision and control (Cat. No. 99CH36304), Vol. 4* (pp. 3346–3351). IEEE.

Nešić, D., Tan, Y., Manzie, C., Mohammadi, A., & Moase, W. (2012). A unifying framework for analysis and design of extremum seeking controllers. In *Proc. of IEEE Chinese control and decision conference* (pp. 4274–4285).

Ochoa, D., Poveda, J., Subbaraman, A., Schmidt, G., & Pour-Safaei, F. (2021). Accelerated concurrent learning algorithms via data-driven hybrid dynamics and non-smooth odes. In *Learning for dynamics and control (L4DC) conference*.

Oliveira, T., & Krstić, M. (2022). *Extremum seeking through delays and PDEs*. SIAM.

Oliveira, T., Krstić, M., & Tsubakino, D. (2016). Extremum seeking for static maps with delays. *IEEE Transactions on Automatic Control*, *62*(4), 1911–1926.

O'Malley, R. (2014). *Historical developments in singular perturbations*. Springer.

Ortmann, L., Maeght, J., Panciatici, P., Dörfler, F., & Bolognani, S. (2022). Online feedback optimization for transmission grid operation. arXiv preprint arXiv:2212.07795.

Pandey, A., & Murray, R. (2023). Robustness guarantees for structured model reduction of dynamical systems with applications to biomolecular models. *International Journal of Robust and Nonlinear Control*, *33*(9), 5058–5086.

Poveda, J. (2023). Averaging in a class of stochastic hybrid dynamical systems with time-varying flow maps. In *2023 IEEE 62nd conference on decision and control (CDC)*. IEEE, arXiv preprint arXiv:2311.13112.

Poveda, J., Benosman, M., Teel, A., & Sanfelice, R. (2021). Robust coordinated hybrid source seeking with obstacle avoidance in multivehicle autonomous systems. *IEEE Transactions on Automatic Control*, *67*(2), 706–721.

Poveda, J., Brown, P., Marden, J., & Teel, A. (2017). A class of distributed adaptive pricing mechanisms for societal systems with limited information. In *56th IEEE conference on decision and control* (pp. 1490–1495).

Poveda, J., & Krstić, M. (2021). Nonsmooth extremum seeking control with user-prescribed fixed-time convergence. *IEEE Transactions on Automatic Control*, *66*(12), 6156–6163.

Poveda, J., Kutadinata, R., Manzie, C., Nešíc, D., Teel, A., & Liao, C. (2018). Hybrid extremum seeking for black-box optimization in hybrid plants: An analytical framework. In *57th IEEE conferece on decision and control* (pp. 2235–2240).

Poveda, J., & Li, N. (2021). Robust hybrid zero-order optimization algorithms with acceleration via averaging in continuous time. *Automatica*, *123*.

Poveda, J., & Quijano, N. (2015). Shahshahani gradient-like extremum seeking. *Automatica*, *58*, 51–59.

Poveda, J., & Teel, A. (2017a). A framework for a class of hybrid extremum seeking controllers with dynamic inclusions. *Automatica*, *76*, 113–126.

Poveda, J., & Teel, A. (2017b). A robust event-triggered approach for fast sampled-data extremization and learning. *IEEE Transactions on Automatic Control*, *62*(10), 4949–4964.

Poveda, J., & Teel, A. (2019). Hybrid mechanisms for robust synchronization and coordination of multi-agent networked sampled-data systems. *Automatica*, *99*, 41–53.

Poveda, J., & Teel, A. (2020). The heavy-ball ode with time-varying damping: persistence of excitation and uniform asymptotic stability. In *Proc. of American control conference* (pp. 773–778).

Praly, L. (2016). Convergence of the gradient algorithm for linear regression models in the continuous and discrete-time cases. In *Int. rep. mines ParisTech*. Centre Automatique et Systèmes, [Online]. Available: https://hal.archives-ouvertes.fr/hal-01423048.

Prieur, C., Goebel, R., & Teel, A. (2007). Hybrid feedback control and robust stabilization of nonlinear systems. *IEEE Transactions on Automatic Control*, *52*(11), 2103–2117.

Prieur, C., Queinnec, I., Tarbouriech, S., Zaccarian, L., et al. (2018). Analysis and synthesis of reset control systems. *Foundations and Trends® in Systems and Control*, *6*(2–3), 117–338.

Rejeb, J., Morărescu, I.-C., Girard, A., & Daafouz, J. (2018). Stability analysis of a general class of singularly perturbed linear hybrid systems. *Automatica*, *90*, 98–108.

Rivera-Ortiz, P., & De Vecchio, D. (2014). Integral action with time scale separation: A mechanism for modularity in biological systems. In *53rd IEEE conference on decision and control* (pp. 49–55). [ISSN: 0191-2216].

Rockafellar, R., & Wets, R. (1998). *Variational analysis*. Springer.

Saberi, A., & Khalil, H. (1984). Quadratic-type Lyapunov functions for singularly perturbed systems. *IEEE Transactions on Automatic Control*, *29*, 542–550.

Sanders, J., Verhulst, F., & Murdock, J. (2007). *Averaging methods in nonlinear dynamical systems, Vol. 59*. Springer.

Sanfelice, R. (2021). *Hybrid feedback control*. Princeton.

Sanfelice, R., Messina, M., Tuna, S., & Teel, A. (2006). Robust hybrid controllers for continuous-time systems with applications to obstacle avoidance and regulation to disconnected set of points. In *Proc. of American control conference* (pp. 3352–3357).

Sanfelice, R., & Teel, A. (2011). On singular perturbations due to fast actuators in hybrid control systems. *Automatica*, 692–701.

Sannuti, P., & Kokotović, P. (1969). Near-optimum design of linear systems by a singular perturbation method. *IEEE Transactions on Automatic Control*, *14*(1), 15–22.

Sarychev, A. (2001). Lie-and chronologico-algebraic tools for studying stability of time-varying systems. *Systems & Control Letters*, *43*(1), 59–76.

Sastry, S., & Bodson, M. (1989). *Adaptive control: stability, convergence, and robustness*. Englewood Cliffs, NJ: Prentice-Hall.

Scheinker, A., & Krstić, M. (2012). Minimum-seeking for clfs: Universal semiglobally stabilizing feedback under unknown control directions. *IEEE Transactions on Automatic Control*, *58*(5), 1107–1122.

Scheinker, A., & Krstić, M. (2014). Extremum seeking with bounded update rates. *Systems & Control Letters*, *63*, 25–31.

Scheinker, A., & Krstić, M. (2017). *Model-free stabilization by extremum seeking*. Springer.

Shujaee, K., & Lehman, B. (1997). Vibrational feedback control of time delay systems. *IEEE Transactions on Automatic Control*, *42*(11), 1529–1545.

Sontag, E. (1989). Smooth stabilization implies coprime factorization. *IEEE Transactions on Automatic Control*, *34*, 435–443.

Sontag, E. (1999). Stability and stabilization: discontinuities and the effect of disturbances. In *Nonlinear analysis, differential equations and control* (pp. 551–598). Springer.

Sontag, E. (2022). Remarks on input-to-state stability of perturbed gradient flows, motivated by model-free feedback control learning. *Systems & Control Letters*, *161*, Article 105138.

Sootla, A., & Anderson, J. (2017). Structured projection-based model reduction with application to stochastic biochemical networks. *IEEE Transactions on Automatic Control*, *62*(11), 5554–5566.

Su, W., Boyd, S., & Candes, E. (2016). A differential equation for modeling nesterov's accelerated gradient method: Theory and insights. *Journal of Machine Learning Research*, *17*(153), 1–43.

Sussmann, H., & Liu, W. (1991). Limits of highly oscillatory controls and the approximation of general paths by admissible trajectories. In *[1991] Proceedings of the 30th IEEE conference on decision and control, Vol. 1* (pp. 437–442).

Suttner, R., & Dashkovskiy, S. (2017). Exponential stability for extremum seeking control systems. *IFAC-PapersOnLine*, *50*(1), 15464–15470.

Suttner, R., & Dashkovskiy, S. (2022). Robustness and averaging properties of a large-amplitude, high-frequency extremum seeking control scheme. *Automatica*, *136*, Article 110020.

Suttner, R., & Krstić, M. (2022). Source seeking with a torque-controlled unicycle. *IEEE Control Systems Letters*, *7*, 79–84.

Tan, Y., Nešić, D., & Mareels, I. (2006). On non-local stability properties of extremum seeking controllers. *Automatica*, *42*(6), 889–903.

Taringoo, F. (2017). Synchronization on Lie groups: coordination of blind agents. *IEEE Transactions on Automatic Control*, *62*(12), 6324–6338.

Taringoo, F., Dower, P., Nesic, D., & Tan, Y. (2018). Optimization methods on Riemannian manifolds via extremum seeking algorithms. *SIAM Journal on Control and Optimization*, *56*(5), 3867–3892.

Teel, A. (2000). Lyapunov methods in non smooth optimization, part ii: persistently exciting finite differences. In *Proc. of IEEE conference on decision and control* (pp. 118–123).

Teel, A., Forni, F., & Zaccarian, L. (2013). Lyapunov-based sufficient conditions for exponential stability in hybrid systems. *IEEE Transactions on Automatic Control*, *58*(6), 1591–1596.

Teel, A., Moreau, L., & Nešić, D. (2003). A unified framework for input-to-state stability in systems with two time scales. *IEEE Transactions on Automation Control*, *48*, 1526–1544.

Teel, A., & Nešić, D. (2000). Averaging with disturbances and closeness of solutions. *Systems & Control Letters*, *40*(5), 317–323.

Teel, A., & Nešić, D. (2010). Averaging theory for a class of hybrid systems. *Dynamics of Continuous, Discrete and Impulsive Systems*, *17*, 829–851.

Teel, A., Peuteman, J., & Aeyels, D. (1999). Semi-global practical asymptotic stability and averaging. *Systems & Control Letters*, *37*(5), 329–334.

Teel, A., Poveda, J., & Le, J. (2019). First-order optimization algorithms with resets and Hamiltonian flows. In *58th IEEE conference on decision and control* (pp. 5838–5843).

Teel, A., & Praly, L. (2000). A smooth Lyapunov function from a class-KL estimate involving two positive semidefinite functions. *ESAIM: Control, Optimisation and Calculus of Variations*, 313–367.

Tsubakino, D., Oliveira, T., & Krstić, M. (2023). Extremum seeking for distributed delays. *Automatica*, *153*, Article 111044.

Vasil'Eva, A., Butuzov, V., & Kalachev, L. (1995). *The boundary function method for singular perturbation problems*. SIAM.

Vela, P., & Burdick, J. (2003). A general averaging theory via series expansions. In *Proceedings of the 2003 American control conference, 2003, Vol. 2* (pp. 1530–1535).

Volosov, V. (1962). Averaging in systems of ordinary differential equations. *Russian Mathematical Surveys*, *17*(6), 1.

Wang, W., & Nesic, D. (2010). Input-to-state stability and averaging of linear fast switching systems. *IEEE Transactions on Automatic Control*, *55*(5), 1274–1279.

Wang, W., Nesic, D., & Teel, A. (2012). Input-to-state stability for a class of hybrid dynamical systems via averaging. *Mathematics of Control, Signals, and Systems volume*, *23*, 223–256.

Wang, W., Teel, A., & Nešić, D. (2012b). Averaging in singularly perturbed hybrid systems with hybrid boundary layer systems. In *2012 IEEE 51st IEEE conference on decision and control (CDC)* (pp. 6855–6860). IEEE.

Wang, W., Teel, A., & Nešić, D. (2012a). Analysis for a class of singularly perturbed hybrid systems via averaging. *Automatica*, *48*(6).

Wang, X.-F., Teel, A., Sun, X.-M., Liu, K.-Z., & Shao, G. (2023). A distributed robust two-time-scale switched algorithm for constrained aggregative games. *IEEE Transactions on Automatic Control*.

Westervelt, E., Grizzle, J., & Koditschek, D. (2003). Hybrid zero dynamics of planar biped walkers. *IEEE Transactions on Automatic Control*, *48*(1), 42–56.

Whitby, M., Cardelli, L., Kwiatkowska, M., Laurenti, L., Tribastone, M., & Tschaikowski, M. (2022). PID control of biochemical reaction networks. *IEEE Transactions on Automatic Control*, *67*(2), 1023–1030.

Wilson, A., Recht, B., & Jordan, M. (2021). A Lyapunov analysis of accelerated methods in optimization. *Journal of Machine Learning Research*, *22*, 1–34.

Xue, W., Fan, J., Lopez, V., Li, J., Jiang, Y., Chai, T., et al. (2019). New methods for optimal operational control of industrial processes using reinforcement learning on two time scales. *IEEE Transactions on Industrial Informatics*, *16*(5), 3085–3099.

Yang, W., Wang, Y.-W., Wen, C., & Daafouz, J. (2020). Exponential stability of singularly perturbed switched systems with all modes being unstable. *Automatica*, *113*, Article 108800.

Ye, M., Han, Q.-L., Ding, L., & Xu, S. (2023). Distributed Nash equilibrium seeking in games with partial decision information: A survey. *Proceedings of the IEEE*, *111*(2), 140–157.

Yu, H., Koga, S., & Oliveira, T. (2021). Extremum seeking for traffic congestion control with a downstream bottleneck. *Journal of Dynamic Systems, Measurement, and Control*, *143*.

Zand, A., Tavazoei, M., & Kuznetsov, N. (2022). Chaos and its degradation-promoting-based control in an antithetic integral feedback circuit. *IEEE Control Systems Letters*, *6*, 1622–1627.

Zhang, J., & Fridman, E. (2023). Lie-brackets-based averaging of affine systems via a time-delay approach. *Automatica*, *152*, Article 110971.

Zhang, Y., Naidu, D., Cai, C., & Zou, Y. (2014). Singular perturbations and time scales in control theories and applications: An overview 2002–2012. *International Journal of the Information Systems and Science*, *9*(1), 1–36.

Zhang, C., Siranosian, A., & Krstić, M. (2007). Extremum seeking for moderately unstable systems and for autonomous vehicle target tracking without position measurements. *Automatica*, *43*(10), 1832–1839.

Zhu, S., Commuri, S., & Lewis, F. (1994). A singular perturbation approach to stabilization of the internal dynamics of multilink flexible robots. In *Proceedings of 1994 American control conference-ACC'94, Vol. 2* (pp. 1386–1390). IEEE.

Zhu, Y., & Fridman, E. (2022). Extremum seeking via a time-delay approach to averaging. *Automatica*, *135*, Article 109965.