Title

Robust Feedback Mechanisms for Fast and Global Decision-Making and Control

Author

Ochoa Tamayo, Daniel Esteban

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Robust Feedback Mechanisms for Fast and Global Decision-Making and Control

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Electrical Engineering
(Intelligent Systems, Robotics, & Control)

by

Daniel Esteban Ochoa Tamayo

Committee in charge:

Professor Jorge I. Poveda, Chair
Professor Jorge Cortés
Professor Miroslav Krstić
Professor Jun-Kun Wang
Professor Yang Zheng

2024

The Dissertation of Daniel Esteban Ochoa Tamayo is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego
2024

DEDICATION

To my mother Zoraida, and my father Julio.

EPIGRAPH

Human beings are not born once and for all
on the day their mothers give birth to them,
but ... life obliges them over and over again
to give birth to themselves.

Gabriel García Marquéz

Five minutes are enough to dream a whole life,
that is how relative time is.

Mario Benedetti

Utopia is on the horizon.
I move two steps closer;
it moves two steps further away.
I walk another ten steps and the horizon runs ten steps further away.
As much as I may walk, I'll never reach it.
So what's the point of utopia?
The point is this: to keep walking.

Fernando Birri

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This dissertation represents not only personal effort but also the collective influence of those who have shaped my intellectual and personal growth over the past five years. Your impact resonates throughout these pages.

I extend my deepest gratitude to my advisor, Professor Jorge I. Poveda, for his invaluable guidance, patience, and encouragement throughout my Ph.D. journey. His mentorship has been instrumental in shaping my research and academic development. I particularly appreciate his guidance in crafting compelling research narratives and his emphasis on seeking fundamental insights that make our work both interesting and rigorous. Our collaboration has been a mutual learning experience, and the path he is forging serves as an inspiration for my own career.

I am profoundly grateful to my dissertation committee members: Professors Jorge Cortés, Miroslav Krstić, Jun-Kun Wang, and Yang Zheng. Their insightful comments and constructive suggestions have significantly enriched this dissertation and broadened my academic horizons.

I extend my heartfelt thanks to all my coauthors for their invaluable contributions and thought-provoking discussions. These collaborations have significantly enhanced my research and this dissertation. The following sections detail the specific contributions and reprints from our collaborative work:

Chapter 2 incorporates material from our work on momentum-based Nash set-seeking. Sections 2.1, 2.2, and 2.3 are reprints of Examples 1, 2, and 3, respectively, from "Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. The dissertation author is the first author of this paper.

Chapter 3 explores accelerated continuous-time approximate dynamic programming. Section 3.2, in full, is a reprint of the material as it appears in "Accelerated Continuous-Time Reinforcement Learning via Data-Assisted Hybrid Control" [2]. The dissertation

author is the first author of this paper.

Chapter 4 extensively builds upon our work on Nash-Set seeking with momentum. Sections 4.1, 4.2, and 4.3, discussing distributed momentum-based dynamics in potential games, are, in part, a reprint of the material from Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. Section 4.4, in full, is a reprint of the material on decentralized pricing dynamics for congestion games from High-Order Decentralized Pricing Dynamics for Congestion Games: Harnessing Coordination to Achieve Acceleration" [3]. The dissertation author was one of the primary investigators of this paper. Section 4.5, in full, is a reprint of our research on robust optimization over networks using distributed restarting of momentum-based dynamics from "Robust Optimization Over Networks Using Distributed Restarting of Accelerated Dynamics" [4]. The dissertation author is the first author of this paper.

Chapter 5 extends our work on distributed Nash-set seeking dynamics. Section 5.1, in full, is a reprint of the material from Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. Section 5.2, in full, is a reprint of our work on decentralized concurrent learning with momentum and restart from Decentralized Concurrent Learning with Momentum and Restart" [5]. The dissertation author is the first author of these papers.

Chapter 6, which explores data-driven hybrid dynamics and nonsmooth ODEs for accelerated concurrent learning, is, in full, a reprint of the material from "Accelerated Concurrent Learning Algorithms via Data-Driven Hybrid Dynamics and Nonsmooth ODEs" [6]. The dissertation author is the first author of this paper.

Chapter 7, discussing high-performance optimal incentive-seeking in transactive control for traffic congestion, is, in full, a reprint of the material from "High-Performance Optimal Incentive-Seeking in Transactive Control for Traffic Congestion" [7]. The dissertation author is the first author of this paper.

Chapter 8, which delves into dynamic gains for asymptotic-behavior shaping in

hybrid dynamic inclusions, has been accepted for publication in the proceedings of the 2024 IEEE Conference on Decision and Control. The chapter will appear as a paper titled "Dynamic Gains for Asymptotic-Behavior Shaping in Hybrid Dynamic Inclusions" [8]. The dissertation author is the first author of this paper.

Chapter 9, examining prescribed-time stability in switching systems with resets, is, in full, a reprint of the material from "Prescribed-Time Stability in Switching Systems with Resets: A Hybrid Dynamical Systems Approach" [9]. The dissertation author is the first author of this paper.

Chapter 10, exploring robust global optimization on smooth compact manifolds via hybrid gradient-free dynamics, is, in full, a reprint of the material from "Robust Global Optimization on Smooth Compact Manifolds via Hybrid Gradient-Free Dynamics" [10]. The dissertation author is the first author of this paper.

Finally, Chapter 11 contains the unpublished material "On Hybrid Vibrational Control: Robust Global Stabilization of the Kapitza's Pendulum" coauthored with Professor J. I. Poveda [11]. The dissertation author is one of the primary investigators of this material.

In each of these projects, I took a leading role in both the research and writing processes, drawing on the expertise and insights of my coauthors. I am particularly grateful to Gerd S. Schmidt, Anantharam Subbaraman, and Farshad R. Pour-Safaei for their contributions to our work on accelerated concurrent learning algorithms. I extend my thanks to Yilan Chen and Jason Marden for our productive collaboration on decentralized pricing dynamics in congestion games. I am also thankful to Nicolas Espitia for his valuable insights on prescribed-time stability in switching systems, and to Muhammad Umar Javed and Xudong Chen for our joint efforts on decentralized concurrent learning. My gratitude extends to César Uribe for his input on distributed optimization in our work on robust momentum-based methods for consensus optimization.

I owe a special debt of gratitude to Nicanor Quijano, whose encouragement during

my Master's program in Colombia and continued friendship and support over the years have been invaluable.

To my friends who provided companionship from Colombia and elsewhere, I extend my sincerest gratitude. Germán, José Sebastián, Víctor, Sofi, Alejandro, Daniel Esteban, Diego, Oscar, Camilo, Felipe, Nicolás, and Jostein—your encouragement alleviated the challenges of being abroad during difficult times. To Caro, for always posing thought-provoking questions when needed most. To Esteban, for being a constant role model, motivating me to face challenges with a smile.

The past two years at the Adaptive Hybrid Systems Laboratory in San Diego have been demanding yet profoundly instructive. I am indebted to Mahmoud, Adel, and Michael for our stimulating discussions. To Vishal, your willingness to help and constant support have been a source of both joy and intellectual growth.

To Samuel, Leidy, and Yolanda for their unwavering support; even when I struggle to express it, my gratitude for you runs deep. To Andrea, my guiding star and source of joy, for rekindling my zest for life and encouraging me to take pride in my achievements. To my father Julio, for nurturing my dreams and teaching me how a well-crafted plan can turn them into reality. Finally, to my mother Zoraida, for showing me that being hardworking and treating people with kindness can go hand in hand.

VITA

Daniel Esteban Ochoa Tamayo

Education

2024	Ph.D. in Electrical Engineering (Intelligent Systems, Robotics, and Control), University of California San Diego
2022	M.Sc. in Electrical and Computer Engineering, University of Colorado Boulder
2019	M.Sc. in Electronics and Computer Engineering, University of Los Andes, Colombia
2017	B.S. in Physics, University of Los Andes, Colombia
2017	B.S. in Electronics Engineering, University of Los Andes, Colombia

Experience

2022-2024	Graduate Student Researcher, University of California San Diego
2019-2022	Graduate Research Assistant, University of Colorado Boulder
2017-2019	Graduate Student Researcher, University of los Andes, Colombia

Selected Publications

Journal Papers

1. **D. E. Ochoa** and J. I. Poveda, "Robust global optimization on smooth compact manifolds via hybrid gradient-free dynamics," Automatica, vol. 171, p. 111916, 2025.
2. **D. E. Ochoa** and J. I. Poveda, "Momentum-based Nash set-seeking over networks via multitime scale hybrid dynamic inclusions," IEEE Transactions on Automatic Control, vol. 69, no. 7, pp. 4245–4260, 2024.
3. **D. E. Ochoa**, M. U. Javed, X. Chen, and J. I. Poveda, "Decentralized concurrent learning with coordinated momentum and restart," Systems & Control Letters, vol. 193, p. 105931, 2024.
4. **D. E. Ochoa**, N. Espitia, and J. I. Poveda, "Prescribed-time stability in switching systems with resets: A hybrid dynamical systems approach," Systems & Control Letters, vol. 193, p. 105910, 2024.
5. M. Abdelgalil, **D. E. Ochoa**, and J. I. Poveda, "Multi-time scale control and optimization via averaging and singular perturbation theory: From ODEs to hybrid dynamical systems," Annual Reviews in Control, vol. 56, p. 100926, 2023.

6. **D. E. Ochoa**, F. Galarza-Jimenez, F. Wilches-Bernal, D. Schoenwald, J. I. Poveda, "Control Systems for Low-Inertia Power Grids: A Survey on Virtual Power Plants". IEEE Access. 2023.
7. **D. E. Ochoa** and J. I. Poveda, "High-performance optimal incentive-seeking in transactive control for traffic congestion," European Journal of Control, vol. 68, p. 100696, 2022, 2022 European Control Conference Special Issue.
8. **D. E. Ochoa**, J. I. Poveda, C. Uribe, and N. Quijano, "Robust optimization over networks using distributed restarting of accelerated dynamics," IEEE Control Systems Letters, vol. 5, pp. 301–306, 2021.
9. J. Martinez-Piazuelo, **D. E. Ochoa**, N. Quijano, and L. F. Giraldo, "A Multi-Critic Reinforcement Learning Method: An Application to Multi-Tank Water Systems". IEEE Access. 2021.

Conference Papers

1. **D. E. Ochoa** and J. I. Poveda, "Dynamic Gains for Transient-Behavior Shaping in Hybrid Dynamic Inclusions," 63rd IEEE Conference on Decision and Control, to appear, 2024.
2. **D. E. Ochoa**, J. I. Poveda, "Prescribed-Time Concurrent Learning with Switching Datasets", 8th IFAC Conference on Analysis and Design of Hybrid Systems, 2024.
3. Y. Chen, **D. E. Ochoa**, J. R. Marden, and J. I. Poveda, "High-Order Decentralized Pricing Dynamics for Congestion Games: Harnessing Coordination to Achieve Acceleration". American Control Conference. 2023.
4. **D. E. Ochoa** and J. I. Poveda, "Accelerated Continuous-Time Approximate Dynamic Programming via Data-Assisted Hybrid Control". 14th IFAC workshop and Adaptive and Learning Control Systems (ALCOS). 2022.
5. **D. E. Ochoa**, J. I. Poveda, A. Subbaraman, G. S. Schmidt, and F. R. Pour-Safaei, "Accelerated concurrent learning algorithms via data-driven hybrid dynamics and nonsmooth ODEs," in Learning for Dynamics and Control. PMLR, 2021.
6. **D. E. Ochoa**, J. I. Poveda, C. A. Uribe, and N. Quijano, "Hybrid robust optimal resource allocation with momentum". Conference on Decision and Control (CDC). 2019.

ABSTRACT OF THE DISSERTATION

Robust Feedback Mechanisms for
Fast and Global Decision-Making and Control

by

Daniel Esteban Ochoa Tamayo

Doctor of Philosophy in Electrical Engineering
(Intelligent Systems, Robotics, & Control)

University of California San Diego, 2024

Professor Jorge I. Poveda, Chair

The implementation of advanced optimization-based control strategies in complex engineering systems promises significant improvements in efficiency and performance. However, the practical implementation of these strategies often faces substantial challenges, including distributed implementation requirements, nonlinear dynamics, system uncertainties, and insufficient robustness margins. This dissertation addresses these challenges through the lens of hybrid dynamical systems, developing robust and efficient control algorithms for a diverse range of applications. Our research explores four key areas, each offering distinct advantages for various scenarios: momentum-based methods for

distributed optimization and real-time decision-making, nonsmooth approaches for fixed and finite-time stability, time-varying methods for prescribed-time convergence, and global optimization on manifolds.

In the domain of distributed optimization and real-time decision-making, we introduce novel hybrid momentum-based algorithms that overcome the limitations of related purely continuous-time approaches used in optimization, and extend their applicability to both potential and nonpotential game settings. These methods provide accelerated convergence rates over traditional gradient-based approaches, and are particularly effective in scenarios with low curvature.

Complementing this work, we investigate nonsmooth dynamics for fixed and finite-time stability. While momentum methods excel under certain conditions, nonsmooth approaches offer the potential for guaranteed convergence within a finite time, independent of initial conditions. Our approach applies both momentum-based and nonsmooth methods to practical problems such as traffic congestion management and accelerated learning, offering a comprehensive comparison of their respective strengths across different scenarios.

Our research extends to time-varying methods, utilizing suitable dynamic gains to shape the transient behavior of hybrid systems with preexisting uniform asymptotic stability properties. By interconnecting these dynamic gains with the original system, we obtain, in particular, prescribed-time stability results, guaranteeing convergence to a compact set within a user-defined finite time that is independent of initial conditions and problem parameters. We demonstrate the effectiveness of this method in complex scenarios, including systems with intermittent feedback, by applying it to switching systems with resets.

We address the challenge of global optimization on compact manifolds using gradient-free dynamics. Our approach overcomes topological obstacles that typically preclude the implementation of smooth, nonsmooth, and even time-varying techniques in these spaces. By harnessing the flexibility of hybrid systems, we develop mechanisms that not only

render the manifold forward invariant but also have provable nonzero robustness margins for global optimization.

Finally, we address the challenge of robust global stabilization of Kapitza's pendulum's naturally unstable upright position. We propose a novel hybrid control approach that leverages multiple oscillating directions to achieve global asymptotic stability. This method overcomes the limitations of traditional smooth control techniques in addressing the pendulum's complex dynamics, demonstrating the power of hybrid systems in stabilizing counterintuitive nonlinear phenomena.

Throughout this dissertation, we employ tools from hybrid systems theory and Lyapunov stability analysis to provide rigorous convergence guarantees. Our theoretical results are substantiated by extensive numerical simulations, demonstrating the efficacy of the proposed methods across various engineering domains.

Introduction

The implementation of advanced optimization-based control strategies in complex engineering systems promises significant improvements in efficiency and performance. However, the practical implementation of these strategies often faces substantial challenges, including distributed implementation requirements, nonlinear dynamics, system uncertainties, and insufficient robustness margins. This dissertation addresses these challenges through the lens of hybrid dynamical systems, developing robust and efficient control algorithms for a diverse range of applications.

The increasing complexity of modern engineering systems, from power grids [12] to autonomous vehicles [13], demands sophisticated control strategies that can handle multiple objectives, constraints, and uncertainties. Optimization-based control has emerged as a powerful paradigm for addressing these challenges [14, 15]. This approach combines the rigor of mathematical optimization with the practical constraints and objectives of control systems, enabling the design of controllers that can handle multi-objective criteria, system constraints, and uncertainties. As control systems become increasingly distributed, large-scale, and operating in uncertain environments, new challenges arise that require innovative approaches.

Traditional centralized control methods often fall short in these scenarios due to communication limitations, scalability issues, and the need for real-time adaptation. Moreover, the presence of nonlinearities, coupled dynamics, and time-varying parameters further complicates the design and analysis of effective control strategies. The transition

from theory to practice in optimization-based control often encounters obstacles that limit the effectiveness of these methods in real-world scenarios. These include computational complexity and real-time implementation constraints, robustness issues in the presence of model uncertainties and disturbances, challenges in distributed implementation and coordination, difficulties in guaranteeing global convergence (especially in non-convex settings), and the need for fast adaptation and convergence in time-critical applications.

This dissertation aims to address these challenges by developing a comprehensive framework that leverages the power of hybrid dynamical systems. Hybrid systems, which combine continuous-time and discrete-time dynamics, offer a flexible and powerful modeling paradigm that can capture complex behaviors and enable innovative control strategies [16]. By exploiting the interplay between continuous and discrete dynamics, we develop algorithms that achieve superior performance, robustness, and convergence properties compared to traditional approaches.

The overarching goal of this dissertation is to develop a comprehensive framework for addressing the challenges in implementing optimization-based control strategies. Specifically, we aim to develop robust and efficient algorithms for distributed optimization and real-time decision-making, design control strategies that achieve fixed-time and finite-time stability, create methods for prescribed-time convergence in complex systems, and solve global optimization problems on manifolds with provable robustness guarantees.

To achieve these objectives, our research explores four key areas, each offering distinct advantages for various scenarios:

Momentum-Based Methods

Momentum-based methods have gained significant attention in optimization and control due to their ability to accelerate convergence over traditional gradient-based approaches [17, 18, 19]. However, their application in distributed and noncooperative settings, where individual agents aim to individually optimize their local cost functions,

presents unique challenges. To address these challenges, in Chapter 2, we begin by exploring the foundations and limitations of Nesterov's Ordinary Differential Equation (ODE), a continuous-time analog of Nesterov's accelerated gradient method [20]. We identify robustness limitations and challenges in distributed real-time decision-making and nonpotential settings. Chapter 3 introduces hybrid dynamics to overcome these robustness limitations, presenting an accelerated continuous-time approximate dynamic programming approach using data-assisted hybrid control. Chapters 4 and 5 extend these concepts to distributed settings, addressing both potential and nonpotential scenarios. We develop momentum-based algorithms for transactive control in congestion games, distributed consensus optimization, and Nash-set seeking in nonpotential games.

Non-Lipschitz Methods for Fixed and Finite-Time Stability

Non-Lipschitz methods offer powerful tools for achieving fixed and finite-time convergence in control systems. In this part, we focus on two main applications: traffic congestion management and accelerated concurrent learning.

Chapter 6 develops data-driven hybrid dynamics and non-Lipschitz ODEs for accelerated concurrent learning. We propose novel algorithms that achieve finite-time and fixed-time convergence in parameter estimation problems while relaxing persistence of excitation requirements. These methods include:

1. Hybrid momentum-based algorithms with periodic and adaptive restarting
2. Finite-time concurrent learning dynamics
3. Fixed-time concurrent learning dynamics

Our analysis employs tools from Lyapunov theory and hybrid systems to establish rigorous stability and convergence guarantees. These non-Lipschitz methods complement the momentum-based approaches, offering additional tools for achieving rapid and guaranteed convergence in optimization-based control problems.

Chapter 7 introduces high-performance optimal incentive-seeking controllers for transactive control in traffic congestion. We present three variants:

1. A gradient-based controller
2. A hybrid momentum-based controller
3. A fixed-time controller

These approaches leverage non-Lipschitz dynamics to achieve rapid convergence to optimal incentives while guaranteeing closed-loop stability.

Time-Varying Methods

Time-varying methods provide powerful tools for shaping the transient behavior of dynamical systems and achieving prescribed-time stability. In this part, we explore these concepts in the context of hybrid and switching systems. Chapter 8 presents a framework for asymptotic-behavior shaping in hybrid dynamic inclusions using dynamic gains. We demonstrate applications in accelerating gradient flows and hybrid dynamics with momentum. Chapter 9 develops a comprehensive theory of prescribed-time stability in switching systems with resets, using a hybrid dynamical systems approach. We introduce novel concepts such as blow-up average dwell-time conditions and explore applications in control and decision-making with intermittent feedback.

Global Asymptotic Guarantees via Hybrid Dynamical Systems

Achieving global asymptotic stability guarantees is crucial in many control and optimization problems, particularly when dealing with nonconvex landscapes or topological constraints. Chapter 10 addresses this challenge by introducing robust global optimization techniques for smooth compact manifolds using hybrid gradient-free dynamics. This approach overcomes topological obstructions that prevent global convergence using smooth ODEs, with applications in various domains including robotics and attitude control.

Chapter 11 examines Kapitza's pendulum, a classic nonlinear system exhibiting counterintuitive behavior where rapid pivot oscillations render the typically unstable inverted position locally asymptotically stable. We address the unique challenges posed by this system, particularly the topological obstructions preventing global asymptotic stability under traditional smooth control methods. By exploring the mathematical foundations and extending the analysis to arbitrary oscillation directions, we propose a novel approach combining multiple oscillating directions with a carefully designed hybrid dynamical system. This strategy demonstrates the power of hybrid control in addressing limitations of nonlinear systems and provides insights into stabilizing systems with intrinsic topological constraints.

Throughout this dissertation, we emphasize the importance of rigorous stability analysis and performance guarantees. Our approach integrates tools from optimization theory, control theory, and dynamical systems to develop algorithms that are both theoretically sound and practically implementable. This work aims to bridge the gap between theoretical advancements and practical applications, addressing the challenges posed by complex, uncertain, and distributed environments.

To establish a solid foundation for the subsequent technical discussions, the next chapter provides a comprehensive introduction to the fundamental concepts and notation of hybrid dynamical systems. This chapter presents essential preliminaries, including formal definitions, solution concepts, and stability notions, which are extensively utilized in the analysis and design of the algorithms presented in later chapters.

Chapter 1

Preliminaries on Hybrid Dynamical Systems

This chapter introduces the fundamental concepts and notations used throughout this dissertation, with a focus on hybrid dynamical systems. We begin by establishing the mathematical notation and definitions for sets, functions, and stability notions that form the foundation of our work. The chapter then delves into the formal framework of hybrid dynamical systems, presenting their basic structure and solution concepts. We discuss various formulations of HDS, including those with time-varying flows and external inputs, which are crucial for modeling the algorithms developed in subsequent chapters. Finally, we introduce key stability and convergence notions for HDS, ranging from uniform global asymptotic stability to semi-global practical stability. These concepts are essential for analyzing the performance and robustness of the optimization-based control algorithms presented in this thesis.

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 $|x|_{\mathcal{A}} := \inf_{y \in \mathcal{A}} |x - y|$, where $|\cdot|$ denotes the standard norm in $\mathbb{R}^n$. Unless stated otherwise, we assume all vectors are column vectors. If $x_i \in \mathbb{R}^{n_i}$ for $i \in \{1, \ldots, k\}$ are vectors, we use $(x_1, \ldots, x_k) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ to denote their concatenation, i.e., $(x_1, x_2, \cdots x_k) \coloneqq [x_1^\top, x_2^\top, \cdots x_k^\top]^\top$. We use $\mathbb{S}^1 \subset \mathbb{R}^2$ to denote the unit circle centered at

the origin, $\mathbb{T}^m \coloneqq \underbrace{\mathbb{S}^1 \times \mathbb{S}^1 \times \ldots \mathbb{S}^1}_{m \text{ times}}$. We use $\mathbb{B}$ 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 $\mathcal{C}^k$ for $k \in \mathbb{N}_{\geq 0}$ if it is k-times continuously differentiable with locally Lipschitz derivatives. We use $\frac{\partial f}{\partial x}$ to denote the Jacobian of a continuously differentiable function $f : \mathbb{R}^m \ni x \to f(x) \in \mathbb{R}^n$. When $n = 1$, we use ∇f to denote the gradient of f and define $\left(\frac{\partial f}{\partial x}(x)\right)^\top = \nabla f(x)$ for all $x \in \mathbb{R}^n$.

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 $\overline{\text{co}}\ \mathcal{X}$ to denote the closed convex hull of $\mathcal{X}$, $\overline{\mathcal{X}}$ to denote the closure of $\mathcal{X}$, and $\text{int}(\mathcal{X})$ to denote its topological 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) non-increasing, 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 $\gamma \in \mathcal{PD}$.

Hybrid Dynamical Systems

In this dissertation, we will model our algorithms as Hybrid Dynamical Systems with time-varying flows, of the form (see [16]):

$$x \in C, \qquad \dot{x} \in F(x), \tag{1.1a}$$

$$x \in D, \qquad x^+ \in G(x), \tag{1.1b}$$

where $x \in \mathbb{R}^n$ is the main state of the system, s is an auxiliary state used to model the evolution of the continuous time, $F : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called the flow map, and $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called the jump map. The sets C and D, called the flow set and the jump set, respectively, condition the points in $\mathbb{R}^n$ where the system can *flow* or *jump* via equations (1.1a) or (1.1b), respectively. In this way, the HDS can be represented by the notation $\mathcal{H} := (C, F, D, G)$. Systems of the form (1.1) can be seen as generalizations of purely continuous-time systems ($D = \emptyset$) and purely discrete-time systems ($C = \emptyset$). Solutions to HDS of the form (1.1) are parameterized by both a continuous-time index $t \in \mathbb{R}_{\geq 0}$, which increases continuously during flows, and a discrete-time index $j \in \mathbb{Z}_{\geq 0}$, which increments by one during jumps. Thus, the notation $\dot{x}$ in (1.1a) represents the derivative $\frac{dx(t,j)}{dt}$; and x^+ in (1.1b) represents the value of x after an instantaneous jump, i.e., $x(t, j+1)$. Naturally, solutions $(x, s) : \mathrm{dom}(x, s) \to \mathbb{R}^n$ to (1.1) are defined on hybrid time domains.

Definition 1.1. (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 $0 = t_0 \leq t_1 \ldots \leq t_J$. The set E is a HTD if for all $(T, J) \in E$, $E \cap ([0, T] \times \{0, \ldots, J\})$ is a compact HTD. Given a HTD E, we use $\sup_t E := \sup \{t \in \mathbb{R}_{\geq 0} : \exists\, j \in \mathbb{Z}_{\geq 0}, \text{ such that } (t, j) \in E\}$.

Given a HTD E, we use

$$\sup_t E \coloneqq \sup\{t \in \mathbb{R}_{\geq 0} : \exists\ j \in \mathbb{Z}_{\geq 0},\ \text{such that } (t,j) \in E\}$$
$$\sup_j E \coloneqq \sup\{j \in \mathbb{Z}_{\geq 0} : \exists\ t \in \mathbb{R}_{\geq 0},\ \text{such that } (t,j) \in E\}.$$

Also, we let $\sup E \coloneqq (\sup_t E, \sup_j E)$, and $\text{length}(E) \coloneqq \sup_t E + \sup_j E$.

Definition 1.2. (Hybrid Arc) A hybrid arc consists of a hybrid time domain, denoted by $\text{dom}(z)$, and a function $z : \text{dom}(z) \to \mathbb{R}^n$, such that $z(\cdot, j)$ is locally absolutely continuous on $I_j \coloneqq \{t\ :\ (t,j) \in \text{dom}(z)\}$ for each $j \in \mathbb{Z}_{\geq 0}$ such that I_j has nonempty interior.

Given a hybrid system $\mathcal{H}$, its solutions are hybrid arcs z that adhere to specific conditions determined by the hybrid time domain $\text{dom}(z)$ and the data of the hybrid system (C, F, D, G).

Definition 1.3. (Solutions to Hybrid Systems) [16, Def. 2.6] A hybrid arc z is a solution to the hybrid dynamical system $\mathcal{H}$ in (1.3) if $z(0,0) \in \overline{C} \cup D$ and

1) For all $j \in \mathbb{Z}_{\geq 0}$ such that $I_j = \{t\ :\ (t,j) \in \text{dom}(z)\}$ has nonempty interior $x(t,j) \in C$ for all $t \in \text{int} I_j$, and $\dot{z}(t,j) \in F(z(t,j))$ for almost all $t \in I_j$;

2) for all $(t,j) \in \text{dom}(z)$ such that $(t, j+1) \in \text{dom}(z)$, $z(t,j) \in D$, and $z(t, j+1) \in G(x(t,j))$.

A solution to the hybrid dynamical system $\mathcal{H}$ is said to be maximal if there does not exist another solution $\tilde{z}$ to $\mathcal{H}$ such that $\text{dom}(z) \subset \text{dom}(\tilde{z})$ and $z(t,j) = \tilde{z}(t,j)$ for all $(t,j) \in \text{dom}(z)$. We use $\mathcal{S}_{\mathcal{H}}(A)$ to denote the set of all maximal solutions x to $\mathcal{H}$ with $x(0,0) \in A \subset \mathbb{R}^n$. To simplify notation, when $A = \{x_0\}$, we write $\mathcal{S}_{\mathcal{H}}(x_0)$. If not set A is mentioned, $x \in \mathcal{S}_{\mathcal{H}}$ means that x is a maximal solution to $\mathcal{H}$. A solution z to $\mathcal{H}$ is said to

be complete if $\mathrm{dom}(z)$ is unbounded.

In some cases, our models will explicitly depend on time. To study these scenarios, we can incorporate an additional auxiliary state $s \in \mathbb{R}_{\geq 0}$ to represent the time variable. This modification alters the dynamics in (1.1) as follows:

$$(x, s) \in C \times \mathbb{R}_{\geq 0}, \qquad \dot{x} \in F(x, s), \qquad \dot{s} = 1, \tag{1.2a}$$

$$(x, s) \in D \times \mathbb{R}_{\geq 0}, \qquad x^+ \in G(x), \qquad s^+ = s, \tag{1.2b}$$

In this formulation, the continuous state s represents time, which increases at a constant rate during flows and remains unchanged during jumps.

Now, to analyze the influence of external disturbances or inputs on the system dynamics, we can extend our hybrid dynamical system model. This extension leads to the following formulation:

$$(x, u) \in \tilde{C} \coloneqq C \times \mathbb{R}^m, \qquad \dot{x} \in F(x, u), \tag{1.3a}$$

$$(x, u) \in \tilde{D} \coloneqq D \times \mathbb{R}^m, \qquad x^+ \in G(x, u), \tag{1.3b}$$

where u represents the input or disturbance, which can affect both the flow and jump dynamics. The following definition is borrowed from [21].

Definition 1.4. A hybrid signal is a function defined on a HTD. A hybrid signal $u : \mathrm{dom}(u) \to \mathbb{R}^n$ is called a hybrid input if $u(\cdot, j)$ is Lebesgue measurable and locally essentially bounded for each j. A hybrid signal $x : \mathrm{dom}(x) \to \mathbb{R}^n$ is called a hybrid arc if $x(\cdot, j)$ is locally absolutely continuous for each j such that the interval $I_j \coloneqq \{t : (t, j) \in \mathrm{dom}(x)\}$ has nonempty interior. A hybrid arc $x : \mathrm{dom}(x) \to \mathbb{R}^n$ and a hybrid input $u : \mathrm{dom}(u) \to \mathbb{R}^n$ form a solution pair (x, u) to (1.3) if $\mathrm{dom}(x) = \mathrm{dom}(u)$, $(x(0, 0), u(0, 0)) \in \tilde{C} \cup \tilde{D}$, and:

1. For all $j \in \mathbb{Z}_{\geq 0}$ such that I_j has nonempty interior, and for almost all $t \in I_j$, $(z(t,j), u(t,j)) \in \tilde{C}$ and $\dot{x}(t,j) \in F(x(t,j), u(t,j))$.

2. For all $(t,j) \in \text{dom}(z)$ such that $(t, j+1) \in \text{dom}(x)$, $(x(t,j), u(t,j)) \in \tilde{D}$ and $x(t, j+1) \in G(x(t,j))$.

A hybrid solution pair (x, u) is said to be maximal if it cannot be further extended. A hybrid solution pair (x, u) is said to be complete if length $\text{dom}(z) = \infty$. This does not necessarily imply that $\sup_t \text{dom}(z) = \infty$, or that $\sup_j \text{dom}(z) = \infty$, although at least one of these two conditions should hold when z is complete. To simplify notation, in this dissertation we use $|u|_{(t,j)} = \sup_{\substack{(0,0) \leq (\tilde{t},\tilde{j}) \leq (t,j) \\ (t,j) \in \text{dom}(z)}} \left|u(\tilde{t}, \tilde{j})\right|$, and we use $|u|_\infty$ to denote $|u|_{(t,j)}$ when $t + j \to \infty$.

Stability and Convergence Notions

To model the different convergence properties of our algorithms, we make use of class $\mathcal{KL}$ functions β, which are continuous functions that satisfy $\lim_{r \to 0^+} \beta(r, \nu) = 0$ for each fixed $\nu \in \mathbb{R}_{\geq 0}$, $\lim_{\nu \to \infty} \beta(r, \nu) = 0$ for each fixed $r \in \mathbb{R}_{\geq 0}$, and which are non-decreasing in its first argument, and non-increasing in the second argument. Class $\mathcal{KL}$ functions are standard in the feedback control; see [22, 16]. Moreover, these functions can model different types of convergence properties depending on the structure of β.

Definition 1.5. Let $\mathcal{A} \subset \mathbb{R}^n$ be a closed set, and suppose every solution of (1.1) satisfies $|x(t,j)|_\mathcal{A} \leq \beta(|x(0,0)|_\mathcal{A}, t + j), \quad \forall \ (t,j) \in \text{dom}(x)$. Then, the set $\mathcal{A}$ is said to be:

(a) Uniformly Globally Asymptotically Stable (UGAS) if β is of class $\mathcal{KL}$.

(b) Uniformly Globally Exponentially Stable (UGES) if $\beta(r, s) = c_1 r e^{-c_2 s}$, $c_1, c_2 > 0$.

(c) Uniformly Globally Finite-time Stable (UGFS) if it is UGAS and there exists a continuous function $T : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $\lim_{s \to T(r)} \beta(r, s) = 0$.

(d) Uniformly Globally Fixed-time Stable (UGFXS) if it is UGFS and there exists $T^* > 0$ such that $T(r) < T^*$ for all $r \in \mathbb{R}_{\geq 0}$.

Note that all the properties listed in Definition 1.5 are stronger than standard convergence notions used in offline optimization or estimation algorithms. In particular, UGAS implies not only convergence in the standard limiting sense, but also *uniform* global stability (in the sense of Lyapunov) *and uniform* global attractivity.

We will also consider ε-parameterized HDS of the form:

$$x \in C_\varepsilon, \quad \dot{x} = F_\varepsilon(x), \quad \text{and} \quad x \in D_\varepsilon, \; x^+ \in G_\varepsilon(x), \tag{1.4}$$

where $\varepsilon > 0$. For these systems we will study *semi-global practical stability* properties as $\varepsilon \to 0^+$.

Definition 1.6. The compact set $\mathcal{A} \subset C \cup D$ is said to be *Semi-Globally Practically Asymptotically Stable* (SGP-AS) as $\varepsilon \to 0^+$ for system (1.4) if $\exists\ \beta \in \mathcal{KL}$ such that for each pair $\delta > \nu > 0$ there exists $\varepsilon^* > 0$ such that for all $\varepsilon \in (0, \varepsilon^*)$ every solution of (1.4) with $|x(0,0)|_{\mathcal{A}} \leq \delta$ satisfies

$$|x(t,j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t + j) + \nu, \tag{1.5}$$

$\forall\ (t, j) \in \mathrm{dom}(x)$. When the function β has an exponential form, we say that $\mathcal{A}$ is *semi-globally practically exponentially stable* (SGP-ES).

Part I

Momentum-Based Methods

Chapter 2

Nesterov's ODE: Foundations and Theory

The quest for efficient optimization algorithms has been a driving force in mathematical research for decades. Among the significant advancements in this field, Nesterov's accelerated gradient method, introduced by Yurii Nesterov in 1983 [20], stands out as a landmark achievement. This method revolutionized the approach to first-order optimization by achieving an optimal convergence rate of $O(1/k^2)$ for smooth convex problems, where k is the number of iterations, significantly outperforming the $O(1/k)$ rate of standard gradient descent.

In recent years, there has been a renewed interest in understanding the fundamental principles behind Nesterov's acceleration. This research has led to the formulation of Nesterov's method as a continuous-time dynamical system, known as Nesterov's ODE (Ordinary Differential Equation) [17]. This continuous-time perspective has provided new insights into the acceleration phenomenon and has become a staple for the theoretical study of momentum-based algorithms.

Nesterov's ODE is a second-order differential equation that can be viewed as the continuous-time limit of Nesterov's accelerated gradient method:

$$\ddot{z} + \frac{3}{t}\dot{z} + c\nabla f(z) = 0 \tag{2.1}$$

where $z \in \mathbb{R}^n$ is the main state, $f : \mathbb{R}^n \to \mathbb{R}^n$ is an objective function to be minimized,

∇f is its gradient, and $c > 0$ is a tunable gain. This ODE encapsulates the key features of Nesterov's acceleration: the presence of a velocity term $\dot{z}$, which provides momentum, and a time-dependent damping coefficient $\frac{3}{t}$, which gradually reduces the influence of the momentum as time increases. The interplay between momentum and dynamic damping leads to the following theorem which characterizes accelerated rates of convergence under Nesterov's ODE towards the solution of the optimization problem characterized by f [17].

Theorem 2.1. (Convergence Rate of Nesterov's ODE) For a convex function f with L-Lipschitz continuous gradient, every solution x to Nesterov's ODE satisfies:

$$f(z(t)) - f(z^\star) \leq \mathcal{O}\left(\frac{1}{t^2}\right) \tag{2.2}$$

where $z^\star$ is a minimizer of f.

The convergence result of Theorem 2.1 establishes a powerful link between Nesterov's ODE and its discrete-time counterpart by mirroring the optimal $O(1/k^2)$ convergence rate of Nesterov's accelerated gradient method. This parallel in convergence behavior opens up new avenues for analysis and algorithm design in optimization theory. By recasting the accelerated gradient method in a continuous-time framework with matching convergence properties, researchers can leverage sophisticated tools from dynamical systems theory to gain deeper insights into the acceleration phenomenon. This theoretical understanding has catalyzed the development of improved discrete-time algorithms and more nuanced continuous-time models. For example, high-resolution ODEs proposed in [23] offer a more detailed characterization of the optimization process, leading to refined acceleration techniques.

The impact of Nesterov's ODE extends far beyond its immediate application in centralized optimization. It has found applications in various fields, serving as a theoretical foundation in machine learning [18], control theory [24, 25], and distributed optimization

[26]. In machine learning, for example, the ODE perspective has led to new insights into the behavior of neural network training algorithms.Researchers have also developed several extensions to Nesterov's ODE to address different optimization scenarios. Modifications to the damping term have led to new families of accelerated methods [27], while adaptations to stochastic settings have broadened the applicability of the ODE framework [28]. These extensions demonstrate the flexibility and generality of the continuous-time approach to acceleration.

Despite these advancements and extensions, Nesterov's ODE remains one of the fundamental models for understanding acceleration in optimization. Its elegance and simplicity make it an ideal starting point for more sophisticated techniques, and its core principles continue to guide research in accelerated optimization methods.

While Nesterov's ODE offers significant acceleration in centralized optimization, its extension to control and decision-making contexts introduces several fundamental challenges. The following section explores three key limitations that arise when implementing Nesterov's method: lack of robustness guarantees, coordination constraints with multiple timers, and instability in nonpotential settings where the driving term $\nabla f(z)$ in (2.1) is replaced by a vector field $\mathcal{G}(z)$ that cannot be expressed as the gradient of a potential function f. Each of these issues poses unique obstacles to the effective application of Nesterov-like methods in distributed optimization and game-theoretic scenarios.

2.1 Robustness Limitations

A primary concern when implementing Nesterov's method is its lack of robustness. This issue becomes particularly evident when the system is subjected to small perturbations or noise, which are inevitable in practical implementations. To illustrate this issue, we consider the following example:

Example 2.1. (Instability Under Small Disturbances) Consider the problem of minimizing a quadratic cost function f whose gradient satisfies

$$\nabla f(z) = \begin{pmatrix} 10 & -5 \\ -5 & 10 \end{pmatrix} z + \begin{pmatrix} -250 \\ -150 \end{pmatrix}. \tag{2.3}$$

The function f is strongly convex and its unique minimizer is given by

$$z^\star = \frac{1}{3} \begin{pmatrix} 130 \\ 110 \end{pmatrix}.$$

Additionally, since the gradient ∇f in (2.3) is Lipschitz, Theorem 2.1 allows us to conclude that every solution z to the ODE (2.1) converges to the minimizer $z^\star$. However, when the ODE is implemented with a perturbed gradient $\nabla f(z) + e(t)$, where $e(t)$ is an arbitrarily small bounded disturbance, highly oscillatory and unstable behavior emerges, as shown in Figure 2.1 (blue line). Indeed, as shown in [29, Thm.1], in this case there exists no $\mathcal{KL}$ function β such that the solutions of the perturbed ODE satisfy the bound:

$$|z(t) - z^\star| \leq \beta(|z(0) - z^\star|, t) + \nu, \tag{2.4}$$

for all $(t) \in \mathrm{dom}(z)$.

The absence of a $\mathcal{KL}$ bound creates significant challenges when integrating Nesterov's method into larger control systems. In real-world applications, optimization algorithms often interact with various components of a control stack, such as state estimators or adaptive controllers. These interactions can introduce minor but persistent errors into the optimization process. The sensitivity of Nesterov's ODE to such perturbations complicates the task of ensuring stable performance in interconnected systems. Moreover,

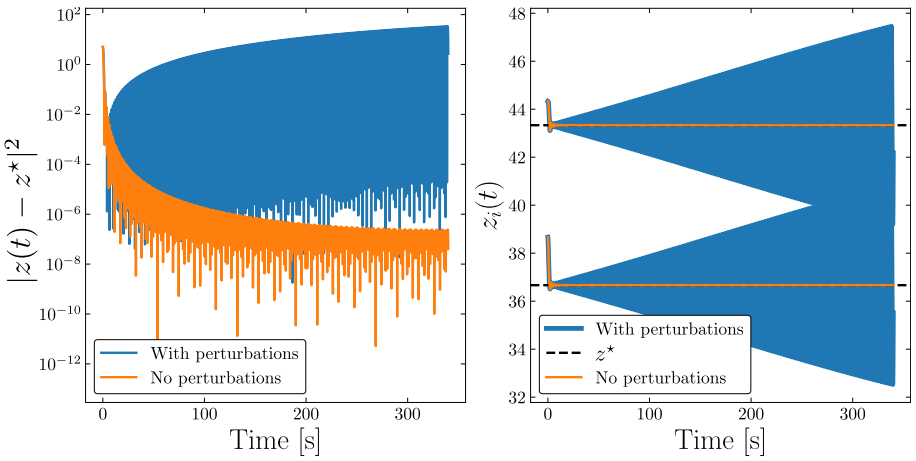


Figure 2.1. Trajectories of Nesterov's ODE with and without perturbation $\varepsilon(t)$ on the gradient $\nabla f(z)$. Instability emerges with a periodic perturbation satisfying $|\varepsilon(t)| \leq \sqrt{2} \times 10^{-1}$ for all t.

this robustness issue impedes the direct application of singular perturbation and averaging techniques, which typically require "reduced" or "average" systems with stability properties characterized by $\mathcal{KL}$ bounds [30]. These techniques play a crucial role in analyzing and designing complex control structures with multiple time scales.

To address these robustness concerns, researchers have investigated "restarting" heuristics in momentum-based optimization algorithms. Studies on restarting mechanisms have been conducted in both discrete-time [27, 31] and continuous-time settings [32, 27]. A significant advancement in this field was the development of a class of hybrid dynamical systems that incorporate discrete-time resets into continuous-time momentum methods [24]. This approach leverages the framework of hybrid dynamical systems [33] to provide a theoretical foundation for analyzing accelerated optimization algorithms with resets. These systems can induce uniform asymptotic stability properties, effectively mitigating the robustness issues inherent in the original Nesterov ODE.

A crucial element in these hybrid dynamics is the isntroduction of a timer state τ with its own dynamics, which replaces the time variable t in the dynamic damping coefficient $3/t$ that multiplies $\dot{z}$ in Nesterov's ODE. This results in the following continuous-time

system:

$$\ddot{z} + \frac{3}{\tau}\dot{z} + c\nabla f(z) = 0, \quad \dot{\tau} = 1,$$

where $\tau \in [T_0, \infty)$ is a timer state used to represent the evolution of time t in the original Nesterov's ODE (2.1), and $T_0 > 0$ is a tunable parameter. The state τ is periodically reset when it reaches an upper bound $T > 0$, following the rule:

$$\tau = T \implies \tau^+ = T_0, \tag{2.5}$$

where $T > T_0 > 0$. This mechanism increases the damping effect after the reset, aiming to prevent the accumulation of errors that could lead to instability. Figure 2.1 (black line) illustrates the effect of implementing these resetting dynamics on Nesterov's ODE in the context of Example 2.1. The figure demonstrates that incorporating resets restores the stability and convergence properties of the algorithm, even in the presence of disturbances. In Chapter 3, we will further formalize this type of hybrid dynamics and explore resetting the velocity $\dot{x}$ based on the monotonicity properties of the gradient $\nabla f(x)$.

2.2 Coordination Limitations for Distributed Implementation

While resets offer a promising solution to robustness issues of Nesterov's ODE in centralized settings, their extension to distributed environments presents new challenges. In distributed settings, a fundamental limitations is the existence of local timers τ_i for each agent $i \in \mathcal{V} \coloneqq \{1, 2, \cdots, N\}$. This localization introduces the challenge of coordinating resets for these timer variables τ_i across the network.

Agents typically have limited access to global information and restricted communication capabilities, complicating the synchronization of timer resets. As we shall see below,

without proper coordination, asynchronous resets of τ_i may disrupt the system's convergence and stability. The resulting inconsistencies in dynamic damping coefficients $3/\tau_i$ across players can lead to varying convergence rates and directions, potentially causing suboptimal solutions or instability. These coordination challenges extend beyond distributed optimization problems. They are particularly relevant in the broader context of games, where agents, or players, may have conflicting objectives and incomplete information about the global state.

To understand these coordination challenges in the context of games, in this dissertation, we consider noncooperative games with $N \in \mathbb{Z}_{\geq 2}$ players, where each player i can control its own action q_i, and has access to the actions q_j of neighboring players $j \in \mathcal{N}_i \coloneqq \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$. The connection between neighboring players in the game is characterized by a directed, connected, and time-invariant graph $\mathbb{G} = \{\mathcal{E}, \mathcal{V}\}$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of players and $\mathcal{E}$ is the set of edges between players. We use $\mathcal{L}$ to denote the in-Laplacian matrix of the graph $\mathbb{G}$. The main goal of each player i is to minimize its own cost function $\phi_i : \mathbb{R}^n \to \mathbb{R}$ by controlling its own action $q_i \in \mathbb{R}^n$. We assume that the costs ϕ_i are *twice continuously differentiable*, and we use $q = (q_1, q_2, \ldots, q_n) \in \mathbb{R}^{nN}$ to denote the overall vector of actions of the game. We also use denote as $q_{-i} \in \mathbb{R}^{n(N-1)}$ the vector of all actions with the action of player i removed. Within this setting, a *Nash equilibrium* (NE) is defined as a point $z^\star = (z_1^\star,\ z_2^\star \ldots, z_N^\star) \in \mathbb{R}^{Nn}$ that satisfies:

$$\phi_i(z_i^\star, z_{-i}^\star) = \inf_{z_i \in \mathbb{R}} \phi_i(z_i, z_{-i}^\star), \quad \forall\ i \in \mathcal{V}. \tag{2.6}$$

In other words, at a Nash equilibrium, no player can unilaterally improve their cost by changing their action. To characterize infinitesimal deviations in the actions of the players q_i and their effect in the cost functions ϕ_i, when the cost functions are differentiable, it is

useful to introduce the *pseudogradient* vector of the game $\mathcal{G} : \mathbb{R}^{Nn} \to \mathbb{R}^{Nn}$ [34, Eq. (3.9)]:

$$\mathcal{G}(z) = \left(\frac{\partial \phi_1(z)}{\partial z_1}, \frac{\partial \phi_2(z)}{\partial z_2}, \ldots, \frac{\partial \phi_n(z)}{\partial z_n} \right). \tag{2.7}$$

In the special case where there exists a potential function $\mathcal{P} : \mathbb{R}^{Nn} \to \mathbb{R}$ such that $\mathcal{G}(z) = \nabla \mathcal{P}(z)$, the game receives the name of *potential game*. In this case, finding the Nash Equilibrium of the the game essentially reduces to minimizing the cost function $\mathcal{P}(z)$ of a distributed optimization problem. By using this connection, we propose a natural extension of Nesterov's ODE (2.1) to the case of noncooperative games where we replace the gradient ∇f with the pseudogradient vector of the game $\mathcal{G}$. The extension, of course, needs to consider the distributed nature of the problem, where the computational capabilities of the players are not necessarily collocated, and where the assumption of a single timer variable τ for the overall system becomes unrealistic. Specifically, we consider the following continuous-time Nash Equilibrium seeking dynamics describing the evolution of the action of the ith player $z_i \in \mathbb{R}^n$:

$$\ddot{z}_i + \frac{3}{\tau_i} \dot{z}_i + c \cdot \frac{\partial \phi_i(z)}{\partial z_i} = 0, \quad \dot{\tau}_i = 1, \quad \forall i \in \mathcal{V}. \tag{2.8}$$

The overall dynamics, describing the evolution of the action profile $z = (z_1, z_2, \cdots, z_N) \in \mathbb{R}^{Nn}$, can then be expressed as

$$\ddot{z} + \frac{3}{2} \mathcal{T}^{-1} \dot{z} + c \cdot \mathcal{G}(z) = 0, \quad \dot{\tau} = \mathbf{1}_N. \tag{2.9}$$

where the matrix $\mathcal{T} \in \mathbb{R}^{N \times N}$ is defined as $\mathcal{T} \coloneqq \operatorname{diag}(\tau_1, \tau_2, \cdots, \tau_N) \otimes I_n$.

With the introduction of new timer variables τ_i, corresponding resetting rules of the from (2.5) might be implemented by each player, with the objective of inducing robustness guarantees into the momentum-based Nash equilibrium seeking mechanism. However, as the following example illustrates, if these resets are not coordinated, i.e., if they are not

guaranteed to occur simultaneously, prohibitively slow converges rates might emerge.

Example 2.2. (Slow Convergence with Uncoordinated Resets) Consider the momentum based dynamics in (2.8) in a potential game with $N = 30$ players, and with a strongly convex potential function f whose strong convexity constant is $\kappa = 0.01$. Each player $i \in \mathcal{V}$ implements the dynamics in (2.8) with $\phi_i = f$ for all $i \in \mathcal{V}$. Players reset $(\tau_i, \dot{x}_i)$ every 25 seconds in their local time frame, using $\tau_i^+ = 0.1$ and $(\dot{x}_i)^+ = 0$. Despite theoretical guarantees of fast convergence for this reset strategy in centralized settings [24, 25], the distributed implementation with uncoordinated resets leads to significantly slower convergence compared to standard pseudogradient flows of the form $\dot{z} = -\mathcal{G}(z)$, as shown in Figure 2.2 (blue line).

This example demonstrates how uncoordinated resets can negate the acceleration benefits of Nesterov's method, even in potential games where a global optimization perspective is applicable. The asynchronous nature of resets across different players can lead to conflicting updates, effectively slowing down the overall convergence process. The challenge lies in maintaining accelerated convergence while ensuring sufficiently coordinated resets across the network, all without resorting to centralized control. This introduces additional complexity in the analysis and design of distributed algorithms, requiring different mathematical tools to prove stability and convergence properties for these hybrid systems that combine continuous-time dynamics with discrete-time resets. We comprehensively study the use of these tools in the context of transactive control in congestion games in Section 4.4, distributed consensus-based optimization in Section 4.5, and Nash Equilibrium seeking in Section 5.1.

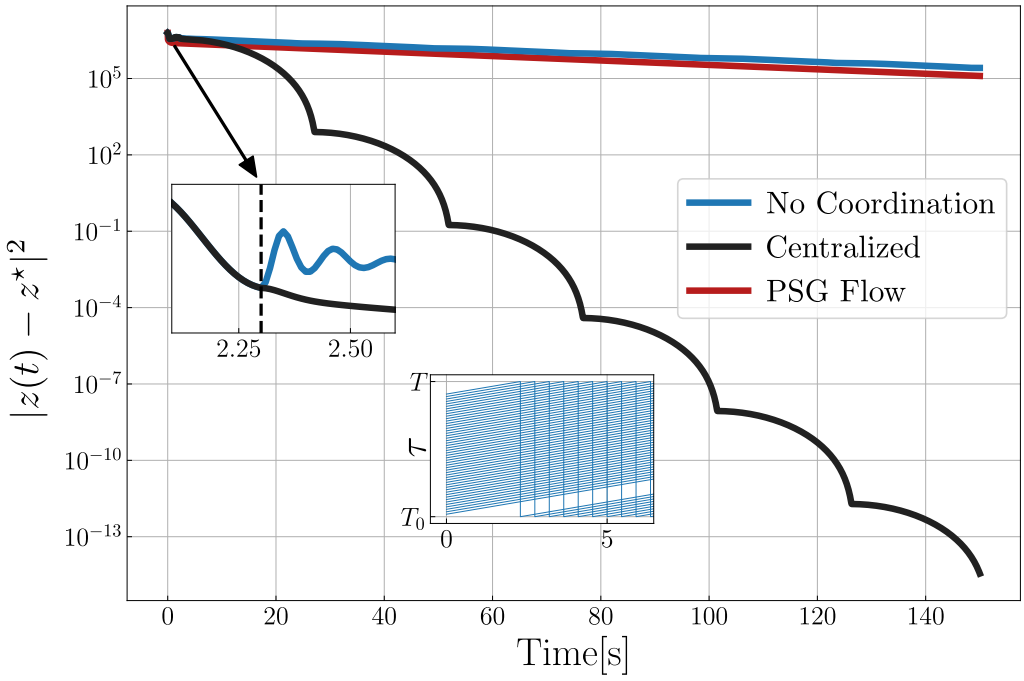


Figure 2.2. Centralized vs non-coordinated resets in a potential-game with $\kappa = 0.01, \ell = 100$ and $N = 30$. The insets show the evolution of the states τ_i with and without coordination mechanisms.

2.3 Limitations in Settings with Non-Symmetric Jacobian

While the challenges of implementing Nesterov-like methods in distributed settings are significant even when studying Nash-equilibrium seeking in potential games, the situation becomes more complex when dealing with scenarios where the vector field $\mathcal{G}$ in (2.9) does not admit a symmetric Jacobian, i.e., when there exists $x \in \mathbb{R}^{Nn}$ such that $\frac{\partial}{\partial x}\mathcal{G}(x) \neq \left(\frac{\partial}{\partial x}\mathcal{G}(x)\right)^\top$. In the context of games, this means that each player's objective function ϕ_i cannot be derived from a single global potential f, leading to what is known as a *non-potential game*. A critical limitation in non-potential games is that Nash-set seeking dynamics of the form $\dot{x} = -\mathcal{G}(x)$, or those taking the form in (2.9), can no longer be interpreted as optimization dynamics on a common landscape. Consequently, many techniques from optimization theory may not apply directly. The following example

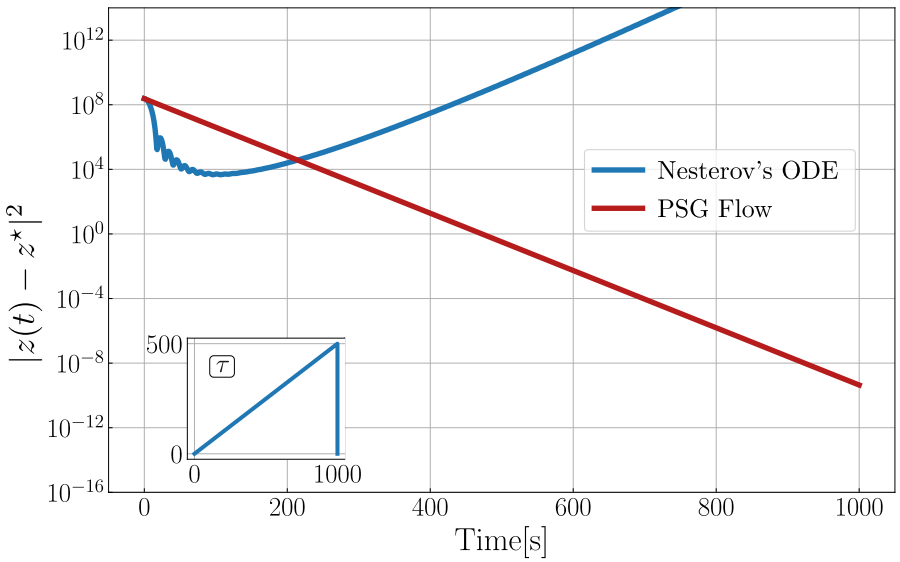


Figure 2.3. Lack of convergence of trajectories of (2.1) in a the non-potential game of Example 2.3 with $\kappa = 0.02,\ \ell = 0.0214,\ n = 30,\ T_0 = 0.1,\ T = 3.74$.

illustrates how momentum-based acceleration techniques that are effective for potential games can lead to instability in settings where $\mathcal{G}$ does not admit a symmetric Jacobian.

Example 2.3. (Lack of Convergence in Non-Potential Games) Consider a non-potential game with $N = 30$ players, characterized by the pseudogradient

$$\mathcal{G}(z) = Az + b, \tag{2.10}$$

with $z = (z_1, z_2, \ldots, z_N)$ with $z_i \in \mathbb{R}$, $A \neq A^\top$, and $\frac{1}{2}(A + A^\top) \succeq \kappa I_N$, where $\kappa = 0.02$, and $b = 0$. This game belongs to a class of games known as strongly monotone quadratic games, which have been extensively studied in the literature [34, 35]. For such games, the standard pseudogradient flow $\dot{x} = -\mathcal{G}(x)$ guarantees exponential convergence to the unique Nash equilibrium [35].

Figure 2.3 illustrates two different implementations. The blue line represents a trajectory generated by the standard pseudogradient flow using the pseudogradient in (2.10). The red line depicts a trajectory produced by system (2.9) using the same

pseudogradient $\mathcal{G}$. While the standard pseudogradient flow converges, system (2.9) generates divergent trajectories. This divergence occurs even when all timers are initially synchronized ($\tau(0) = T_0\mathbf{1}_N$ for $T_0 > 0$) and resets are implemented ($\tau_i^+ = T_0$ whenever $\tau_i = T$) every 25 seconds.

While Example 2.3 demonstrates that instability persists even with infrequent resets, Chapter 5 will reveal that implementing sufficiently frequent resets can enhance convergence and restore stability in nonpotential games.

Before presenting a solution for the cases with non-symmetric Jacobians in Chapter 5, we will first explore approaches to address the existing challenges of Nesterov's ODE in Chapters 3 and 4 in centralized settings and potential games. Our focus there will be on developing hybrid mechanisms that provide robustness, stability, and acceleration. These approaches will aim to strike a balance between harnessing the benefits of momentum and ensuring stability by employing reset policies tailored to the specific structure of the game under consideration.

Acknowledgements

Sections 2.1, 2.2, and 2.3, are a reprint of Examples 1, 2, and 3, respectively, in "Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. The dissertation author was the first author of this paper.

Chapter 3

Overcoming Robustness Limitations via Hybrid Dynamics

3.1 Hybrid Momentum-Based Mechanisms for Robust Centralized Optimization

To tackle the robustness issues of Nesterov's ODE, we begin by considering the hybrid dynamical system introduced in [24], which combines continuous-time dynamics of the form (2.1) with discrete-time resets in a *centralized setting*. To this end, we first introduce the change of coordinates

$$q = z, \quad p = \frac{t}{2}\dot{z} + z, \tag{3.1}$$

which transforms Nesterov's ODE in (2.1) into the following form:

$$\dot{x} = \begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \end{pmatrix} = F_0(x) \coloneqq \begin{pmatrix} 2\tau^{-1}(p-q) \\ -2k\tau\nabla f(q) \\ \eta \end{pmatrix}, \tag{3.2a}$$

where $x \coloneqq (q, p, \tau) \in \mathbb{R}^{2n+1}$, $k = \frac{c}{4}$, $c > 0$ is as definied in (2.1), and where we have let the rate of change of τ be a tunable parameter $\eta > 0$ to broaden the flexibility of the approach. To incorporate resets/restarts into these dynamics, the continuous-time dynamics in (3.2a)

are only allowed to flow when $z \in C_1$, with the flow set C_1 defined as follows:

$$C_1 = \{(q, p, \tau) \in \mathbb{R}^{2n+1} \ : \ \tau \in [T_0, T]\}, \tag{3.2b}$$

where $T > T_0 > 0$ are resetting parameters that condition the frequency of resets. The discrete-time dynamics modeling the reset/restarts are given by

$$x^+ = \begin{pmatrix} q^+ \\ p^+ \\ \tau^+ \end{pmatrix} = G_0(x) \coloneqq \begin{pmatrix} q \\ \alpha p + (1-\alpha) q \\ T_0 \end{pmatrix}, \tag{3.2c}$$

where $\alpha \in \{0, 1\}$ models a resetting policy whose value is chosen depending on the convexity properties of the function f. The choice $\alpha = 0$ leads to resets of the form $p^+ = q$, which corresponds to $(\dot{x})^+ = 0$ for the original Nesterov's ODE. On the other hand, $\alpha = 1$ corresponds to keeping p constant during reset events. The discrete-time dynamics (3.2c) are implemented whenever x belongs to the jump set

$$D_0 = \{(q, p, \tau) \in \mathbb{R}^{2n+1} \ : \ \tau = T\}. \tag{3.2d}$$

The data introduced in (3.2) defines the hybrid dynamical system

$$\mathcal{H}_0 = (C_0, F_0, D_0, G_0) \tag{3.3}$$

conforming to the general structure outlined in (1.1). The following theorem characterizes the stability properties of a suitable compact set $\mathcal{A}_0$ under the hybrid dynamical system $\mathcal{H}_0$, in cases where the cost function f is either convex or strongly convex [25, Thm. 1]:

Theorem 3.1. (Uniform Asymptotic Stability of the Regularized Nesterov's ODE) Consider the hybrid dynamical system $\mathcal{H}_0 = (C_0, F_0, D_0, G_0)$ defined by (3.2). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable cost/potential function and define the compact set

$$\mathcal{A}_0 \coloneqq \{(q, p, \tau) \in \mathbb{R}^{2n+1} \ : \ p = q = q^\star, \tau \in [T_0, T]\}, \tag{3.4}$$

where $q^\star$ is a minimizer of f. Then:

a) If f is convex, the set $\mathcal{A}_0$ is UGAS for $\mathcal{H}_0$. Moreover, any maximal solution $z = (q, p, \tau)$ to $\mathcal{H}_0$ satisfies the following bound:

$$f(q(t, j)) - f(q^\star) \leq \frac{c_j}{\tau(t, j)^2}, \tag{3.5}$$

for all $(t, j) \in \mathrm{dom}(z)$, where $\{c_j\}_{j=1}^{\infty} > 0$ is a sequence of monotonously decreasing positive constants satisfying $c_j \to 0$.

b) If f is κ-strongly convex, ∇f is ℓ-globally Lipschitz, and the resetting parameters (T_0, T) satisfy

$$T^2 - T_0^2 \geq \frac{1}{2k\kappa},$$

then $\mathcal{A}_0$ is UGES for $\mathcal{H}_0$.

This theorem establishes both the stability of the hybrid system and its convergence rates. For convex functions, it achieves a semi-accelerated $O(1/\tau(t, j)^2)$ convergence rate during intervals of flow, matching the performance of Nesterov's accelerated method in the initial interval. For strongly convex functions, it guarantees exponential convergence.

Furthermore, since F_0 and G_0 are continuous, and C_0 and D_0 are closed, it follows

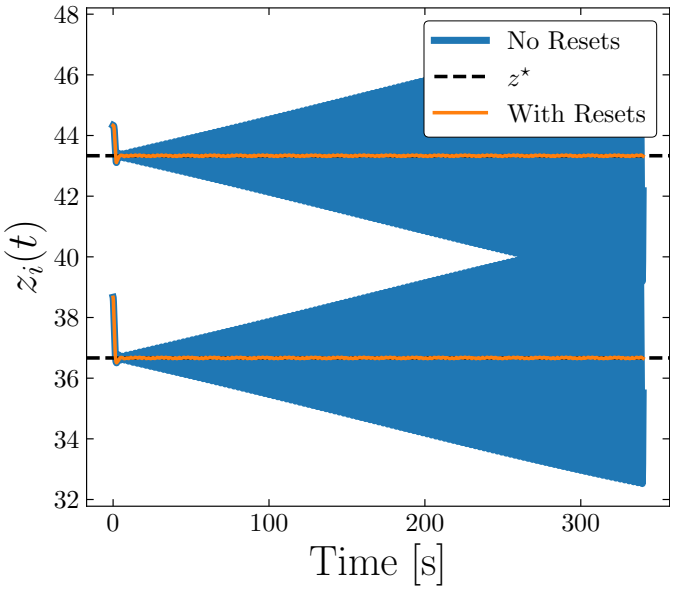


Figure 3.1. Trajectories of Nesterov's ODE, in the same scenario of Example Example 2.1, with and without resets under a periodic perturbation $\varepsilon(t)$ on the gradient $\nabla f(z)$ satisfying $|\varepsilon(t)| \leq \sqrt{2} \times 10^{-1}$ for all t.

that $\mathcal{H}_0$ is well-posed as defined in [33, Ch. 6]. By applying [33, Theorem 7.21], this well-posedness property of the HDS $\mathcal{H}_0$, combined with the uniform stability results from Theorem 3.1, leads to the following corollary. This corollary directly addresses the robustness issues of Nesterov's ODE outlined in Section 2.1.

Lemma 3.1. Let $e : \mathbb{R}_{\geq 0} \to \mathbb{R}^{2n+1}$ be a measurable function satisfying $\sup_{t \geq 0} |e(t)| \leq \bar{e}$, with $\bar{e} > 0$. Then, under conditions the conditions of Theorem 3.1-a) and Theorem 3.1-b), the dynamics

$$z + e \in C_0, \quad \dot{z} \in F_0(z + e) + e, \tag{3.6a}$$

$$z + e \in D_0, \quad z^+ \in G_0(z + e) + e, \tag{3.6b}$$

render the set $\mathcal{A}$ semi-globally practically asymptotically stable as $\bar{e} \to 0^+$.

The result of Lemma 3.1 addresses the robustness issues of Nesterov's ODE in centralized settings, as outlined in Chapter 2. To demonstrate this, we incorporate resets

into Nesterov's ODE using an auxiliary timer variable $\tilde{\tau}$. We initialize $\tilde{\tau}$ with $T_0 = 0.1$ and reset it every 4.9 time units. Simulating the same scenario presented in Example 2.1, we obtain the results shown in Figure 3.1. As observed, the incorporation of suitable resets restores the system's stability in the presence of disturbances. The next section employs the hybrid dynamics defined in (3.2) for the accelerated convergence of the weights of suitable neural networks in the context of adaptive dynamic programming.

3.2 Accelerated Continuous-Time Reinforcement Learning via Data-Assisted Hybrid Control

Recent technological advances in computation and sensing have incentivized the development and implementation of data-assisted feedback control techniques previously deemed intractable due to their computational complexity. Among these techniques, reinforcement learning (RL) has emerged as a practically viable tool with remarkable degrees of success in robotics [36], autonomous driving [13], water-distribution systems [37], among other cyber-physical applications, see [38]. These types of algorithms, are part of a large landscape of adaptive systems that aim to control a plant while simultaneously optimizing a performance index in a model-free way, with closed-loop stability guarantees.

In this section, we focus on a particular class of infinite horizon RL problems from the perspective of approximate optimal control and approximate adaptive dynamic programming (AADP). Specifically, we study the optimal control problem for nonlinear continuous-time and control-affine deterministic plants, interconnected with approximate adaptive optimal controllers [39] in an actor-critic configuration. These types of adaptive controllers aim to find, in real time, the solution to the Hamilton-Jacobi-Bellman (HJB) equation by measuring the output of the nonlinear dynamical system while making use of two approximation structures:

- a critic, used to estimate the optimal value function of the optimal control problem,

and

- an actor, used to estimate the optimal feedback controller.

Our goal is to design online adaptive dynamics for the real-time tuning of the aforementioned structures, while simultaneously achieving closed-loop stability and high transient performance. To achieve this, and motivated by the widespread usage of momentum-based gradient dynamics in practical RL settings [40], we study continuous-time actor-critic dynamics inspired by a class of ordinary differential equations (ODEs) that can be seen as continuous-time counterparts of Nesterov's accelerated optimization algorithm [19]. Such types of algorithms have gained popularity in optimization and related fields due to the fact that they can minimize smooth convex functions at a rate of order $\mathcal{O}(1/t^2)$ [18]. The main source for the acceleration property in these ODEs comes from the addition of momentum to gradient-based dynamics, in conjunction with a vanishing dynamic damping coefficient. However, as recently shown in [25] and [29], the non-uniform convergence properties that emerge in these types of dynamics complicates their use in feedback systems with plant dynamics in the loop. in this chapter, we overcome these challenges by incorporating resets into the proposed momentum-based algorithms, similar to restarting heuristics studied in the machine learning literature, see [17] and [31]. Our resulting actor-critic controller is naturally modeled by a hybrid dynamical system that incorporates continuous-time and discrete-time dynamics, which we analyze using tools from [33].

A traditional assumption in the literature of continuous-time actor-critic RL is that the regressors used in the parameterizations satisfy a persistence of excitation condition along the trajectories of the plant. However, in practice, this condition can be difficult to verify a priori. To circumvent this issue, in this chapter we consider a data-assisted approach, where a finite amount of past "sufficiently rich" recorded data is used to guarantee asymptotic learning in the closed-loop system. As a consequence, the resulting data-assisted hybrid control algorithm concurrently uses real-time and recorded data,

similar in spirit to concurrent-learning (CL) techniques [41]. By using Lyapunov-based tools for hybrid dynamical systems, we analyze the interconnection of an actor-critic neural-network (NN) controller and the nonlinear plant, establishing that the trajectories of the closed-loop system remain ultimately bounded around the origin of the plant and the optimal actor and critic NN parameters. Since the resulting closed-loop system has suitable regularity properties in terms of continuity of the dynamics, our stability results are in fact robust with respect to arbitrarily small additive disturbances that can be adversarial in nature, or that can arise due to numerical implementations. To the best knowledge of the authors, these are the first theoretical stability guarantees of continuous-time accelerated actor-critic algorithms for neural network-based adaptive dynamic programming controllers in nonlinear deterministic settings.

The Time-Invariant Hamilton-Jacobi-Bellman Equation

Consider a control-affine nonlinear dynamical plant

$$\dot{x} = f(x) + g(x)u, \tag{3.7}$$

where $x \in \mathbb{R}^n$ is the state of the system, $u \in U \subset \mathbb{R}^m$ is the input, and $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are locally Lipschitz functions. Our goal is to design a stable algorithm able to find –in real time– a control law u^* that minimizes the cost functional $V : \mathbb{R}^n \times \mathcal{U}_V \to \mathbb{R}$ given by:

$$V(x_0, u) \coloneqq \int_0^\infty r\Big(x\big(\tau\big), u\left(x(\tau)\right)\Big)d\tau, \tag{3.8}$$

where $x\big(t\big)$ represents a solution to (3.7) from the initial condition $x(0) = x_0$, that results from implementing a feedback law u, belonging to a class of admissible control laws $\mathcal{U}_V$ characterized as follows:

Definition 3.1. [42, Definition 1] Given the dynamical system in (3.7), a feedback control $u : \mathbb{R}^n \to \mathbb{R}^m$ is *admissible* with respect to the cost functional V in (3.8) if

- u is continuous,
- u renders system (3.7) UAS,
- $V(x_0, u) < \infty$ for all $x_0 \in \mathbb{R}^n$. □

We denote the set of *admissible feedback laws* as $\mathcal{U}_V$.

In (3.8), we consider cost functions $r : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ of the form $r(x, u) \coloneqq Q(x) + R(u)$, where the state-cost is given by $Q(x) \coloneqq x^\top \Pi_x x$ with $\Pi_x \succ 0$, and the control-cost is given by $R(u) \coloneqq u^\top \Pi_u u$ with $\Pi_u \succ 0$. To find the optimal control law that minimizes (3.8), we study the *Hamiltonian function* $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ related to (3.7) and (3.8), given by

$$H(x, u, \nabla V) \coloneqq \nabla V^\top (f(x) + g(x)u) + Q(x) + R(u). \tag{3.9}$$

Using (3.9), a necessary optimality condition for u^* is given by Pontryagin's maximum principle [43]:

$$u^*(x) = \arg\min_{u \in \mathcal{U}_V} H(x, u, \nabla V^*) \implies u^*(x) = -\frac{1}{2}\Pi_u^{-1} g(x)^\top \nabla V^*(x), \tag{3.10}$$

where V^* represents the *optimal value function*:

$$V^*(x) \coloneqq \inf_{u \in \mathcal{U}_V} V(x, u(\cdot))$$

On the other hand, under the assumption that V^* is continuously differentiable, the optimal value function can be shown to satisfy the Hamilton-Jacobi-Bellman equation [39,

Ch. 1.4]:

$$\frac{\partial V^*}{\partial t} = -H(x, u^*, \nabla V^*) \quad \forall x \in \mathbb{R}^n.$$

Since the functional in (3.8) does not have an explicit dependence on t, it follows that $\frac{\partial V^*}{\partial t} = 0$, and hence $H(x, u^*, \nabla V^*) = 0$, meaning that for all $x \in \mathbb{R}^n$, the following holds:

$$\nabla {V^*}^\top \big(f(x) + g(x)u^*(x)\big) + Q(x) + R\Big(u^*(x)\Big) = 0. \tag{3.11}$$

The time-invariant Hamilton-Jacobi-Bellman equation in (3.11), enables a state-dependent characterization of optimality. Therefore, by using the optimal control law in (3.10), and assuming that the system dynamics (3.7) are known, the form (3.11) could be leveraged to find V^*. Unfortunately, finding an explicit closed-form expression for V^*, and thus for the optimal control law, is, in general, an intractable problem. However, the utility of (3.11) is not completely lost. As we shall show in the following sections, online and historical "measurements" of (3.11) can be leveraged in real time to estimate the optimal control law u^* while concurrently rendering a neighborhood of the origin of system (3.7) asymptotically stable.

Data-Assisted Critic Dynamics

To leverage the form of (3.11), we consider the following parameterization of the optimal value function $V^*(x)$:

$$V^*(x) = {\theta_c^*}^\top \phi_c(x) + \epsilon_c(x) \quad \forall x \in K, \tag{3.12}$$

where $K \subset \mathbb{R}^n$ is a compact set, $\theta_c^* \in \mathbb{R}^{l_c}$, $\phi_c : \mathbb{R}^n \to \mathbb{R}^{l_c}$ is a vector of continuously differentiable basis functions, and $\epsilon_c : \mathbb{R}^n \to \mathbb{R}$ is the approximation error. The parameterization (3.12) is always possible on compact sets due to the continuity properties of V and the universal approximation theorem [44]. This parametrization results in an optimal

Hamiltonian of the form $H_p^* \coloneqq H(x, u^*, \frac{\partial \phi_c}{\partial x}^\top \theta_c^* + \nabla \epsilon_c)$ given by:

$$H_p^*(x) = \theta_c^{*\top} \psi(x, u^*(x)) + Q(x) + R\left(u^*(x)\right) + \nabla \epsilon_c(x)^\top \left(f(x) + g(x)u^*(x)\right), \tag{3.13}$$

where we defined $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{l_c}$ as:

$$\psi(x, u) \coloneqq \frac{\partial \phi_c(x)}{\partial x} \left(f(x) + g(x)u\right). \tag{3.14}$$

We note that the explicit dependence of $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{l_c}$ on the control action u, defined in (3.14), is a fundamental departure from the previous approaches studied in the context of concurrent learning (CL) NN actor-critic controllers, such as those considered in [45] and [46]. In particular, we note that in the context of CL the data used to estimate the optimal value function V^* is generated from measurements of the optimal Hamiltonian which, by definition, incorporates the optimal control law u^*. Hence, the need to include u as part of the regressor vectors ψ becomes crucial; this dependence characterizes how far our recorded measurements of a Hamiltonian are from the optimal Hamiltonian H_p^*. Indeed, this distance will explicitly emerge in our convergence and stability analysis. Naturally, the dependence of (3.14) on u will impose stronger conditions on the recorded data needed to estimate V^*.

Assuming we have access to ϕ_c, we can define a *critic* neural network as:

$$\hat{V}(x) \coloneqq \theta_c^\top \phi_c(x), \quad \forall x \in K, \tag{3.15}$$

which will serve as an approximation of the optimal value function V^* in (3.12). This critic NN results in an estimated Hamiltonian:

$$H\left(x, u, \nabla \hat{V}\right) \coloneqq \theta_c^\top \psi\left(x, u\right) + Q(x) + R(u), \tag{3.16}$$

which we will use to design the update dynamics of the critic parameters θ_c. In particular, our goal is to use previously recorded data from trajectories of the plant to ensure asymptotic stability of the set of optimal critic parameters $\{\theta_c^*\}$, while simultaneously enabling the incorporation of instantaneous measurements from the plant. To achieve this goal, we will assume enough "richness" properties in the recorded data, a notion that is captured by a relaxed (and finite-time) version of *persistence of excitation* (PE); see [41] and [47].

Assumption 3.1. Let $\{\psi(x_k, u^*(x_k))\}_{k=1}^N$ be a sequence of recorded data, and define:

$$\Lambda \coloneqq \sum_{k=1}^{N} \Psi(x_k, u^*(x_k))\Psi(x_k, u^*(x_k))^\top, \quad \Psi(x,u) \coloneqq \frac{\psi(x,u)}{1+\psi(x,u)^\top\psi(x,u)}. \tag{3.17}$$

There exists $\underline{\lambda} \in \mathbb{R}_{>0}$ such that $\Lambda \succeq \underline{\lambda} I_n$, i.e., the data is $\underline{\lambda}$-sufficiently-rich ($\underline{\lambda}$-SR). □

Remark 3.1. In this section, we study reinforcement learning dynamics that do not make explicit usage of exploration signals with standard PE properties, which can be difficult to guarantee in practice. Instead, we assume access to samples obtained by observing the action of optimal values $u^*(x_k)$ acting on the plant. Note however that this does not imply knowledge of the optimal control policy as a whole, but only of a finite number of *demonstrations* from an "expert" policy. Similar requirements commonly arise in the literature of imitation learning, or inverse reinforcement learning, and have been recently shown in practice to reduce the exploratory requirements of online reinforcement learning algorithms, with mild assumptions in the sampling of the demonstrations. For recent discussions on these topics in the discrete-time stochastic reinforcement learning setting we refer the reader to [48] and [49].

Now, we consider the instantaneous and data-dependent errors of the estimated Hamiltonian with respect to the optimal one:

$$\begin{aligned} e^i\left(\theta_c, x, u\right) &\coloneqq H\left(x, u, \nabla \hat{V}\right) - H\left(x, u^*(x), \nabla V^*\right) \\ &= \theta_c^\top \psi\left(x, u\right) + Q(x) + R\left(u\right), \\ e_k^d(\theta_c) &\coloneqq H\left(x_k, u^*(x_k), \nabla \hat{V}\right) - H\left(x_k, u^*(x_k), \nabla V^*\right) \\ &= \theta_c^\top \psi\left(x_k, u^*(x_k)\right) + Q(x_k) + R\left(u^*(x_k)\right), \end{aligned}$$

where we used the fact that $H\left(x, u^*(x), \nabla V^*\right) = 0$. Moreover, we define the *joint instantaneous and data-dependent* error as:

$$e\left(\theta_c, x, u\right) \coloneqq \frac{1}{2}\left(\rho_i \frac{e^i\left(x, \theta_c, u\right)^2}{\left(1 + |\psi(x,u)|^2\right)^2} + \rho_d \sum_{k=1}^{N} \frac{e_k^d(\theta_c)^2}{\left(1 + |\psi\left(x_k, u^*(x_k)\right)|^2\right)^2}\right), \tag{3.18}$$

where $\rho_i \in \mathbb{R}_{\geq 0}$ and $\rho_d \in \mathbb{R}_{>0}$ are tunable gains. Since we are interested in designing real-time training dynamics for the estimation of the optimal parameters θ_c^*, we compute the the gradient of (3.18) with respect to θ_c as follows:

$$\begin{aligned} \nabla_{\theta_c} e(\theta_c, x, u) = \rho_i &\left(\Psi(x,u)\Psi(x,u)^\top \theta_c + \frac{\psi(x,u)\left[Q(x) + R(u)\right]}{\left(1 + \psi(x,u)^\top \psi(x,u)\right)^2} \right) \\ &+ \rho_d \left(\Lambda \theta_c + \sum_{k=1}^{N} \frac{\psi(x_k, u^*(x_k))\left[Q(x_k) + R\left(u^*(x_k)\right)\right]}{\left(1 + \psi\left(x_k, u^*(x_k)\right)^\top \psi\left(x_k, u^*(x_k)\right)\right)^2} \right), \end{aligned} \tag{3.19}$$

where Λ and Ψ are defined in Assumption 3.1.

The "propagated" error to the HJB equation that results from the approximate parametrization of V^* in (3.12), is given by:

$$\epsilon_{\text{HJB}}(x) \coloneqq H(x, u^*(x), \nabla V^*) - H\left(x, u^*, \frac{\partial \phi_c(x)}{\partial x}^\top \theta_c^*\right)$$

$$= -\nabla \epsilon_c^\top(x)\Big(f(x) + g(x)u^*(x)\Big). \tag{3.20}$$

The following assumption is standard, and it is satisfied when the involved functions are continuous and K is compact.

Assumption 3.2. There exist $\overline{\phi_c},\ \overline{d\phi_c},\ \overline{\epsilon_c},\ \overline{d\epsilon_c},\ \overline{\epsilon_{\text{HJB}}}, \overline{g} \in \mathbb{R}_{>0}$ such that

$$|\phi_c(x)| \leq \overline{\phi_c}, \quad \left|\frac{\partial \phi_c(x)}{\partial x}\right| \leq \overline{d\phi_c}, \quad |\epsilon_c(x)| \leq \overline{\epsilon_c},$$
$$|\nabla \epsilon_c(x)| \leq \overline{d\epsilon_c},\ |\epsilon_{HJB}(x)| \leq \overline{\epsilon_{HJB}},\ |g(x)| \leq \overline{g} \quad \forall x \in K,$$

where K is the same set considered in (3.12). □

Critic Dynamics via Data-Driven Hybrid Momentum-Based Control

To design fast asymptotically stable dynamics for the estimate θ_c, we propose a new class of momentum-based critic dynamics inspired by accelerated gradient flows with restarting mechanisms, such as those studied in [19] and [31]. Specifically, we consider the following hybrid dynamics with a similar structure to the HDS (3.2), now with state $y \coloneqq (\theta_c, p, \tau)$ and data defined by:

$$C_0^c \coloneqq \left\{y \in \mathbb{R}^{2l_c+1} \ : \ \tau \in [T_0, T]\right\}, \qquad F_0^c(y, x, u) \coloneqq \begin{pmatrix} \frac{2}{\tau}(p - \theta_c) \\ -2k_c \nabla_{\theta_c} e(\theta_c, x, u) \\ \frac{1}{2} \end{pmatrix}, \tag{3.21a}$$

$$D_0^c \coloneqq \left\{y \in \mathbb{R}^{2l_c+1} \ : \ \tau = T\right\}, \qquad G_0^c(y) \coloneqq \begin{pmatrix} \theta_c \\ \theta_c \\ T_0 \end{pmatrix}, \tag{3.21b}$$

where $k_c \in \mathbb{R}_{>0}$ is a tunable gain, and (p, τ) are auxiliary states that are periodically reset every time $\tau = T$ via the jump map (3.21b), with $\infty > T > T_0 > 0$. The dynamical

system in (3.21) flows in continuous time according to (3.21a) whenever the timer variable τ is in $[T_0, T]$. As soon as τ hits T, the algorithm (3.21) resets the timer variable to T_0, as well as the momentum variable p to θ_c, while leaving θ_c unaffected. Accordingly, after the first reset, the system exhibits periodic resets every $\Delta T = 2(T - T_0)$ intervals of time. The following assumption provides data-dependent tuning guidelines for the resetting frequency of the timer variable τ, which will be leveraged in our stability results.

Assumption 3.3. The tunable parameters $(T_0, T, k_c, \rho_i, \rho_d)$ satisfy $2\rho_d\underline{\lambda} > \rho_i$ and

$$T_0^2 + \frac{1}{2k_c\underline{\lambda}\rho_d} < T^2 < \frac{8\rho_d\underline{\lambda}}{k_c\rho_i^2}, \tag{3.22}$$

where $\underline{\lambda}$ is the level of richness of the recorderd data defined in Assumption 3.1. □

For system (3.21), we study stability properties with respect to the compact set:

$$\mathcal{A}_c \coloneqq \mathcal{A}_{\theta_c,p} \times [T_0, T], \tag{3.23a}$$

$$\mathcal{A}_{\theta_c,p} \coloneqq \left\{(\theta_c, p) \in \mathbb{R}^{2l_c} : p_c = \theta_c,\ \theta_c = \theta_c^*\right\}. \tag{3.23b}$$

The following theorem is the first main result of this paper. All the proofs are presented in Appendix A.1.

Theorem 3.2. Given a number l_c of basis functions ϕ_c parametrizing the critic NN, and a compact set $K \subset \mathbb{R}^n$, suppose that Assumptions 3.1, 3.2 and 3.3 are satisfied. Then, there exists $(\kappa, c) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ and class-$\mathcal{K}_\infty$ functions γ_1 and γ_2, such that for every solution $y = (\theta_c, p, \tau)$ to (3.21) with initial condition $y(0,0) = (\theta_c(0,0), p(0,0), \tau(0,0))$, and using the control policy $u(\cdot) \in \mathcal{U}_V$ on the plant, the critic parameters θ_c satisfy

$$|\theta_c(t,j) - \theta_c^*| \leq \kappa e^{-c(t+j)}|y(0,0)|_{\mathcal{A}_c} + \gamma_2\left(|\tilde{u}(x(t,j))|\right) + \gamma_1(\overline{\epsilon_{\text{HJB}}}), \tag{3.24}$$

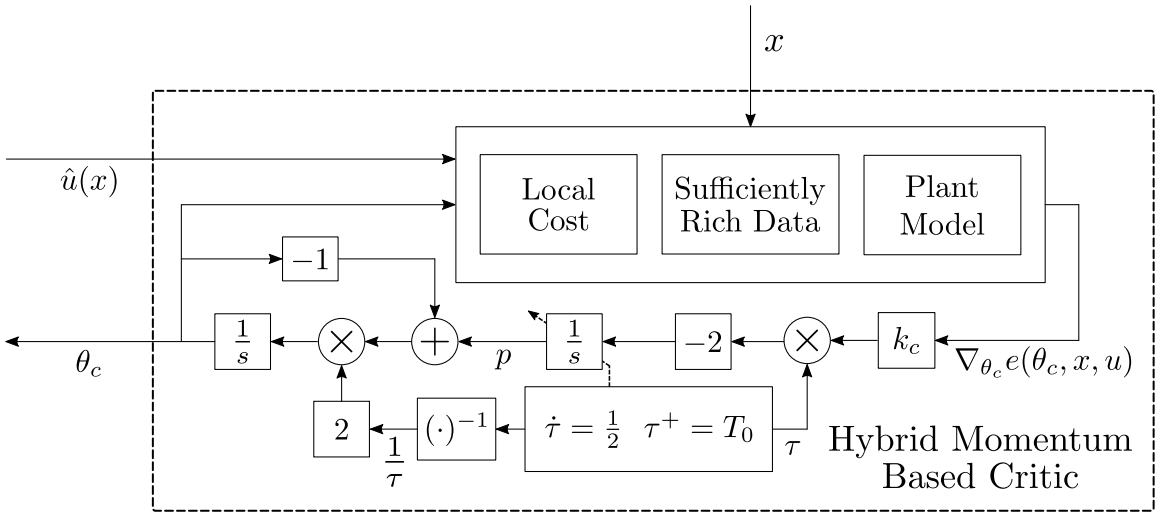


Figure 3.2. Proposed Hybrid Momentum Based Dynamics for the training of the Critic subsystem

where $\tilde{u}(x(t,j)) := u(x(t,j)) - u^*(x(t,j))$, for all $(t,j) \in \mathrm{dom}\,(y)$ □

The presence of a residual optimal-control mismatch term in (3.24) of the form $\gamma_2(|u(x) - u^*(x)|)$, represents a crucial difference with respect to previous CL adaptive dynamic approaches, such as those studied in [45] and [39, Ch. 4]. This term is a direct byproduct of our definition of ψ in (3.14), its dependence on the control action u, and its appearance in the error gradient (3.19). In principle, the emergence of this term in Theorem 1 is agnostic to the particular gradient-based update dynamics for the critic NN, regardless of the inclusion or not of momentum. Since $\gamma_2 \in \mathcal{K}$, the larger the difference between the nominal input u and the optimal feedback law u^*, the greater the residual error in the convergence of θ_c. In particular, the bound (3.24) describes a semi-global practical input-to-state stability property that, to the best knowledge of the authors, is novel in the context of CL-based RL. In the next section we will show that the residual error $\gamma_2(|\tilde{u}|)$ can be removed by incorporating an additional actor NN in the system.

Remark 3.2. In contrast to standard data-driven gradient-descent dynamics for the estimation of the optimal value function V^*, which can achieve exponential rates of convergence proportional to $\underline{\lambda}$ (cf. [46, 41]), under the assumptions of Theorem 3.2 the critic update dynamics (3.21) can achieve exponential convergence with rates proportional to $\sqrt{\underline{\lambda}}$. As shown in [25], momentum-based dynamics of this form can achieve these rates using the restarting parameter

$$T = T^* := e\sqrt{\frac{1}{2k_c\rho_d\underline{\lambda}} + T_0^2}. \tag{3.25}$$

This property is particularly useful in settings where the level of richness of the data-set is limited, i.e., when $\underline{\lambda} \ll 1$, which is common in practical applications.

Theorem 3.2 guarantees exponential convergence to a neighborhood of the optimal parameters $\{\theta_c^*\}$ that define the optimal value function V^*. Consequently, by continuity, and on compact sets, $\hat{V}$ would converge to an ϵ-approximation of V^*, which can be leveraged by the control law (3.10) to stabilize system (3.7). However, as noted in [50], implementing only critic structures for the control of nonlinear dynamical systems of the form (3.7) can lead to poor closed-loop transient performance. To tackle this issue, we consider an auxiliary dynamical system, called the *actor*, which will serve as an estimator of the optimal controller that acts on the plant.

Actor Dynamics

Using the optimal value parametrization described in Section 3.2 the optimal control law can written as:

$$u^*(x) = -\frac{1}{2}\Pi_u^{-1} g(x)^\top \left[\frac{\partial \phi_c(x)}{\partial x}^\top \theta_c^* + \nabla \epsilon_c(x)\right], \quad \forall x \in K. \tag{3.26}$$

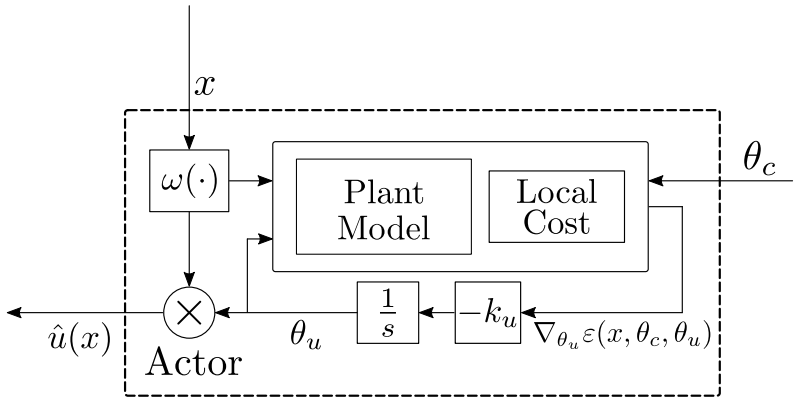


Figure 3.3. Actor Subsystem

Therefore, using $\frac{\partial \phi_c(x)}{\partial x}$ and $g(x)$ we can implement an actor neural-network given by:

$$\hat{u}(x) = \omega(x)^\top \theta_u, \tag{3.27}$$

where $\omega : \mathbb{R}^n \to \mathbb{R}^{l_c \times m}$ is defined as:

$$\omega(x) \coloneqq -\frac{1}{2}\frac{\partial \phi_c(x)}{\partial x} g(x) \Pi_u^{-1}. \tag{3.28}$$

To guarantee convergence of $\hat{u}$ to u^*, we design update dynamics for $\theta_u \in \mathbb{R}^{l_c}$ based on the minimization of the error:

$$\begin{aligned}
\varepsilon(x, \theta_c, \theta_u) &\coloneqq \frac{1}{2}\left[\alpha_1 \frac{\varepsilon_a(x,\theta_c,\theta_u)^\top \varepsilon_a(x,\theta_c,\theta_u)}{1 + \mathrm{Tr}\{\omega(x)^\top \omega(x)\}} + \alpha_2 \varepsilon_b(\theta_c,\theta_u)^\top \varepsilon_b(\theta_c,\theta_u)\right], \\
\varepsilon_a(x, \theta_c, \theta_u) &\coloneqq \hat{u}(x) - \omega(x)^\top \theta_c = \omega(x)^\top (\theta_u - \theta_c), \\
\varepsilon_b(\theta_c, \theta_u) &\coloneqq \theta_u - \theta_c,
\end{aligned} \tag{3.29}$$

which satisfies:

$$\nabla_{\theta_u} \varepsilon(x, \theta_c, \theta_u) = \Omega(x)(\theta_u - \theta_c),$$

where

$$\Omega(x) \coloneqq \alpha_1 \frac{\omega(x)\omega(x)^\top}{1 + \mathrm{Tr}\{\omega(x)^\top \omega(x)\}} + \alpha_2 I \in \mathbb{R}^{l_c \times l_c} \quad \forall x \in \mathbb{R}^n. \tag{3.30}$$

Based on these definitions, we consider the following gradient-descent dynamics for the actor neural-network:

$$\dot{\theta}_u = F_u(\theta_u, x, \theta_c) \coloneqq -k_u \nabla_{\theta_u} \varepsilon(x, \theta_c, \theta_u), \tag{3.31}$$

where $k_u \in \mathbb{R}_{>0}$ is a tunable gain. A scheme representing these update dynamics is shown in Figure 3.3.

Momentum-Based Actor-Critic Feedback System

Consider the closed-loop, shown in Figure 3.4, resulting from the interconnection between the plant (3.7), the critic update dynamics (3.21), the actor update dynamics (3.31) and the feedback law in (3.27), and given by:

$$\dot{x} = f(x) + g(x)\hat{u}(x), \qquad x^+ = x, \tag{3.32a}$$

$$\dot{y} = F_0^c(y, x, \hat{u}(x)), \qquad y^+ = G_0^c(y), \tag{3.32b}$$

$$\dot{\theta}_u = F_u(\theta_u, x, \theta_c), \qquad \theta_u^+ = \theta_u, \tag{3.32c}$$

and with flow set and jump set given by $C_0^{adp} = \mathbb{R}^n \times C_0^c \times \mathbb{R}^{l_c}$ and $D_0^{adp} = \mathbb{R}^n \times D_0^c \times \mathbb{R}^{l_c}$ respectively, where C_0^c and D_0^c are as defined in (3.21). Let $z \coloneqq (x, y, \theta_u)$ be the overall state of the closed-loop system, and define:

$$\mathcal{A} \coloneqq \{0\} \times \mathcal{A}_c \times \{\theta_c^*\}.$$

The following is the main result of this paper.

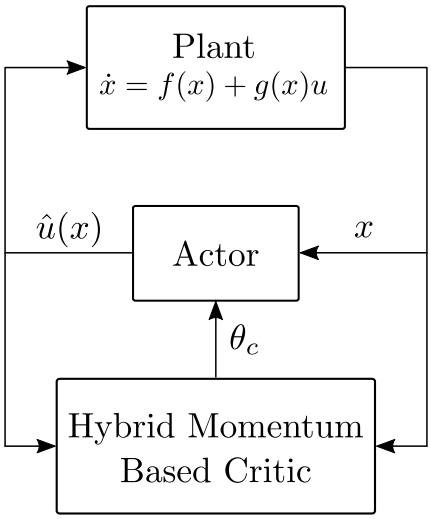


Figure 3.4. Diagram of the closed-loop system

Theorem 3.3. Given the vector of basis functions $\phi_c : \mathbb{R}^n \to \mathbb{R}^{l_c}$ parametrizing the critic NN and a compact set $K_z \coloneqq K \times K_y \times K_\theta \subset \mathbb{R}^n \times \mathbb{R}^{2l_c+1} \times \mathbb{R}^{l_c}$, where K is given as in (3.12), suppose that Assumption 3.1-3.3 are satisfied. Then, there exists $\beta \in \mathcal{KL}$, $\gamma \in \mathcal{K}$ and tunable parameters $(\rho_i, \rho_d, k_c, k_u, \alpha_1, \alpha_2)$, such that for every solution $z = (x, y, \theta_u)$ to the closed-loop system (3.32), with initial condition $z(0,0) = (x(0,0), y(0,0), \theta_u(0,0)) \in K_z$, there exists $\tilde{T} > 0$ such that for all $(t, j) \in \mathrm{dom}(z)$:

$$|z(t,j)|_{\mathcal{A}} \leq \beta(|z(0,0)|_{\mathcal{A}}, t+j) + \gamma(\left|\left(\overline{\epsilon_{\mathrm{HJB}}}, \overline{d\epsilon_c}\right)\right|) + \nu,$$

for all $0 \leq t + j \leq \tilde{T}$, and

$$|z(t,j)|_{\mathcal{A}} \leq \gamma(\left|\left(\overline{\epsilon_{\mathrm{HJB}}}, \overline{d\epsilon_c}\right)\right|) + \nu, \quad \forall\ \tilde{T} \leq t + j,$$

for some $\nu > 0$ constant. □

Theorem 3.3 establishes asymptotic convergence to a neighborhood of the compact set $\mathcal{A}$ as $\left(\overline{\epsilon_{\mathrm{HJB}}}, \overline{d\epsilon_c}\right) \to 0$ from any compact set K_z modulo some error ν, under a suitable

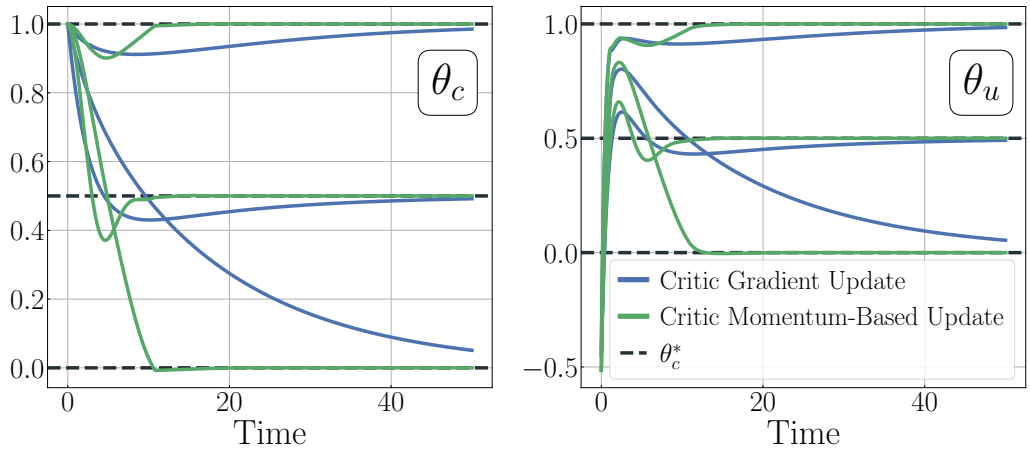


Figure 3.5. Convergence of the weights of the critic's (left) and actor's (right) networks to the optimal values.

choice of tunable parameters. To the best knowledge of the authors this is the first result providing stability certificates for continuous-time actor-critic reinforcement learning using recorded data and accelerated value-function estimation dynamics with momentum. In addition, since the resulting closed-loop system in (3.32) is given by a well-posed hybrid system, the stability results are robust with respect to arbitrarily small additive disturbances on the states and dynamics [33, Ch. 7]; see Lemma 3.1.

Numerical Example: To illustrate our theoretical results we study the following nonlinear control-affine plant:

$$\dot{x} = f(x) + g(x)u, \tag{3.33a}$$

$$f(x) = \begin{bmatrix} -x_1 + x_2 \\ -\frac{1}{2}\Big(x_1 - x_2\big(1 - \cos(2x_1 + 2)^2\big)\Big) \end{bmatrix}, \tag{3.33b}$$

$$g(x) \coloneqq \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}, \tag{3.33c}$$

with local state and control costs given by $Q(x) = x^\top x$ and $R(u) = u^2$ [46]. The optimal value function for this setting is given by $V^*(x) = \frac{1}{2}x_1^2 + x_2^2$ with optimal control law given

by $u^*(x) = -(\cos(2x_1)+2)x_2$. Using this information, we choose $\phi_c(x) = (x_1^2, x_1x_2, x_2^2)$, and we implement the prescribed hybrid momentum-based dynamics in (3.21) for the update of the critic neural network, and the update dynamics for the actor described in (3.31). We obtain the results shown in Figure 3.5 with $x(0,0) = (-10, 10)$, $\theta_c(0,0) = (1,1,1)$ and $\theta_u \in [0,1]^3$. We compare the results with the case in which the critic neural-network is updated with the gradient-descent dynamics of [45], and where the sufficiently rich data is a set of 16 data points obtained by sampling the dynamics (3.33) in a grid around the origin of size 4×4. In our simulations we use $T_0 = 0.1, T = 5.5$ for the momentum-based dynamics in (3.21). These particular values are obtained by using the level of richness $\underline{\lambda}$ of the data-set, and the inequalities in (3.22) in order to ensure compliance with Assumption 3.3. For both reinforcement learning dynamics we use $k_c = 1, k_u = 1, \rho_d = 1$ and $\rho_i = 1$. As shown in the figure both update dynamics are able to converge to $\{\theta_c^*\}$, with $\theta_c^* = (1/2, 0, 1)$ describing the optimal value function V^*. However, the hybrid-based dynamics are able to significantly improve the transient performance of the learning mechanism.[1]

We conclude this section by noting that, while the results presented in this chapter address the robustness issues of Nesterov's ODE in centralized settings via the incorporation of resets on the timer state τ, extending them to distributed settings introduces new challenges in coordinating those resets across agents.

Acknowledgements

Section 3.2, in full, is a reprint of the material as it appears in "Accelerated Continuous-Time Reinforcement Learning via Data-Assisted Hybrid Control" [2]. The dissertation author was the first author of this paper.

[1]The code used to implement this simulation can be found in the following repository: https://github.com/deot95/Accelerated-Continuous-Time-Approximate-Dynamic-Programming-through-Data-Assisted-Hybrid-Control

Chapter 4

Distributed Momentum-Based Dynamics in Potential Games

The limitations of Nesterov's ODE discussed in the previous chapter necessitate an alternative set of tools for applying momentum-based dynamics of the form (2.1) in control and distributed decision-making contexts. This chapter introduces a class of hybrid dynamical systems incorporating distributed and coordinated restarting/resetting mechanisms. These systems are specifically designed to address the challenges of robustness and coordination in potential games and distributed optimization problems discussed in Chapter 2, Sections 2.1 and 2.2. The discussion and proposed solutions for problems arising in nonpotential settings are deferred to Chapter 5.

4.1 A Robust Coordination Mechanism for Distributed Resets

To overcome the limitations of momentum-based dynamics of the form (2.8) in distributed decision-making, we first employ the change of coordinates introduced in (3.1) and incorporate resets to the individual dynamics of players presented in the distributed noncooperative game formulation of Section 2.2. This transformation yields the following

continuous-time dynamics for player $i \in \mathcal{V} = \{1, 2, \cdots, N\}$:

$$\tau_i \in [T_0, T) \Longrightarrow \begin{pmatrix} \dot{q}_i \\ \dot{p}_i \\ \dot{\tau}_i \end{pmatrix} = \mathcal{F}_i(x_i) \coloneqq \begin{pmatrix} \frac{2}{\tau_i}(p_i - q_i) \\ -2k\tau_i \nabla_{q_i}\phi_i(q) \\ \eta \end{pmatrix}, \tag{4.1}$$

where $x_i = (q_i, p_i, \tau_i) \in \mathbb{R}^{2n+1}$, $k = \frac{c}{4}$, and $c > 0$ is as defined in (2.8). By introducing individual resets to obtain nonzero robustness margins, we obtain the following discrete-time dynamics for agent $i \in \mathcal{V}$:

$$\tau_i = T \Longrightarrow \begin{pmatrix} q_i^+ \\ p_i^+ \\ \tau_i^+ \end{pmatrix} = \mathcal{R}_i(x) \coloneqq \begin{pmatrix} q_i \\ \alpha_i p_i + (1 - \alpha_i) q_i \\ T_0 \end{pmatrix}. \tag{4.2}$$

In (4.2), the parameters $\alpha_i \in \{0, 1\}$ model the different individual reset policies of the players. Assuming *access to Oracles that provide real-time evaluations* of the gradient $\nabla_{q_i}\phi_i(q)$, players can implement the hybrid dynamics (4.1)-(4.2) in a fully decentralized fashion by running their own timers τ_i to enable the flows in (4.1) and trigger the jumps in (4.2). However, as demonstrated in Example 2.2 in Chapter 2, lack of coordination between player resets can hinder the acceleration properties expected from using momentum, even in potential games when all players implement the same reset policy α_i.

To address this issue, we endow each player with a distributed hybrid coordination mechanism for the resets. The coordination mechanism is implemented in a distributed fashion, with each player $j \in \mathcal{V}$ using a set-valued coordination mapping $\mathcal{C}_j : \mathbb{R}_{\geq 0} \rightrightarrows \mathbb{R}_{\geq 0}$, defined as

$$\mathcal{C}_j(\tau_j) \coloneqq \begin{cases} T & \text{if } \tau_j \in (T_0 + r_j, T] \\ \{T_0, T\} & \text{if } \tau_j = T_0 + r_j \\ T_0 & \text{if } \tau_j \in [T_0, T_0 + r_j) \end{cases}, \tag{4.3}$$

where the individual parameter $r_j > 0$ satisfies

$$r_j \in \left(0, \frac{T - T_0}{N}\right) \quad \forall j \in \mathcal{V}. \tag{4.4}$$

Using $\mathcal{C}_j$, the coordination mechanism functions as follows: whenever the timer of player i satisfies $\tau_i = T$, two events occur:

1. Player $i \in \mathcal{V}$ resets its own state $x_i = (q_i, p_i, \tau_i) \in \mathbb{R}^{2n+1}$ using the dynamics (4.2), and

2. Player i sends a pulse to its neighbors $j \in \mathcal{N}_i$, who then update their state $x_j = (q_j, p_j, \tau_j)$ as follows:

$$q_j^+ = q_j, \qquad p_j^+ = p_j, \qquad \tau_j^+ \in \mathcal{C}_j(\tau_j). \tag{4.5}$$

Since player i can only signal its neighbors, the remaining players $j \notin \mathcal{N}_i$ maintain their states constant after these two events, i.e., $x_j^+ = x_j$, for all $j \notin \mathcal{N}_i$.

The combination of continuous-time dynamics with momentum (4.1), and the set-valued discrete-time dynamics that model the coordinated resets leads to a HDS of the form (1.1), where multiple resets can happen simultaneously (in the continuous-time domain) when more than two players satisfy the condition $\tau_i = T$. To ensure that this system has suitable robustness properties we need to guarantee that small disturbances in the states, including in τ_i, do not lead to drastic changes in the behavior of the players. This property can be asserted by working with *well-posed* HDS in the sense of [16, Ch. 7]. Roughly speaking, for a HDS to be well-posed, a suitable (graphically) convergent sequence of solutions of the overall system must also converge (in a graphical sense) to another solution of the hybrid system. In the context of (4.1)-(4.5), we need to guarantee, among others, that for each $\tau_0 \in [T_0, T]$, and each graphically convergent sequence of

solutions $\{\tau_k\}_{k\in\mathbb{N}}$ with individual components $\tau_{i,k}$ satisfying

$$0 \leq \tau_{1,k}(0,0) \leq \ldots \leq \tau_{N,k}(0,0) < \tau_0, \quad \forall\, k \in \mathbb{N}, \tag{4.6}$$

and

$$\lim_{k\to\infty} \tau_{1,k}(0,0) = \ldots = \lim_{k\to\infty} \tau_{N,k}(0,0) = \tau_0, \tag{4.7}$$

the sequence $\{\tau_k\}_{k\in\mathbb{N}}$ must converge (graphically) to a function $\tilde{\tau}$ that is also a solution starting from the initial condition $\tilde{\tau}_1(0,0) = \tilde{\tau}_2(0,0) = \ldots = \tilde{\tau}_N(0,0) = \tau_0$. Thus, when $\tau_0 = T$, the above conditions imply that players will reset their timers $\tau_{i,k}$ sequentially with smaller and smaller times between resets as $k \to \infty$. It follows that in the limit, resets must also be sequential with no time between resets. Since the sequence is determined by the initial conditions, a well-posed model of the coordination mechanism must take into account *every* possible order of sequential resets of the timers τ_i. In other words, if multiple players simultaneously satisfy the condition $\tau_i = T$, then we need to consider all possible sequential resets induced by such players. As discussed in [51], this behavior is unavoidable in well-posed multi-agent HDS with decentralized discrete-time dynamics.

Following the above discussion, we construct a suitable jump map and a jump set that implement the overall coordination mechanism, including the individual reset policies $\alpha_i \in \{0,1\}$. Specifically, we introduce a new set-valued mapping $G^0 : \mathbb{R}^{2nN+N} \rightrightarrows \mathbb{R}^{2nN+N}$, which is defined to be non-empty only when $\tau_i = T$ and $\tau_j \in [T_0, T)$ with $j \neq i$, for each $i \in \mathcal{V}$, and has elements given by

$$\begin{aligned} G^0(x) \coloneqq \Big\{ & (v_1, v_2, v_3) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^{N} : (v_{1,i}, v_{2,i}, v_{3,i}) = \mathcal{R}_i(x_i), \\ & v_{1,j} = q_j, v_{2,j} = p_j, v_{3,j} \in \mathcal{C}_j(\tau_j),\ \forall\, j \in \mathcal{N}_i, \\ & v_j = x_j, \forall\, j \notin \mathcal{N}_i \Big\}, \end{aligned} \tag{4.8}$$

where $x \coloneqq (x_1, x_2, \cdots, x_N)$, $(q_i, p_i, \tau_i) = x_i \in \mathbb{R}^{2n+1}$, and where the *reset map* $\mathcal{R}_i$ and the *coordination mapping* $\mathcal{C}_j$ are defined in (4.2) and (4.3), respectively. The reset mechanisms of the players make use of three positive tunable parameters (η, T_0, T), which satisfy $0 < T_0 < T$ and $0 < \eta \leq \frac{1}{2}$, and which are selected *a priori* by the system designer.

4.2 Hybrid Momentum-Based Nash-Set Seeking Dynamics

Using the construction of G^0 in (4.8), we define the jump map of the overall hybrid system as

$$x^+ \in G_1(x) \coloneqq \overline{G^0}(z), \tag{4.9a}$$

where $\overline{G^0}$ is the outer-semicontinuous hull of G^0, [52, pp. 154], i.e., the unique set-valued mapping that satisfies $\text{graph}(G_1) = \text{cl}(\text{graph}(G^0))$. By construction, the mapping G_1 is locally bounded and outer-semicontinuous in $\mathbb{R}^n \times \mathbb{R}^n \times [T_0, T]^N$. Additionally, it preserves the sparsity properties of the graph $\mathbb{G}$ and guarantees that any pair of resets of the form (4.2) occur sequentially, thus satisfying conditions (4.6)-(4.7). Thus, using the jump map (4.9a), we can now define the *hybrid momentum-based Nash-Set-Seeking* (HM-NSS) dynamics

$$\mathcal{H}_1 \coloneqq (C_1, F_1, D_1, G_1), \tag{4.9b}$$

with overall state $x = (p, q, \tau) \in \mathbb{R}^{2nN+N}$, and jump map given by

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \end{pmatrix} = F_1(x) = \begin{pmatrix} 2\mathcal{T}^{-1}(p-q) \\ -2k\mathcal{T}\mathcal{G}(q) \\ \eta \mathbf{1}_N \end{pmatrix}, \tag{4.9c}$$

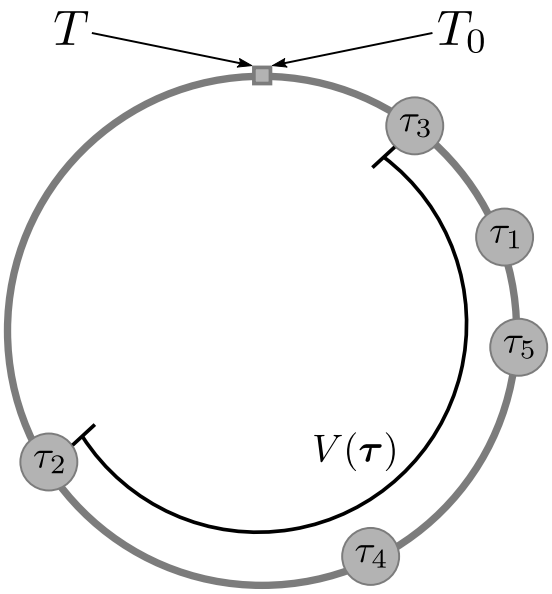


Figure 4.1. Representation of the Lyapunov function used in proof of Lemma B.5

where $p \coloneqq (p_1, p_2, \ldots, p_N)$, $\tau \coloneqq (\tau_1, \tau_2, \ldots, \tau_N) \in \mathbb{R}^N$, $\mathcal{G} : \mathbb{R}^{nN} \to \mathbb{R}^{Nn}$ denotes the pseudogradient of a game, and

$$\mathcal{T} \coloneqq \operatorname{diag}(\tau \otimes \mathbf{1}_n). \tag{4.9d}$$

The flow set C_1 is defined as:

$$C_1 \coloneqq \Big\{ x = (q, p, \tau) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N \ : \ \tau \in [T_0, T]^N \Big\}, \tag{4.9e}$$

the jump map G_1 is given by (4.9a), and the jump set is

$$D_1 \coloneqq \Big\{ x = (q, p, \tau) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N \ : \ x \in C_1, \ \max_{i \in \mathcal{V}} \tau_i = T \Big\}. \tag{4.9f}$$

Under suitable regularity assumptions on the pseudogradient $\mathcal{G}$, and by construction of the HDS $\mathcal{H}_1$, we can establish the following lemma. This lemma outlines key regularity properties of the HDS and demonstrates fixed-time synchronization of the timers τ_i in the system, provided we can rule out finite escape times. To prove the synchronization result,

we employ a Lyapunov function V defined on a circle formed by identifying the endpoints of the interval $[T_0, T]$ by considering the infimum of arc lengths touching all timers τ_i (Figure 4.1). For a complete proof and detailed analysis, see Appendices B.1 and B.3.

Lemma 4.1. Assume that the pseudogradient $\mathcal{G}$ is continuous. Then:

1) The HDS $\mathcal{H}_1$ is well-posed in the sense of [16, Def. 6.29].

2) If, additionally, $\mathcal{G}$ is ℓ-Lipschitz continuous, then every maximal solution of $\mathcal{H}_1$ is complete, and there are at most N jumps in any continuous time interval of length $(T - T_0)/\eta$. Furthermore, for each solution $x = (q, p, \tau)$ and for all $(t, j) \in \operatorname{dom}(x)$ satisfying $t + j \geq (T - T_0)/\eta + N$, it follows that $\tau(t, j) \in \mathcal{A}_{\text{sync}}$, where the compact set $\mathcal{A}_{\text{sync}}$ is defined as follows:

$$\mathcal{A}_{\text{sync}} \coloneqq \left(\{T_0, T\}^N\right) \cup (\mathbf{1}_N \cdot [T_0, T]) \,. \tag{4.10}$$

Figure 4.2 illustrates a block-diagram representation of the hybrid dynamics for each player. These dynamics, and closely related modifications, will be employed throughout the rest of the chapter in several contexts, including Distributed Nash Equilibrium Seeking, Congestion games, and Distributed Consensus-Based Optimization.

4.3 Nash Equilibrium Seeking in Potential Games

The study of equilibria in noncooperative multi-agent systems has garnered significant attention due to its wide-ranging applications in engineering and socio-economic contexts, from optimizing resource allocation in communication networks [53] to managing energy markets [54]. Among the various notions of equilibria related to game-theoretic models, the Nash equilibrium (NE), introduced in [55], has become ubiquitous in many engineering and socio-technical systems, such as transportation networks and energy markets. Recall from Section 2.2 that a NE is defined as an action profile $q^* \in \mathbb{R}^{nN}$ that

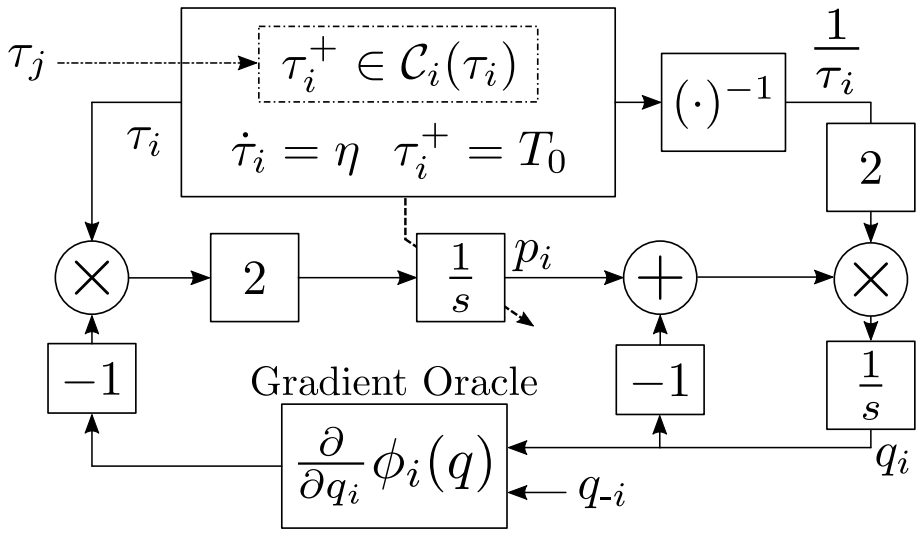


Figure 4.2. Scheme of Individual hybrid momentum-based distributed dynamics dynamics. Periodic coordinated resets restart the state p_i and the timer τ_i.

satisfies

$$\phi_i(q_i^*, q_{-i}^*) = \inf_{q_i \in \mathbb{R}^n} \phi_i(q_i, q_{-i}^*), \quad \forall\, i \in \mathcal{V}. \tag{4.11}$$

where $\phi_i : \mathbb{R}^n \to \mathbb{R}$ denotes the cost function of player $i \in \mathcal{V} \coloneqq \{1, 2, \cdots, N\}$, in a noncooperative game with N players. To converge to this equilibrium, researchers have developed a variety of deterministic and stochastic Nash set seeking (NES) algorithms over the past decades [56, 57, 58, 59, 34, 60, 61, 62].

Potential games [63] have emerged as a particularly tractable and insightful class of game-theoretic models, offering a unique bridge between game theory and optimization theory [64]. These games are characterized by the existence of a global potential function that encapsulates the incentives of all players, thereby aligning individual strategic choices with the optimization of a collective objective [65]. The defining feature of potential games is the existence of a function $P : \mathbb{R}^{nN} \to \mathbb{R}$ such that for all players $i \in \mathcal{V}$ and all strategy profiles $q = (q_1, q_2, \cdots, q_N) \in \mathbb{R}^{nN}$:

$$\frac{\partial P(q)}{\partial q_i} = \frac{\partial \phi_i(q)}{\partial q_i}, \tag{4.12}$$

where $\phi_i : \mathbb{R}^n \to \mathbb{R}$ is the cost function for player i. This structure implies that the pseudogradient of the game is equal to the gradient of the potential function, i.e., $\nabla P(q) = \mathcal{G}(q)$. Consequently, the problem of finding Nash equilibria in potential games can be recast as an optimization problem, allowing researchers to leverage powerful tools from optimization theory.

In the context of game-theoretic control system design, many results in the literature are inspired by or related to the time-invariant pseudogradient (PSG) flows studied by Rosen in [34, Eq. (3.9)], which take the form $\dot{q} = -\mathcal{G}(q)$. In potential games, PSG flows are equivalent to gradient descent on the potential function, inheriting its convergence properties. For general convex potential functions, these flows converge to Nash equilibria at a rate of $\mathcal{O}(1/t)$. In the case of strongly convex potential functions, convergence accelerates to an exponential rate of $\mathcal{O}(e^{-\kappa t})$, where κ denotes the strong convexity of the potential function P.

Despite these guarantees, PSG flows can exhibit slow convergence in practice, particularly when the potential function has shallow curvature. Motivated by this limitation and leveraging promising results from centralized optimization, in this section we present theoretical certificates demonstrating that the hybrid dynamical system $\mathcal{H}_1$, introduced in (4.9), can efficiently and robustly find the set of Nash equilibria $\mathcal{A}_{\text{NE}}$.

For a potential game with a *convex* potential function $P : \mathbb{R}^{nN} \to \mathbb{R}$ and pseudogradient $\mathcal{G}$, it follows that q^* is a NE if and only if $\mathcal{G}(q^*) = 0$ [66, Prop. 2.1]. Formally, we define the set of Nash equilibria as:

$$\mathcal{A}_{\text{NE}} \coloneqq \{q \in \mathbb{R}^{nN} \ : \ \mathcal{G}(q^*) = 0\}. \tag{4.13}$$

We study the stability and convergence properties of the HDS $\mathcal{H}_1$ with respect to a compact set $\mathcal{A}$, which includes both the Nash equilibria and the synchronization of the timer states

τ_i. This set is defined as:

$$\mathcal{A} \coloneqq \mathcal{A}_{qp} \times \mathcal{A}_{\text{sync}}, \tag{4.14}$$

where

$$\mathcal{A}_{qp} \coloneqq \{(q,p) \in \mathbb{R}^{2nN} : p = q, q \in \mathcal{A}_{\text{NE}}\}, \tag{4.15}$$

and $\mathcal{A}_{\text{sync}}$ is as defined in (4.10).

The following theorem presents the main result of this section, establishing uniform asymptotic stability of and accelerated convergence towards the set $\mathcal{A}$ defined in (4.14) under the hybrid dynamical system $\mathcal{H}_1$; the proof is presented in Appendix B.1. These results demonstrate the efficacy of our proposed approach in addressing the limitations of traditional PSG flows studied by Rosen in [34], as well as those of decentralized implementations of Nesterov's ODE that do not incorporate distributed and coordinated resets.

Theorem 4.1. Let P be the potential function of a potential game with pseudo-gradient $\mathcal{G}$. Assume that that P is convex and radially unbounded, and that $\mathcal{G}$ is ℓ-Lipschitz continuous, and consider the HDS $\mathcal{H}_1$ in (4.9) with $k = 1$ and the resetting parameters $0 < T_0 < T$ satisfying

$$T^2 - T_0^2 > \frac{\rho_J}{2} \cdot (1 - \underline{\alpha}), \tag{4.16}$$

where $\underline{\alpha} \coloneqq \min\{\alpha_i\}_{i \in \mathcal{V}}$, and $\alpha_i \in \{0, 1\}$ denotes the individual resetting policy of agent i as defined in (4.2).

(i_1) If $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N) = \mathbf{1}_N$ and $\rho_J \geq 0$ then the set $\mathcal{A}$ is UGAS. Additionally, for any solution $x = (q, p, \tau)$ of the HDS $\mathcal{H}_1$, during flows, the potential function

satisfies the bound

$$P(q(t,j)) - P(\mathcal{A}_{\mathrm{NE}}) \leq \frac{c_j}{\tau_i^2(t,j)}, \text{ for all } i \in \mathcal{V}, \tag{4.17}$$

all $(t,j) \in \mathrm{dom}\,(x)$ satisfying $t + j \geq (T - T_0)/\eta + N$, and where $\{c_j\}_{j=0}^{\infty}$ is a sequence of monotonically decreasing positive numbers that depends on $x(0,0)$ and satisfies $c_j \to 0^+$.

(i_2) If $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N) \in \{0,1\}^N$, P is κ-strongly convex and $\rho_J = \kappa^{-1}$, then the set $\mathcal{A}$ is UGES, and there exists $\lambda > 0$ such that for each compact set $K_0 \subset C_1 \cup D_1$ there exists $M_0 > 0$ such that for all solutions x with $x(0,0) \in K_0$, and for all $(t,j) \in \mathrm{dom}(x)$ the following bound holds:

$$|q(t,j) - q^*| \leq M_0 e^{-\lambda(t+j)}. \tag{4.18}$$

(i_3) If $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N) = \mathbf{0}_N$, P is κ-strongly convex and $\rho_J = \kappa^{-1}$, then the set $\mathcal{A}$ is UGES, and for each compact set $K_0 \subset C_1 \cup D_1$ there exists $M_0 > 0$ such that all solutions x with $x(0,0) \in K_0$, and for all $(t,j) \in \mathrm{dom}\,(x)$ the following bound holds:

$$|q(t,j) - q^*| \leq \frac{T}{T_0}\sqrt{\frac{\ell}{\kappa}}\,(1 - \gamma(\rho_J))^{\zeta(j)/2}\, M_0,$$

where $\zeta(j) \coloneqq \max\{0, \lfloor \frac{j-N}{N} \rfloor\}$ and

$$\gamma(\rho_J) \coloneqq \left(1 - \frac{T_0^2}{T^2} - \frac{\rho_J}{2T^2}\right) \tag{4.19}$$

satisfies $\gamma(\rho_J) \in (0,1)$.

The results of Theorem 4.1 establish robust NSS for $\mathcal{H}_1$ in potential games with

convex and strongly convex potential functions. Similar to the centralized case in Chapter 3, and unlike Nesterov's ODE, for the hybrid dynamics $\mathcal{H}_1$ there exists a class $\mathcal{KL}$ function β such that a bound of the form (10.9) holds under small bounded additive disturbances on the dynamics. This effectively rules out the instability under disturbances presented in Chapter 2, Section 2.1. The bounds of Theorem 4.1 also establish suitable semi-acceleration properties in both monotone and strongly monotone games. Such bounds will eventually hold since the UGAS result also implies that for all times (t, j) such that $t + j \leq (T - T_0)/\eta + N$, the trajectories remain (uniformly) bounded, and Lemma 4.1 guarantees completeness of solutions. Specifically, solutions of $\mathcal{H}_1$ exhibit a "transient phase", where the momentum coefficients synchronize to each other, followed by a "semi-acceleration phase" where the system behaves as having one global momentum coefficient coordinating the overall network. Figures 4.3 and 4.4 illustrate the advantages of the hybrid NES dynamics $\mathcal{H}_1$ in potential games compared to a decentralized implementation of Nesterov's ODE (see (2.8)) that does not implement the robust coordinated resetting mechanism incorporated in $\mathcal{H}_1$.

Remark 4.1. When all players implement the reset protocol $\alpha_i = 1$, item (i_1) establishes a *semi-acceleration* property of order $\mathcal{O}(1/\tau^2)$ that holds during intervals of flow happening with hybrid times $(t, j) \in \mathrm{dom}\,(x)$ satisfying $t+j \geq (T-T_0)/\eta+N$. Since such intervals of flows have a length proportional to $T - T_0$, they can be made arbitrarily large by increasing T. Moreover, if all players initialize their coefficients as $\tau_i(0, 0) = T_0$, then during the first interval of flow we have that $P(q(t, 0))-P(q^*) \leq \frac{d_0}{t^2}$, for all $(t, 0) \in \mathrm{dom}(x)$, where $d_0 > 0$ is fully determined by the initial conditions of the system and the properties of $\mathcal{G}$. To the best knowledge of the authors, the result of Theorem 4.1-(i_1) is the first in the literature that establishes UGAS *and* this type of acceleration property in *distributed* NES dynamics. Centralized convergence results without resets were recently studied independently in [67].

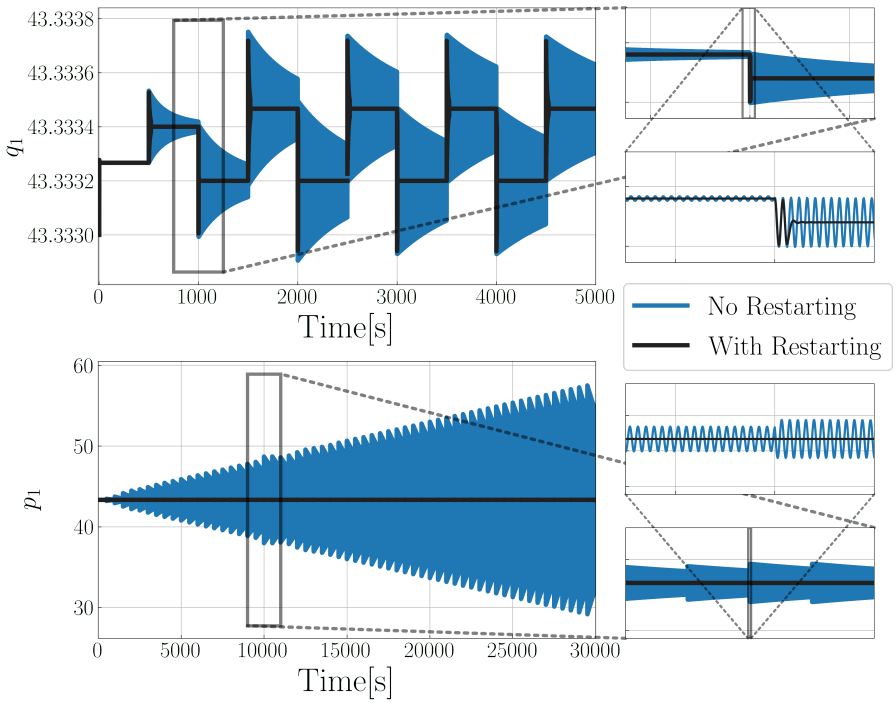


Figure 4.3. Instability of (2.1) in when using perturbed gradients. The instability can be removed by incorporating resets, which generate the stable trajectories shown in black.

> **Remark 4.2.** For potential games with κ-strongly-convex potential function P, the reset policy $\alpha_i = 0,\ \forall i \in \mathcal{V}$, guarantees exponential NSS with rate of convergence dictated by $1 - \gamma(\kappa^{-1})$. In this case, by borrowing results from the literature on centralized accelerated optimization [25, 31], we can consider a "quasi-optimal" restarting parameter $T = e\sqrt{\frac{1}{2\kappa} + T_0^2}$, which guarantees exponential convergence of order $\mathcal{O}(e^{-\sqrt{\kappa}t})$ whenever $T_0 \ll 1$. Finally, the result of item (i_2) shows that the stability and convergence properties of $\mathcal{H}_1$ are robust to heterogeneous reset policies in the game.

The analysis presented in this and the subsequent sections for potential games establishes a robust foundation for addressing complex multi-agent optimization scenarios with momentum-based methods. In the following sections, we demonstrate the versatility of our approach by presenting specializations and closely related hybrid mechanisms to

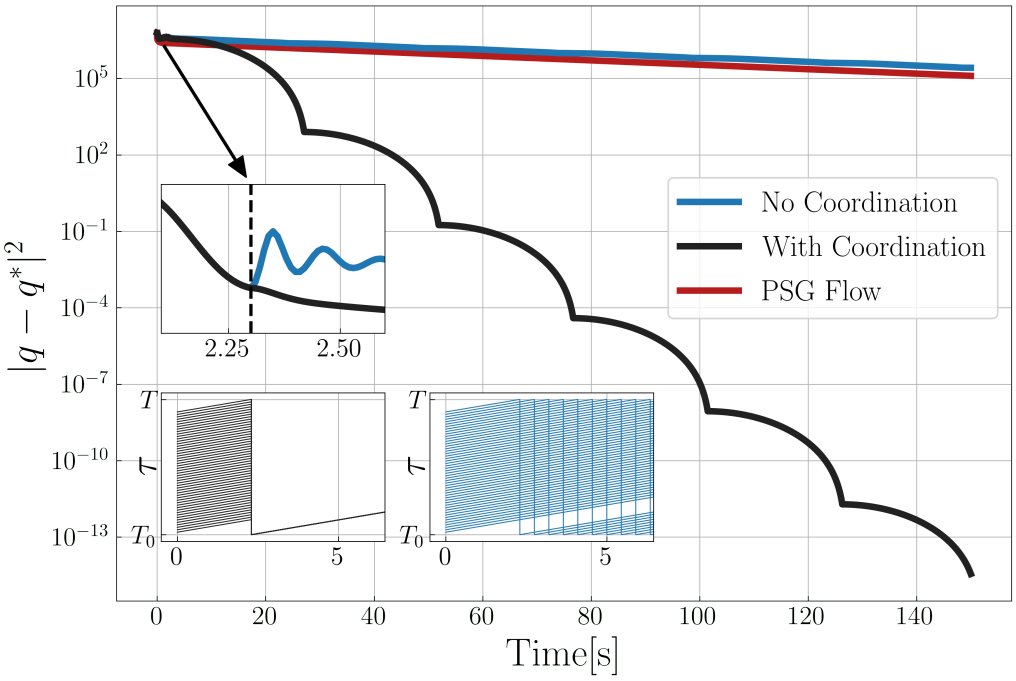


Figure 4.4. Coordinated vs non-coordinated resets in a quadratic κ-strongly monotone potential-game with $\kappa = 0.01, \ell = 100$ and $n = 30$. The insets show the evolution of the states τ_i with and without coordination mechanisms.

$\mathcal{H}_1$ in the context of congestion games and Consensus-Based Optimization over networks. These applications illustrate the broad applicability of our method in diverse distributed decision-making scenarios.

Building upon this foundation, Chapter 5 extends these concepts to non-potential games and network settings. There, we tackle more challenging problems, including those involving partial information and model-free learning. This progression from potential games to more complex scenarios underscores the scalability and adaptability of our hybrid dynamical systems approach in solving a wide range of distributed optimization problems.

4.4 Transactive Control in Congestion Games

In recent years, there has been a growing interest in the study of decentralized resource allocation problems with competitive users in large-scale network systems, in-

cluding transportation networks, power grids, and the internet [68, 69, 70]. The growth in the scale of such networks difficulties the implementation of centralized solutions to problems that involve trade-offs between social system-level efficiency and selfish individual performance. Many efforts have been devoted to address these issues, and to the design of localized control laws that ensure desirable global system performance under the presence of self-interested agents [71, 72, 73, 74]. Game theory provides a collection of mathematical tools that are instrumental for the analysis and design of such systems [75, 69, 76, 77]. In games, self-interested rational decision-makers are usually referred to as *players* with a *strategy set*, or *resource set*, and an individual *payoff function*. Players often use a distributed *learning algorithm* to iteratively update their actions until convergence to a suitable equilibrium point is achieved. The most common equilibrium of interest is the so-called *Nash Equilibrium* (NE) [75], which describes a profile of actions where players have no unilateral incentive to deviate. However, it is well-known that while NEs can provide a notion of individual optimality for the players, they can lead to poor social outcomes as measured by networked-wide welfare functions. Therefore, to maximize the performance of multi-agent engineering systems controlled via local feedback laws, it is essential to align the emerging Nash equilibria with the socially optimal point of the system. In the static scenario (i.e., no iterative learning dynamics), this methodology is referred to in the literature as *mechanism design* [76, 78].

In this section, we focus on a particular class of games referred to as *congestion games*, where a fixed amount of resources must be allocated among n different strategies, and the payoff related to each strategy depends on the total allocation. In such types of games, the notion of Nash flow, or Wardrop equilibria, has been used to characterize resource allocations that are optimal from the individual strategy point of view. However, since Wardrop equilibria might not be socially optimal, social planners are faced with the challenge of designing suitable incentives (e.g., tolls in transportation systems, prices in power systems, etc) such that the emerging Nash flows are also socially optimal. To solve

this challenge, different types of dynamic pricing algorithms have been considered in the literature [79, 80, 81, 82, 83]. To guarantee that the system continuously operates at its optimal point, pricing algorithms must react quickly to changes in traffic demand, weather conditions, road accidents, etc. This adaptability requirement, similar to the "alertness" property of feedback control systems, has motivated the development of different recursive algorithms that iteratively update the incentives as the system operates [84, 85, 86, 87, 88, 89].

Recently, the work in [90] introduced a class of decentralized gradient-based pricing dynamics (G-PD) that achieve global convergence to socially optimal incentives in a class of affine congestion games. As shown in [90], these distributed welfare gradient dynamics guarantee exponential convergence, with a rate of convergence of order $\mathcal{O}(\kappa)$, where κ defines the strong monotonicity properties of the flow map. However, this convergence rate can be quite slow in problems where $\kappa \ll 1$. In such situations, one may hypothesize that *momentum-based dynamics* that incorporate momentum might achieve a better transient performance compared to first-order gradient-based algorithms.

To study this hypothesis we consider a congestion game [91] with N possible strategies, where $i \in \mathcal{V} \coloneqq \{1, \dots, N\}$ denotes the i^{th} strategy of the game. Additionally, we assign to each strategy i a node in a graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E}$ is the set of edges or links. We let $z_i \in [0, 1]$ be the proportion of a fixed resource that is allocated to the i^{th} strategy. The vector of allocations is then defined as $z \coloneqq (z_1, \dots, z_N)$, which belongs to the simplex

$$\Delta = \{z \in \mathbb{R}^N \ : \ \mathbf{1}_N^\top z = 1, \ z_i \geq 0\}.$$

Moreover, we consider that each strategy has an associated cost of the form:

$$\tilde{c}_i(z_i, q_i) = c_i(z_i) + \delta \cdot q_i, \tag{4.20}$$

with $(z_i, q_i) \in \Delta \times \mathbb{R}$, and where c_i represents the incentive-free cost of choosing strategy i, the scalar $q_i \in \mathbb{R}$ denotes an external incentive input, and $\delta \in \mathbb{R}_{>0}$ is a sensitivity parameter. We define the vector of incentives as $q \coloneqq (q_1, \cdots, q_N) \in \mathbb{R}^N$, and we use $\tilde{c}(z, q) = (\tilde{c}_1(z_1, q_1), \ldots, \tilde{c}_N(z_N, q_N))$ to denote the vector of costs of the congestion game under the influence of the external input q. Under suitable monotonicity properties on the cost functions, congestion games are potential games with potential function [92, Sec 2.4]

$$P_q(z) = \sum_{i=1}^{N} \int_0^{z_i} \tilde{c}_i(\zeta, q_i) d\zeta. \tag{4.21}$$

For every fixed exogenous value $q \in \mathbb{R}^N$, a *Nash flow* of the congestion game corresponds to a particular resource allocation $z_q^f \in \Delta$ that minimizes P_q. Hence, by the KKT conditions, a Nash flow must satisfy

$$-\tilde{c}_i\left(z_{i,q}^f, q\right) + \mu + \lambda_i = 0 \qquad \forall i \in \mathcal{V} \tag{4.22a}$$

$$z_q^f \in \Delta, \quad \lambda_i z_{i,q}^{Nf} = 0, \quad \lambda_i \geq 0 \qquad \forall i \in \mathcal{V}. \tag{4.22b}$$

In general, Nash flows might not be socially optimal. To formally quantify the social optimality of a given allocation z, the concept of social welfare is introduced.

Social welfare: The social welfare is defined as:

$$W(z) \coloneqq -\sum_{i=1}^{N} c_i(z_i) z_i. \tag{4.23}$$

Consequently, a *socially optimal* flow z^* corresponds to the social state that maximizes (4.23). Thus, by the KKT conditions, z^* satisfies:

$$-\frac{\partial c_i}{\partial z_i}(z^*) z_i^* - c_i(z_i^*) + \tilde{\mu} - \tilde{\lambda}_i = 0, \qquad \forall i \in \mathcal{V}, \tag{4.24a}$$

$$z^* \in \Delta, \quad \tilde{\lambda}_i z_i^* = 0, \quad \tilde{\lambda}_i \geq 0, \qquad \forall i \in \mathcal{V}. \tag{4.24b}$$

We design a class of dynamic pricing mechanism to render Nash flows socially optimal with better transient-performance. To this end, we consider a subclass of congestion games that satisfy the following assumption.

Assumption 4.1. For the congestion game with $N \in \mathbb{Z}_{\geq 1}$ strategies, the following conditions are satisfied:

(i_1) *Affine costs:* There exists a positive definite diagonal matrix $A \in \mathbb{R}^{N\times N}$ and $b \in \mathbb{R}^N$ such that $c(z) = Az + b$ for all $z \in \mathbb{R}^N$.

(i_2) *Full utilization:* For every incentive $q \in \mathbb{R}^N$, the corresponding Nash flow $z_q^f \in \text{relint}(\Delta)$, i.e, $\left(z_q^f\right)_i \neq 0$ for all $i \in \mathcal{V}$.

Remark 4.3. While Assumption 4.1 is conservative, affine congestion games are commonly found in various societal systems, including parallel network routing and traffic problems [78]. Moreover, since sets of incentives are bounded in most practical applications, it is reasonable to assume that every strategy $i \in \mathcal{V}$ will receive a positive allocation, resulting in the full utilization scenario.

Assumption 4.1 ensures the strong convexity of the potential function in (4.21), which implies that conditions (4.22) hold for a unique Nash Flow z_q^f for every $q \in \mathbb{R}^N$. We refer to the mapping $\mathcal{O}(q) \coloneqq z_q^f$ as the *oracle mapping*. The following lemma provides a characterization of the oracle mapping for the types of games that satisfy Assumption 4.1. All proofs are presented in Appendix B.2.

Lemma 4.2. Under Assumption 4.1 the oracle mapping $\mathcal{O}(\cdot)$ satisfies:

$$\mathcal{O}(q) = -Q(b + \delta \cdot q) + \alpha, \tag{4.25}$$

where

$$\alpha \coloneqq \frac{A^{-1}\mathbf{1}_N}{\mathbf{1}_N^\top A^{-1}\mathbf{1}_N},$$

and where $Q \in \mathbb{R}^{N\times N}$ is the Laplacian matrix of a graph with adjacency matrix

$$\mathbb{A} \coloneqq \frac{A^{-1}\mathbf{1}_N\mathbf{1}_n^\top A^{-1}}{\mathbf{1}_N^\top A^{-1}\mathbf{1}_N},$$

that is

$$Q \coloneqq \left(I - A^{-1}\frac{\mathbf{1}_N\mathbf{1}_N^\top}{\mathbf{1}_N^\top A^{-1}\mathbf{1}_N}\right)A^{-1}. \tag{4.26}$$

We assume that the dynamics describing the convergence to the Nash flow under a given incentive q are instantaneous, and therefore can be omitted. This assumption can be justified using singular perturbation techniques for multi-time scale dynamical systems, see Section placeholder. Now, note that Assumption 4.1-(i_1) guarantees strong concavity of the welfare function W, which implies that the system of equations (4.24) are satisfied for a unique socially optimal flow z^*. However, according to the following lemma, the incentives q that generate this resource allocation state may not be unique, see also [78, 90].

Lemma 4.3. Suppose that Assumption 4.1 holds. Then, the set of incentives that generate socially optimal Nash flows via the costs (4.20) is given by

$$\mathcal{A}_q \coloneqq \{q \in \mathbb{R}^N \ : \ q = q^* + \mu\mathbf{1}_N,\ \mu \in \mathbb{R}\}, \tag{4.27}$$

where $q^* = \frac{-b}{2\delta}$.

Momentum-Based Dynamics for Congestion Games

To control the incentives, we assume that the dynamic pricing mechanism has access to *on-the-fly* measurements of the information tuple

$$\mathcal{I} \coloneqq \{\mathcal{O}(q), c\left(\mathcal{O}(q)\right), \frac{\partial c}{\partial z}\mathcal{O}(q)\},$$

where $\mathcal{O}(\cdot)$ was defined in (4.25). Due to the large-scale nature of congestion games, it may not be practical to measure $\mathcal{I}$ centrally. Instead, we assume that each node i has access to its own information and to the information of the neighboring nodes characterized by the communication graph $\mathbb{G}$. In this way, each node will implement individual dynamics based on the received information to influence the Nash Flow of the congestion game. To simplify our presentation, we make the following assumption on the graph $\mathbb{G}$. However, we stress that this assumption can be relaxed.

Assumption 4.2. $\mathbb{G}$ is connected and undirected.

For games satisfying Assumption 4.1 and a graph satisfying Assumption 4.2, the work in [90] introduced the so-called *distributed welfare-gradient dynamics*, given by

$$\dot{q} = \gamma \mathcal{L}\mathcal{G}^{\mathcal{O}}(q), \tag{4.28}$$

where $\gamma > 0$ is a scalar gain, and where $\mathcal{G} : \mathbb{R}^N \to \mathbb{R}^N$ and $\mathcal{G}^{\mathcal{O}} : \mathbb{R}^N \to \mathbb{R}^N$ are defined as:

$$\mathcal{G}(z) \coloneqq c(z) + \frac{\partial c(z)}{\partial z} z \tag{4.29}$$

$$\mathcal{G}^{\mathcal{O}}(q) \coloneqq \left(\mathcal{G} \circ \mathcal{O}\right)(q). \tag{4.30}$$

As shown in [90], these dynamics render the set $\mathcal{A}_q$ exponentially stable. However, in some cases, the exponential convergence can be prohibitively slow since it is dictated by the

strong monotonicity constant of the mapping $q \mapsto \mathcal{L}\mathcal{G}^{\mathcal{O}}(q)$. To address this issue, and to achieve better transient performance in a decentralized fashion, we use dynamics of the form (4.9) specialized to the setting of congestion games.

The resulting dynamics, which we denote with the name *high-order pricing dynamics* (HOPD), take inspiration from the hybrid dynamical system presented in (4.9), and are described by a hybrid dynamical system with data

$$\mathcal{H}_1^{\text{HOPD}} \coloneqq (C_1^{\text{HOPD}}, F_1^{\text{HOPD}}, D_1^{\text{HOPD}}, G_1^{\text{HOPD}}), \tag{4.31}$$

and state $x = (q, p, \tau) \in \mathbb{R}^{3N}$, where $p \in \mathbb{R}^N$ is a momentum state, and $\tau \in \mathbb{R}^N$ corresponds to a set of timers $\{\tau_i\}_{i \in \mathcal{V}}$ which coordinate the evolution of the distributed dynamics. Specifically, the flow map F_1^{HOPD} of the proposed HOPD is given by

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \end{pmatrix} = F_1^{\text{HOPD}}(x) = \begin{pmatrix} 2\mathcal{T}^{-1}(p - q) \\ 2\gamma \mathcal{L}\mathcal{T}\mathcal{G}^{\mathcal{O}}(q) \\ \frac{1}{2}\mathbf{1}_N \end{pmatrix}, \tag{4.32}$$

where $\gamma \in \mathbb{R}_{>0}$ is again a tunable gain, and $\mathcal{T} \coloneqq \text{diag}(\tau)$. Note that the p-dynamics maintain the sparsity of the communication infrastructure imposed by the graph $\mathbb{G}$.

The flow set C_1^{HOPD} is defined as

$$C_1^{\text{HOPD}} \coloneqq \mathbb{R}^n \times \ker(\mathcal{L})^{\perp} \times [T_0, T]^N, \tag{4.33}$$

where (T_0, T) are tunable parameters which satisfy $T > T_0 > 0$. The proposed HOPD

algorithm employs individual restarting mechanisms given by

$$\tau_i = T \Longrightarrow \begin{pmatrix} q_i^+ \\ p_i^+ \\ \tau_i^+ \end{pmatrix} = \mathcal{R}_i^{\mathrm{HOPD}}(x_i) \coloneqq \begin{pmatrix} p_i \\ p_i \\ T_0 \end{pmatrix}, \tag{4.34}$$

where $x_i = (q_i, p_i, \tau_i) \in \mathbb{R}^3$. While a fully decentralized implementation of these dynamics might seem appealing due to its simplicity, as demonstrated in Chapter 2, Section 2.2, ensuring coordinated and synchronized restarting is crucial to achieve good transient performance and to fully exploit the advantages of incorporating momentum. To achieve this coordinated behavior we implement the distributed coordination map G_1 defined in (4.9a), with the local restarting mechanism (4.2) replaced by (4.34). We denote the resulting distributed coordination map by G_1^{HOPD}.

Finally, the jump set is defined as

$$D_1^{\mathrm{HOPD}} \coloneqq \mathbb{R}^N \times \ker(\mathcal{L})^{\perp} \times D_\tau, \tag{4.35}$$

where

$$D_\tau \coloneqq \left\{ \tau \in \mathbb{R}^N \ : \ \max_{i \in \mathcal{V}} \tau_i = T \right\}. \tag{4.36}$$

The following Lemma characterizes the synchronization certificates for the timer variable τ under the HOPD.

Lemma 4.4. Let

$$\mathcal{A}_\tau \coloneqq [T_0, T]\mathbf{1}_N \cup \left(\{T_0, T\}^N \right) \tag{4.37}$$

represent the set of points in the set $[T_0, T]^N$ where the timer variables are synchronized with a common value in (T_0, T), and, where the value of the timers can only differ from each other during jumps by taking values in the set $\{T_0, T\}$. Then, the

HDS $\mathcal{H}_1^{\text{HOPD}}$ renders the set

$$\mathcal{A}_{\text{sync}} \coloneqq \ker\left(\mathcal{L}\right)^{\perp} \times \ker\left(\mathcal{L}\right)^{\perp} \times A_{\tau}, \tag{4.38}$$

UGFxS with convergence bound $T^* \coloneqq 2\left(T - T_0\right) + N$.

By leveraging Lemma 4.4, the following theorem, corresponding to the main result of this section, characterizes stability certificates for the the hybrid dynamics $\mathcal{H}_1^{\text{HOPD}}$ with respect to the set $\mathcal{A}_q$ defined in (4.27).

Theorem 4.2. Suppose that Assumption 4.1 is satisfied, and assume that the tunable parameters (T_0, T) satisfy

$$T^2 - T_0^2 > \frac{1}{2\sigma_2\left(\mathcal{L}\right)\gamma\delta}, \tag{4.39}$$

where $\sigma_2(\mathcal{L})$ is the minimum non-zero singular value of the laplacian matrix of the graph $\mathcal{L}$. Then, the set

$$\mathcal{A} \coloneqq \left\{(q, p, \tau) \in \mathbb{R}^{3N} \ : \ p = q, \ q \in \mathcal{A}_q, \ \tau \in \mathcal{A}_{\tau}\right\}, \tag{4.40}$$

is UGES under the HDS $\mathcal{H}_1^{\text{HOPD}}$. Additionally, for every $i \in \mathcal{V}$, and for all solutions x, the following bound holds during flows:

$$|q(t,j)|_{\mathcal{A}_q}^2 \leq \frac{T}{T_0}\sqrt{\frac{\sigma_N(\mathcal{L})}{\sigma_2(\mathcal{L})}}\left(1 - \eta\right)^{\tilde{j}/2} M_0,$$

where M_0 is a constant that depends on the initial conditions, $\tilde{j} \coloneqq \max\{0, \lfloor \frac{j-N}{N} \rfloor\}$, $\sigma_N(\mathcal{L})$ is the maximum singular value of the laplacian matrix $\mathcal{L}$, and

$$\eta \coloneqq 1 - \frac{T_0^2}{T^2} - \frac{1}{2T^2\gamma\delta\sigma_2\left(\mathcal{L}\right)}.$$

Remark 4.4. Theorem 4.2 guarantees exponential stability of $\mathcal{A}_q$ with convergence rate conditioned by $1-\eta$. By following similar ideas to the literature on centralized accelerated optimization [25, 28, 93, 94], we can find a "quasi-optimal" restarting parameter $T = e\sqrt{\frac{1}{2\gamma\delta\sigma_2(\mathcal{L})} + T_0^2}$, which guarantees exponential convergence of order $\mathcal{O}\left(e^{-\sqrt{\sigma_2(\mathcal{L})}}\right)$ as $T_0 \to 0^+$.

The key technical elements of the HOPD that allow us to achieve the result of Theorem 4.2 are the resets and the coordination of the resetting timers τ_i. Indeed, as shown in [25] for centralized optimization problems, when high-order dynamics implement vanishing damping, the resulting stability properties can be lost under arbitrarily small disturbances. On the other hand, the incorporation of resets induces suitable uniformity properties in the convergence, which in turn, guarantees a minimum margin of robustness against additive disturbances [33, Ch.7]. Moreover, the reset condition (4.39) permits to leverage the decrease of a suitable Lyapunov function during resets in order to achieve accelerated convergence of order $\mathcal{O}(\sigma_2(\mathcal{L}))$, which is particularly advantageous when $\sigma_2(\mathcal{L}) \ll 1$.

Finally, we note that this result relies on achieving fixed-time synchronization of the resetting timers τ_i via the coordination map (4.3). Without such a coordination mechanism, the performance of the HOPD can be substantially inferior when compared to traditional gradient-based algorithms.

Numerical Example

We illustrate our results with a simple numerical example. We consider a total resource of 1 that needs to be allocated among 4 different nodes, i.e., $\mathcal{V} \coloneqq \{1,2,3,4\}$. The matrices and parameters that describe the payoffs of the underlying game are:

$$A = 2I_4, \quad b = \mathbf{1}_4, \quad \gamma = 0.1, \quad \delta = 1.$$

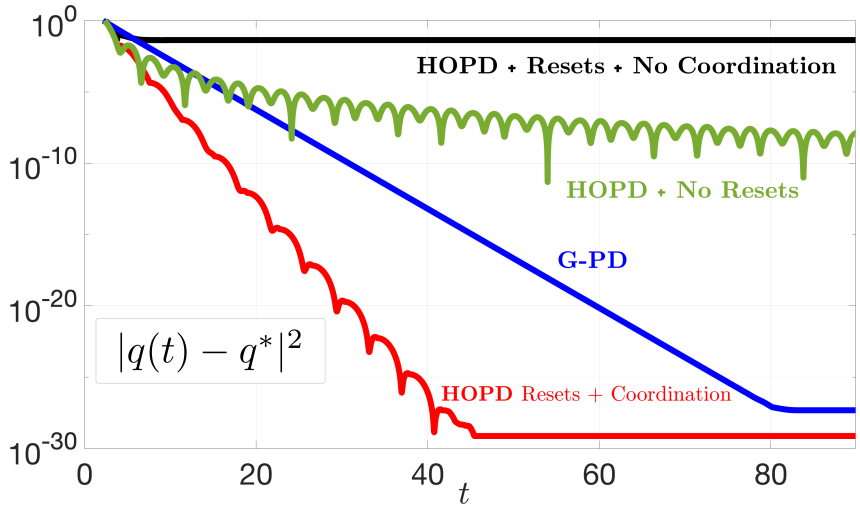


Figure 4.5. Evolution of the incentives over time using a ring communication graph.

We first simulate the G-PD of [90] and the HOPD using a communication graph characterized by a ring. Next we consider three different scenarios for the HOPD: (a) First, we consider the situation where the HOPD are implemented without resets. (b) Second, we implement the HOPD with uncoordinated resets; (c) Finally, we implement the complete HOPD with coordinated resets. All the results are presented in Figure 4.5, which shows the evolution in time of the squared error of the incentive q. As observed, the HOPD with coordinated resets generate substantially better performance compared to the standard G-PD of [90], achieving the same "steady state" error in half of the time. In this case, the restarting frequency of the HOPD was selected to be 4s. Another important observation from our numerical experiments is that using HOPD with *uncoordinated* resets generates substantially worse performance compared to the standard first-order G-PD. This poor performance is shown in the black curve of Figure 4.5. This observation highlights the role of coordination whenever dynamics with momentum and resets are implemented in multi-agent systems. Finally, we repeat our numerical example in a system with a communication graph characterized by a path. The results are presented in Figure 4.7. The restarting frequency was selected as 8s. As it can be observed, the graph's structure

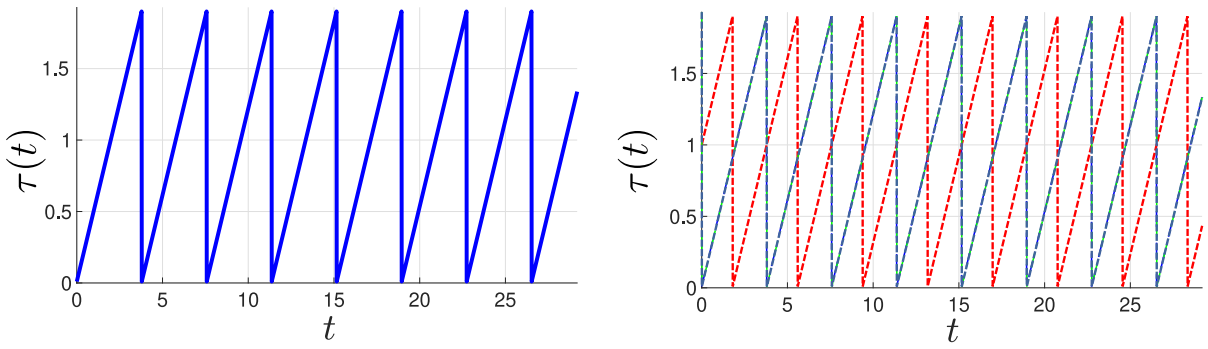


Figure 4.6. Coordinated (left) vs Uncoordinated (right) Resetting Timers in the HOPD.

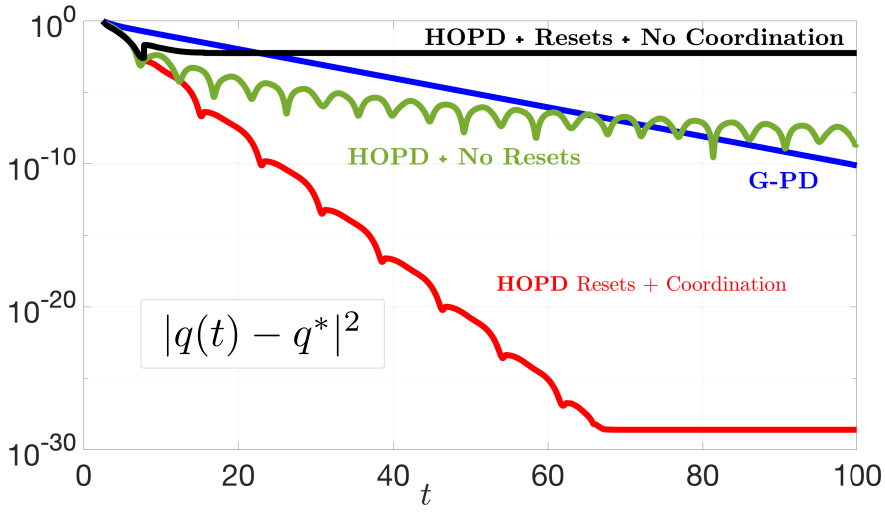


Figure 4.7. Evolution of the incentives over time using a line communication graph.

affects the transient performance of the HOPD. In particular, in this case the difference between the performance of the first-order G-PD dynamics and the HOPD algorithm is more pronounced.

In this section, we have specialized the dynamics in (4.9) to the case of games via a suitable modification of the HDS $\mathcal{H}_1$ that leverages the structure of a subclass of congestion games. This specialization leads to a class of high-order pricing dynamics for the solution of dynamic incentive problems in congestion games under a full utilization assumption. The dynamics incorporate momentum and, when combined with coordinated resets, can achieve better transient performance compared to first-order Welfare gradient

dynamics.

4.5 Accelerated Consensus-Based Optimization

In this section, we study hybrid mechanisms of the form (4.9) for the accelerated, efficient and robust solution of accelerated distributed optimization problems over network systems characterized by connected and undirected graphs $\mathbb{G} \coloneqq (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of nodes, and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges. We consider the setting where each node i has a local function $f_i : \mathbb{R}^n \to \mathbb{R}$, and the network cooperates to find a common point $\zeta^* \in \mathbb{R}^n$ that minimizes a global function defined as the summation of the local costs. This distributed optimization problem can be written as

$$\min_{z_1, z_2, \ldots, z_n \in \mathbb{R}^n} \quad \sum_{i=1}^{N} f_i(z_i), \quad \text{s.t.} \quad z_i = z_j, \quad \forall\, i, j \in \mathcal{V}, \tag{4.41}$$

which is also known in the literature as the *consensus-optimization* problem [95], and which has been shown to be relevant for several engineering applications in areas such as power systems, transportation systems, water distribution systems, and distributed network control, see [96] and references therein.

Discrete-time and continuous-time approaches to solve problem (4.41) have been extensively studied using gradient descent and Newton-based dynamics in [97], [98], primal-dual dynamics [99], and projected dynamics [100], to name just a few. However, a persistent challenge in the solution of problem (4.41) is to achieve fast rates of convergence without sacrificing essential robustness properties of the algorithms. As recently shown in [24, 29], this task is not trivial given that certain classes of accelerated continuous-time algorithms, such as Nesterov's ODE [19, 101, 18], can be destabilized under arbitrarily small disturbances on the states or gradients. Since these disturbances are unavoidable in practice, there is an urgent need for the development of robust, accelerated and distributed algorithms for the solution of problem (4.41).

In the literature of accelerated centralized optimization, one of the approaches that has received significant attention during the last years is the incorporation of restarting techniques. As a mater of fact, as shown in [19],[31], [25], [29], and [102], accelerated algorithms with restarting techniques can achieve exponential convergence rates in strongly convex optimization problems without having perfect knowledge of the condition number of the cost function. Moreover, restarting can also be used to induce suitable robustness properties in the Nesterov's ODE, provided the combination of the continuous-time dynamics and the discrete-time dynamics is carefully carried out [24]. While these ideas have been explored and validated in centralized optimization problems, as mentioned in [103], it remains an open question whether or not similar techniques could be pursued for distributed optimization problems of the form (4.41). As we will show in this chapter, the answer to this question turns out to be positive.

In this section, we formulate and analyze the *robust and* distributed *restarting-based accelerated dynamics for the solution of network optimization problems of the form* (4.41). Since our restarting dynamics combine continuous-time dynamics and discrete-time dynamics, they are modeled as set-valued hybrid dynamical systems [33], for which stability, convergence, and robustness properties can be established using Lyapunov functions and the hybrid invariance principle. The construction of this hybrid system is not trivial due to the distributed nature of the system, which allows for multiple discrete-time updates in the network happening simultaneously in the standard time domain. In contrast to existing results that use projections or primal-dual approaches, we follow a complete dual approach that allows us to recast problem (4.41) as an unconstrained optimization problem with a suitable Laplacian-dependent structure on the dynamics of the momentum variables [103]. This reformulation, allows us to establish sufficient graph-dependent restarting conditions for the solution of the primal problem.

Momentum-Based Dynamics with Distributed Restarting

To solve problem (4.41), let $z = (z_1, z_2, \cdots, z_n)$ be the concatenation of the local decision variables of the nodes of the network $\mathbb{G}$. Define the global cost function $F(z) \coloneqq \sum_{i=1}^{N} f_i(z_i)$, and let $\mathbf{L} \coloneqq \mathcal{L} \otimes I_n \in \mathbb{R}^{nN} \times \mathbb{R}^{nN}$, where $\mathcal{L}$ is the Laplacian matrix of the graph $\mathbb{G}$. We will make the following assumption on Problem (4.41):

> **Assumption 4.3.** The local cost functions f_i are $\mathcal{C}^k$ and μ_i-strongly convex, i.e., there exists $\mu_i > 0$ such that for any $x, y \in \mathbb{R}^n$, $f_i(y) \geq f_i(x) + \nabla f_i(x)^\top (y - x) + \frac{\mu_i}{2}|y - x|^2$. The graph $\mathbb{G}$ is undirected, connected, and time-invariant.

By Assumption 4.3, the extended Laplacian matrix $\mathbf{L}$ satisfies $\mathbf{L} = \mathbf{L}^\top$, $\ker(\mathbf{L}) = \operatorname{span}(\mathbf{1}_{nN})$, and $\ker(\mathbf{L})^\perp = \left\{z \in \mathbb{R}^{nN} : \mathbf{1}_{nN}^\top z = 0\right\}$. Thus, we can write Problem (4.41) as

$$\min_{z \in \mathbb{R}^{nN}} F(z), \quad \text{s.t.} \quad \mathbf{L}z = \mathbf{0}_{nN}, \tag{4.42}$$

where F is also $\bar{\mu}$-strongly convex with $\bar{\mu} \coloneqq \min_{i \in \mathcal{V}} \mu_i$. When the local gradients ∇f_i are also globally L_i-Lipschitz (a condition that we do not necessarily assume), the global gradient ∇F is globally $\bar{L}$-Lipschitz with $\bar{L} = \max_{i \in \mathcal{V}} L_i$.
To solve problem (4.42), we consider its dual problem:

$$\min_{q \in \mathbb{R}^{nN}} \phi(q), \quad \text{with} \quad \phi(q) \coloneqq \max_{z \in \mathbb{R}^{nN}} \{\langle \mathbf{L}q,\, z\rangle - F(z)\}, \tag{4.43}$$

which, as shown in [103], has zero duality gap under Assumption 4.3. By defining the mapping $h : \mathbb{R}^{nN} \to \mathbb{R}^{nN}$ as $h(u) \coloneqq \arg\max_{z \in \mathbb{R}^{nN}} \{\langle u,\, z\rangle - F(z)\}$, the gradient of the dual function ϕ can be computed as follows

$$\nabla \phi(q) = \mathbf{L} h\big(\mathbf{L}q\big). \tag{4.44}$$

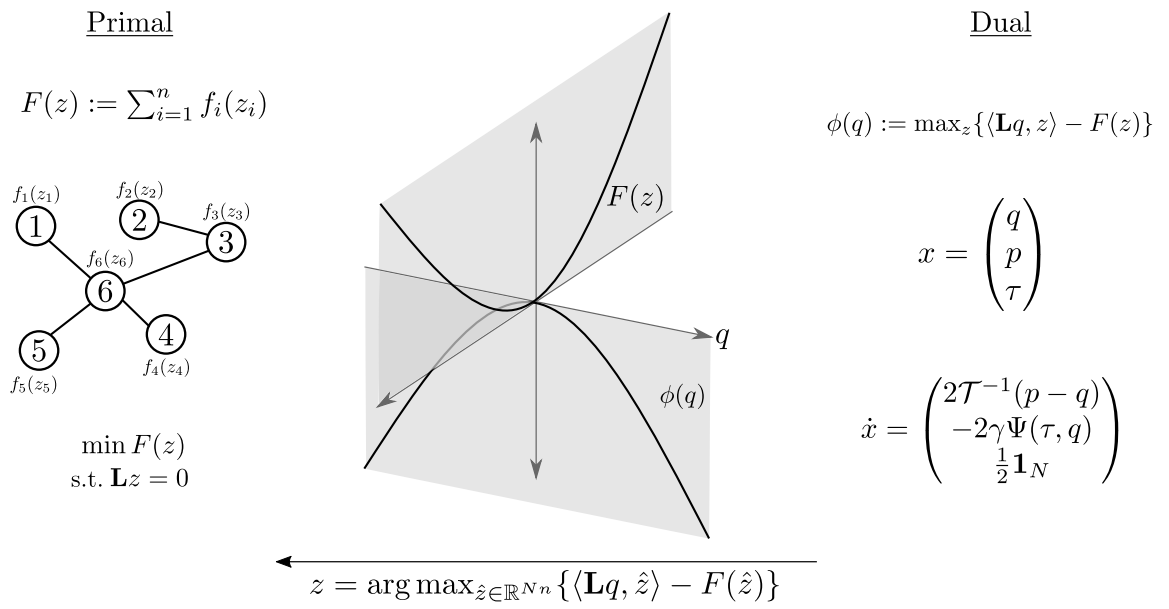


Figure 4.8. Diagram of the approach used to solve the Consensus-Based Optimization problem

Let $\mathcal{L}^2 \coloneqq \mathcal{L}^\top \mathcal{L}$, and denote as $\lambda^+_{\min}(\mathcal{L}^2)$ the smallest positive eigenvalue of $\mathcal{L}^2$, and $\lambda_{\max}(\mathcal{L}^2)$ as its largest eigenvalue. The next lemma follows directly from the results in [103].

> **Lemma 4.5.** The function ϕ is convex. Moreover, if F is $\bar{\mu}$-strongly convex and k-th continuously differentiable at the point $z \coloneqq h(\mathbf{L}x)$, then ϕ is also k-th continuously differentiable at x, and $\nabla\phi$ is globally ℓ_ϕ-Lipschitz with $\ell_\phi = \lambda_{\max}(\mathcal{L}^2)/\bar{\mu}$. If, additionally, ∇F is globally $\bar{L}$-Lipschitz, then ϕ is also μ_ϕ-strongly convex on $\ker(\mathbf{L})^\perp$, with $\mu_\phi = \lambda^+_{\min}(\mathcal{L}^2)/\bar{L}$.

Let $\mathcal{A}_\phi \subset \mathbb{R}^{nN}$ be the set of solutions of Problem (4.43) subject to $q \in \ker(\mathbf{L})^\perp$, i.e., the set defined as

$$\mathcal{A}_\phi \coloneqq \left\{ q \in \ker(\mathbf{L})^\perp \subset \mathbb{R}^{nN} \ : \ \phi(q) = \min_{\tilde{q} \in \mathbb{R}^{nN}} \phi(q) \right\}. \tag{4.45}$$

By Lemma 4.5, if ∇F is not globally Lipschitz, the set $\mathcal{A}_\phi$ may not necessarily be bounded. Therefore, we will make the following technical assumption on ϕ.

Assumption 4.4. The level sets of the dual function ϕ with domain restricted to $\ker(\mathbf{L})^{\perp}$ are bounded.

Finally, when ∇F is also globally $\bar{L}$-Lipschitz, we define the condition numbers of $\mathcal{L}^2$, and F, respectively, as $\kappa_{\mathcal{L}^2} \coloneqq \lambda_{\max}(\mathcal{L}^2)/\lambda^{+}_{\min}(\mathcal{L}^2)$, and $\kappa_F \coloneqq \bar{L}/\bar{\mu}$. These condition numbers will play an important role in the linear convergence properties of our algorithms.

Figure 4.8 illustrates our approach to Problem (4.41). We leverage the dual problem formulation (4.43) and gradient computation (4.44) to develop an algorithm using hybrid momentum-based dynamics similar to (4.9). This algorithm constructs dynamics for the dual variable $x \in \mathbb{R}^{nN}$ to solve the consensus-optimization problem. We recover the primal variable z by using the map h as follows:

$$z = h(\mathbf{L}q) = \arg\max_{\hat{z}\in\mathbb{R}^{nN}} \langle \mathbf{L}q, \hat{z}\rangle - F(\hat{z}). \tag{4.46}$$

Hybrid Dynamics with Distributed Restarting

We solve Problem (4.43) by considering a class of algorithms termed *Hybrid Accelerated Restarting Distributed Dynamics (HARDD)* with data

$$\mathcal{H}_1^{\text{HARDD}} \coloneqq (C_1^{\text{HARDD}}, F_1^{\text{HARDD}}, D_1^{\text{HARDD}}, G_1^{\text{HARDD}}). \tag{4.47}$$

To define these dynamics, each node $i \in \mathcal{V}$ is endowed with three local states (q_i, p_i, τ_i), where $q_i, p_i \in \mathbb{R}^n$, and $\tau_i \in \mathbb{R}$ is a local timer. The overall network has states $x \coloneqq (q_1, q_2, \cdots, q_N) \in \mathbb{R}^{nN}$, $p \coloneqq (p_1, p_2, \cdots, p_N) \in \mathbb{R}^{nN}$, and $\tau \coloneqq (\tau_1, \tau_2, \ldots, \tau_N) \in \mathbb{R}^N$. Using $p \coloneqq (x, y, \tau)$, and building upon the momentum-based HDS $\mathcal{H}_1$ in (4.9), we formulate the

continuous-time dynamics of the algorithms as

$$\dot{x} = \begin{pmatrix} 2\mathcal{T}^{-1}(p-q) \\ -2\gamma\Psi(\tau, q) \\ \frac{1}{2}\mathbf{1}_N \end{pmatrix} \tag{4.48}$$

where $\gamma \in \mathbb{R}_{>0}$ is a tunable gain, $\mathcal{T} \coloneqq \operatorname{diag}(\tau \otimes \mathbf{1}_n)$, and $\Psi : \mathbb{R}^n \times \mathbb{R}^{nN} \to \mathbb{R}^{nN}$ is a mapping to be defined below. The dynamics (4.48) are allowed to evolve in the flow set:

$$x \in C_1^{\text{HARDD}} \coloneqq \mathbb{R}^{nN} \times \ker(\mathbf{L})^{\perp} \times [T_0, T]^N, \tag{4.49}$$

where $T_0 > 0$ and $T > 0$ are tunable parameters. The proposed HARDD employ individual restarting mechanisms of the form

$$\tau_i = T \implies \begin{pmatrix} q_i^+ \\ p_i^+ \\ \tau_i^+ \end{pmatrix} = \mathcal{R}_i^{\text{HARDD}}(p_i) = \begin{pmatrix} (1-\alpha)q_i + \alpha p_i \\ y_i \\ T_0 \end{pmatrix}, \tag{4.50}$$

where $\alpha \in \{0, 1\}$ represents the resetting policy of the overall system, and $x_i = (q_i, p_i, \tau_i) \in \mathbb{R}^{2n+1}$. To achieve a suitable coordination of the individual restarting mechanisms and enable full exploitation of the the advantages of incorporating momentum, we implement the distributed coordination map G_1 defined in (4.9a), with the local restarting mechanism (4.2) replaced by (4.34). We denote the resulting distributed coordination map by G_1^{HARDD}. The corresponding discrete-time dynamics $x^+ \in G_1^{\text{HARDD}}(x)$ are allowed to evolve in the jump set

$$D_1^{\text{HARDD}} \coloneqq \mathbb{R}^{nN} \times \ker(\mathbf{L})^{\perp} \times D_\tau, \tag{4.51}$$

where D_τ is again given by

$$D_\tau \coloneqq \left\{ \tau \in \mathbb{R}^N \ : \ \max_{i \in \mathcal{V}} \tau_i = T \right\}. \tag{4.52}$$

The constant α in (4.50) characterizes two different restarting algorithms: If ∇F is globally Lipschitz, then $\alpha = 1$; otherwise, $\alpha = 0$. We define the mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^{nN} \to \mathbb{R}^{nN}$ in (4.48) as follows:

$$\Psi(\tau, q) \coloneqq \mathbf{L} \cdot \mathcal{T} \cdot h\big(\mathbf{L}q\big). \tag{4.53}$$

The map $\Psi(\tau, q)$ preserves the sparsity properties of the graph because the computation $h(\mathbf{L}x)$ can be carried out locally by each node, and the matrix $\mathcal{T} = \operatorname{diag}(\tau),\ \tau \in \mathbb{R}^N$ is diagonal. Thus, the complete vector field F_1^{HARDD} in (4.48) can be computed in a distributed way.

Remark 4.5. Previous implementations similar to HARDD, without resets, relied on a centralized scalar timer τ that grows unbounded [104]. As discussed in Chapter 2, Section 2.1, such dynamics cannot guarantee UGAS of the set $\mathcal{A}_\phi$ due to vanishing damping terms, making them susceptible to destabilizing disturbances.

When resets are implemented but not properly coordinated in distributed systems, as elaborated in Chapter 2, Section 2.2, prohibitively slow convergence may emerge. For distributed Consensus-Based Optimization, given the structure of problem (4.43), these issues are further amplified. Without robust coordination of individual resets, the algorithm may exhibit unstable behavior, as we illustrate below in Figure 4.10. Our HARDD approach utilizes Ψ as defined in (4.53), where in general $\Psi(q, \tau) \neq \mathcal{T}\nabla\phi(q)$ due to non-commutativity of $\mathbf{L}$ and $\mathcal{T}$. This distinction, together with the incorporation of distributed and coordinated restarting mechanisms, enables the robust and distributed implementation of the HARDD algorithm in the dual domain.

We are now ready to present the main result of this section, which establishes tuning guidelines for the HDS $\mathcal{H}_1^{\text{HARDD}}$, expressed in terms of the parameters of the primal Problem (4.41). In particular, we consider the following conditions on the tunable parameters (γ, T_0, T):

$$\alpha T_0 < \frac{1}{2}\sqrt{\frac{\bar{\mu}}{\gamma \lambda_{\max}(\mathcal{L}^2)}}, \tag{4.54a}$$

$$T > \left((2\kappa_F \kappa_{\mathcal{L}^2})^{\frac{\alpha}{2}}\right) T_0, \tag{4.54b}$$

where $\alpha \in \{0, 1\}$ denotes the resetting policy of the system. We state the stability properties of the HARDD dynamics with respect to the compact set

$$\mathcal{A} \coloneqq \mathcal{A}_{q,p} \times \mathcal{A}_\tau, \tag{4.55}$$

where the sets $\mathcal{A}_{q,p}$ and $\mathcal{A}_\tau$ are defined as follows:

$$\mathcal{A}_{q,p} : = \left\{q, p \in \mathbb{R}^{nN} \ : \ p = q, \ q \in \mathcal{A}_\phi\right\},$$
$$\mathcal{A}_\tau : = [T_0, T] \cdot \mathbf{1}_N \cup \{T_0, T\}^N.$$

Note that the set $\mathcal{A}_\tau$ describes a "synchronization" condition. We will also use $F^* := F(z^*)$ to denote the optimal value of Problem (4.42), where $z^* := \mathbf{1}_n \otimes \zeta^* \in \mathbb{R}^{nN}$, and $\zeta^* \in \mathbb{R}^n$ is the unique solution of the primal Problem (4.41). All proofs are presented in Appendix B.3.

Theorem 4.3. Suppose that Assumptions 4.3 and 4.4 hold, and let the parameters (γ, T_0, T) satisfy (4.54). Then, the following properties hold:

(P1) Every solution p of the HDS $\mathcal{H}_1^{\text{HARDD}}$ has an unbounded time domain and it is uniformly non-Zeno with at most N jumps in any time interval of length

$2(T - T_0)$.

(P2) Every solution p of the HDS $\mathcal{H}_1^{\text{HARDD}}$ satisfies $|\tau(t,j)|_{\mathcal{A}_\tau} = 0$ for all $(t,j) \in \text{dom}(p)$ satisfying $t + j \geq N + 2(T - T_0)$.

(P3) The compact set $\mathcal{A}$ defined in (4.55) is UGAS.

Additionally, for each compact set of initial conditions $K_0 \subset C_1^{\text{HARDD}} \cup D_1^{\text{HARDD}}$, and every solution $x = (q, p, \tau)$ to the HDS $\mathcal{H}_1^{\text{HARDD}}$ satisfying $x(0,0) \in K_0$, the following bounds hold for the primal hybrid arc $z(t,j) = h(\mathbf{L}q(t,j))$:

(P4) If $\alpha = 1$, there exists $c_0, \lambda_0 > 0$ such that $F(z(t,j)) - F^* \leq c_0 e^{-\lambda_0(t+j)}$, for all $(t,j) \in \text{dom}(p)$.

(P5) If $\alpha = 0$, $\tau(0,0) \in \mathcal{A}_\tau$, and $x(0,0) \in \ker(\mathbf{L})^\perp$, then $F(z(t,j)) - F^* \leq \dfrac{c_j}{\tau_i(t,j)}$, for all $i \in \mathcal{V}$, and all $(t,j) \in \text{dom}(p)$, where $\{c_j\}_{j=0}^{\infty}$ is a monotonically decreasing sequence of positive numbers.

To the knowledge of the authors, Theorem 4.3 provides the first network-dependent restarting conditions in the literature of accelerated distributed optimization over networks. The conditions in (4.54) show the dependence of the restarting parameters $0 < T_0 < T$ on the condition numbers of the primal cost function F, the graph $\mathbb{G}$. Note that these conditions can always be satisfied by taking T sufficiently large, and T_0 sufficiently small. However, as stated in property (P2), the larger $T - T_0$ is selected, the longer it will take the network to synchronize the timers. Property (P5) recovers the result of [104] which uses a centralized timer, and Property (P4) establishes linear convergence when the gradient of the primal function is globally Lipschitz. Note that when $\alpha = 0$, the conditions in (4.54) are directly satisfied with $T > T_0 > 0$.

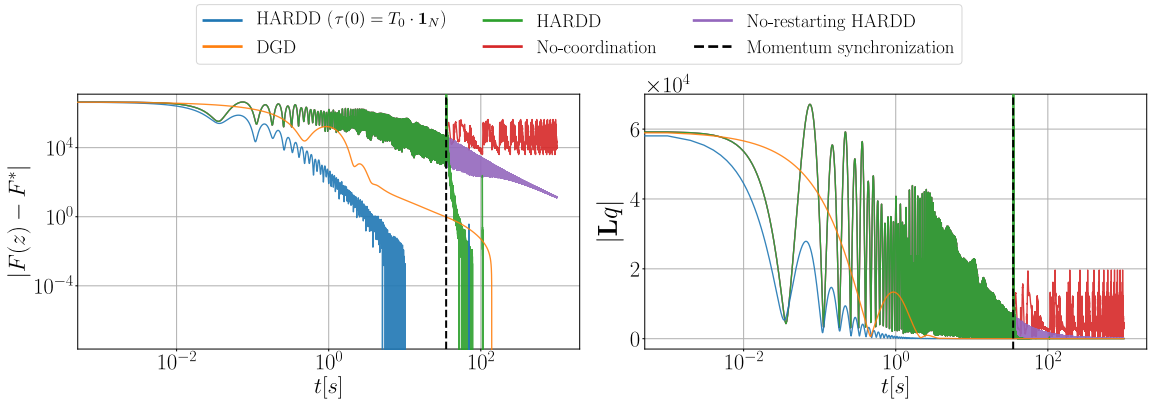


Figure 4.9. (Left) Evolution in time of the sub-optimality measure, and (Right) evolution in time of the consensus distance for different algorithms and/or initializations.

> **Remark 4.6.** (Nonzero Robustness Margins) Since the HARDD algorithms are modeled by a well-posed HDS $\mathcal{H}_1^{\text{HARDD}}$, property (P3) will be preserved, in a semi-global practical way, under *arbitrarily* small (possibly time-varying) perturbations on the states and dynamics of the algorithm. This property, which follows by [33, Lemma 7.20], is fundamental for feedback control applications where measurement disturbances are unavoidable.

Numerical Examples

We apply the HARDD algorithm with $\alpha = 1$ to solve a distributed linear regression problem over a network. The problem is specified by:

$$\min_{z \in \mathbb{R}^{nN}} \frac{1}{2nl} \sum_{i=1}^{n} \|H_i z_i - b_i\|, \quad \text{s.t.} \quad \mathbf{L}z = \mathbf{0}_{nN}, \tag{4.56}$$

where $H_i \in \mathbb{R}^{l \times n}$ and $b_i \in \mathbb{R}^l$ contain the data, l is the number of data points available per node/agent and n is the dimension of such data points. We implement the HARDD algorithm by discretizing the flows using a Runge-Kutta method of 4-th order (RK4) with step-size $dt = 1 \times 10^{-3}$.

As shown in [105], this discretization method preserves the main convergence

properties of well-posed hybrid dynamics, provided the step size is sufficiently small. Indeed, as shown in [104], RK4 can also preserve acceleration. Figure 4.9 presents the numerical results when implementing the discretization on a ring graph and with $\gamma = 1/4$, $T = 35.1$, $T_0 = 0.1$, $N = 5$, $n = 8,$ and $l = 10$. The data is generated by sampling from a normal distribution with mean 1.5 and standard deviation 3[1]. In the figure, the blue trajectory corresponds to the HARDD dynamics with synchronous initialization, i.e., $\tau_i = T_0 = 0.1$ for all $i \in \mathcal{V}$. As expected, this initialization exhibits the best behavior. The green line indicates the behavior of the HARDD dynamics with asynchronous initialization of the timers. After the synchronization event $|\tau|_{\mathcal{A}_\tau} = 0$ occurs, the sub-optimality measure decreases rapidly. For the sake of comparison, we also show the solutions obtained without incorporating the restarting mechanism (purple line), with individual restarting and no coordination between nodes/agents (red line), and using the standard *synchronous* distributed gradient descent [97]. Finally, we note that the red trajectories (restarting with no coordination) exhibit the worst behavior. Figure 4.10 illustrates the importance of coordination in the restarting mechanism; in this case without coordination the dual variables do not converge. The purple trajectory (no restarting) is slow compared to the green trajectory, and arbitrarily small disturbances can destabilize this algorithm, as we show in the example below.

Now, we compare the HARDD algorithm with and without restarting in order to stress the importance that this mechanism has in the robustness properties of the algorithm. To do so, we first note that for the dynamics with no restarting there are no convergence guarantees. However, even if these dynamics do converge, they will suffer from the same limitation of Nesterov's ODE due to the unbounded growth of the timers τ_i, i.e., they will have zero margins of robustness with respect to arbitrarily small perturbations. To explicitly show this behavior, we consider the Consensus-Based Optimization problem

[1]The code used to generate the figures in this section can be found in https://github.com/deot95/HARDD

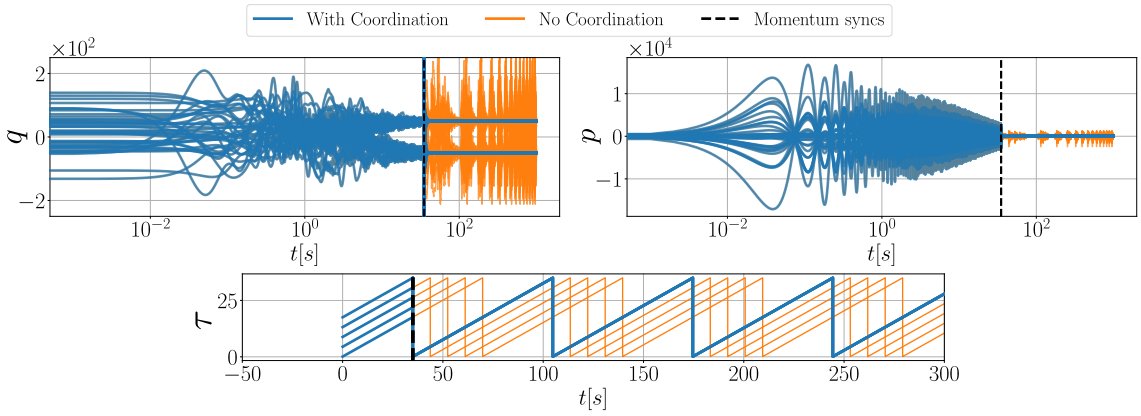


Figure 4.10. Evolution in time of the dual-variables for the HARDD algorithm with and without coordination.

with cost function

$$F(z) = \sum_{i=1}^{2} f_i(z_i) = \frac{1}{2} z^\top z, \qquad z \in \mathbb{R}^{2\times 2},$$

$N = 2$, and $n = 2$. The associated dual optimization problem is given by $\phi(q) = \frac{1}{2} q^\top \mathbf{L}^2 q$, where $\mathbf{L} = \mathcal{L} \otimes I_n \in \mathbb{R}^{4\times 4}$ is the communication matrix of the 2-agent system. We simulate the HARDD algorithm with ε-perturbed measurements of the state, i.e., with

$$\Psi(\tau, q) = \mathbf{L}\mathbf{D}(\tau \otimes \mathbf{1}_p) h\Big(\mathbf{L}\big(x + \varepsilon(t)\big)\Big), \tag{4.57}$$

where $\sup_{t\geq 0} |\varepsilon(t)| = 0.001$ and $\varepsilon(t) \in \ker(\mathbf{L})^\perp$ for all $t \geq 0$. In particular, we consider the following disturbance:

$$\varepsilon(t) = 0.001\epsilon(t)[1, 1, -1, -1]^\top, \tag{4.58}$$

where $\epsilon : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a square periodic signal with unitary amplitude and period equal to 1×10^4. For the case when $T - T_0 = \infty$, (the same setup that generates the purple line in Figure 4.9) the effect of the disturbance ε induces the instability shown in Figure 4.11-(a). Conversely, when the restarting is activated, we obtain the robust stable behavior shown in Figure 4.11-(b).

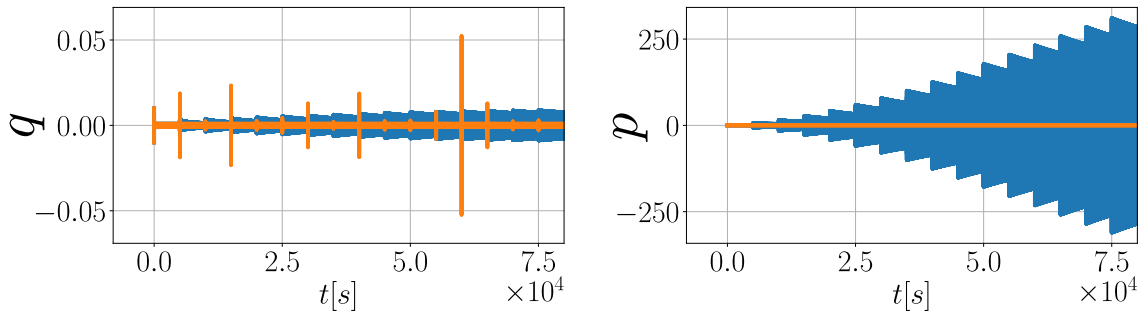


Figure 4.11. (Blue) Instability of the optimization dynamics under disturbance $\varepsilon(t)$ when no restarting is implemented, i.e., $T - T_0 = \infty$. As observed, the trajectories of the system diverge. (Orange) Robust asymptotic stability under the same disturbance $\varepsilon(t)$ with the HARDD algorithm and $T - T_0 = 35$.

In this chapter, we introduced hybrid dynamical systems that model distributed momentum-based dynamics with coordinated resets. This approach addresses the robustness and coordination challenges that momentum-based dynamics with vanishing damping coefficients face in potential games and distributed optimization settings. We applied our method to two key areas: momentum-based transactive control in congestion games and distributed Consensus-Based Optimization over networks. Our approach demonstrated improved performance and robustness compared to traditional gradient-based methods, as supported by theoretical stability guarantees and numerical examples.

The next chapter will explore the adaptation of these hybrid momentum-based dynamics to nonpotential games, as well as their application in a subclass of distributed learning problems with asymmetric communication constraints. As explained in Chapter 2, Section 2.3, in these settings, dynamics without resets may fail to provide suitable convergence properties, highlighting the importance of our hybrid approach.

Acknowledgements

Sections 4.1, 4.2, and 4.3, in part, are a reprint of the material as it appears in "Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. Section 4.5, in full, is a reprint of the material as it appears in "Robust

Optimization Over Networks Using Distributed Restarting of Accelerated Dynamics" [4]. The dissertation author is the first author of these papers. Section 4.4, in full, is a reprint of the material as it appears in "High-Order Decentralized Pricing Dynamics for Congestion Games: Harnessing Coordination to Achieve Acceleration" [3]. The dissertation author was one of the primary investigators of this paper.

Chapter 5

Distributed Momentum-Based Dynamics in Settings with Non-Symmetric Jacobian

Building upon our exploration of Momentum-Based Nash Set Seeking dynamics in potential games, we now turn our attention to the more complex domain of *non-potential games*. The absence of a potential function that establishes a common cost landscape for the players in these games introduces unique challenges that require careful consideration and preclude the direct application of results from the optimization literature. This fundamental difference necessitates a reevaluation of Nash Set Seeking (NSS) strategies and presents new obstacles, particularly in distributed and model-free settings.

Indeed, the incorporation of momentum with time-varying damping, as introduced in equation (2.8) and explained in Chapter 2, offers both opportunities and risks when applied to non-potential games. In Chapters 3 and 4, we demonstrated suitable acceleration and stability properties of the hybrid momentum-based mechanisms in approximate dynamic programming, potential games, congestion games, and distributed consensus optimization. However, as detailed in Chapter 2 (Section 2.3), their direct application to non-potential settings can lead to instabilities. These instabilities persist even with the introduction of slow resets, highlighting the limitations of straightforward adaptations of these methods to non-potential settings.

Despite these challenges, the use of resets in momentum-based dynamics with time-varying damping is not entirely precluded in non-potential games. While the aforementioned

features might suggest that momentum-based dynamics are problematic for games, it so happens that distributed implementations of Nesterov's ODE can, in fact, be employed to efficiently and robustly find Nash Equilibria (NE) in a decentralized manner. This is achievable when these dynamics are combined with suitable distributed discrete-time dynamics that persistently reset some of the players' states in a *coordinated* fashion. However, in contrast to optimization problems [19, 31, 25], for general (non-potential) noncooperative games, the frequency of the *resets must occur within a specific frequency band* to simultaneously achieve stability and acceleration.

5.1 Momentum-Based Nash Set-Seeking over Networks via Multi-Time Scale Hybrid Dynamic Inclusions

In game-theoretic control system design, recent approaches for non-potential games have aimed to incorporate momentum with time-varying damping in the dynamics for Nash set seeking (NSS). However, these methods have not fully addressed their applicability to distributed setups. Some studies have examined momentum-based dynamics with maximally monotone operators $\mathcal{G}$ via Yosida regularizations [106]. Yet, these approaches typically rely on centralized computations or global information, rendering them unsuitable for distributed implementations where players have limited information.

The derivation of stability and robustness guarantees for momentum-based dynamics in non-potential games is a crucial step that extends beyond mere theoretical exploration. It represents a fundamental milestone in the formulation of extended Nash set seeking (NSS) algorithms with additional mechanisms. For instance, in the context of pseudogradient (PSG) flows, stability and convergence results in potential games have been instrumental in designing extended NSS algorithms. These algorithms integrate additional mechanisms based on fast consensus dynamics [62, 107], projections [35], inertia [108], proportional feedback terms [60], tracking terms [109], and adaptive dynamics [110]. The recent work

in [67] and its references provide further insights into these developments. While these results have significantly advanced the design of NSS dynamics, existing approaches still suffer from fundamental transient limitations inherited from PSG flows. These limitations are particularly pronounced in games where the pseudogradient has shallow monotonicity properties.

To address these challenges, we focus on providing stability and robustness certificates for momentum-based dynamics in the context of non-potential games. In particular, we examine dynamics of the form:

$$\dot{q} = \frac{2}{\tau}(p - q), \quad \dot{p} = -2\tau\mathcal{G}(q), \quad \dot{\tau} = \eta, \tag{5.1}$$

where $\tau(0) = T_0 \geq 0$ and $\eta > 0$. As demonstrated in Chapter 2, these dynamics are related to Nesterov's ODE describing the evolution of a state x, when $\mathcal{G}$ is a gradient operator, and via the transformation $q = x,\ p = \frac{\tau}{2}\dot{x} + x$.

As we have illustrated in Chapter 3 and 4 , using this type of dynamics with properly implemented distributed and coordinated resets offers a promising alternative to traditional approaches in potential games. Building upon these insights, we now seek to extend this framework to non-potential games. This approach has the potential to yield improved convergence rates compared to pseudogradient (PSG) flows, while retaining the flexibility to incorporate additional mechanisms crucial for real-world applications.

However, extending these benefits to non-potential games presents new challenges and raises two critical questions:

1) Is it possible to adapt systems of the form (5.1) to provide robust and efficient solutions for NSS problems in non-potential games when $\mathcal{G}$ is the pseudogradient map of a non-potential game?

2) Can these dynamics be extended effectively to network games and model-free settings

without sacrificing stability or convergence guarantees?

In this section, we provide answers to the above questions by using the hybrid dynamical system $\mathcal{H}_1$ formulated in Chapter 4 (Section 4.2). In particular, the following original contributions constitute the main results of this section:

i) We establish *robust asymptotic stability* properties for the hybrid dynamical system $\mathcal{H}_1$ in the context of non-cooperative and *nonpotential* games with N players. This analysis leads to the first Nash Set Seeking (NSS) algorithms with continuous-time *dynamic* momentum and *robust* asymptotic stability guarantees in such games. The structure of the resulting algorithms incorporates three key elements:

a) A class of distributed *continuous-time* pseudogradient-based dynamics with *time-varying* momentum coefficients, inspired by (5.1), which forms the flow map of $\mathcal{H}_1$;

b) Distributed periodic *discrete-time* resets implemented by the players, which are captured in the jump map of $\mathcal{H}_1$ and incorporate *heterogeneous* reset policies allowing players to decide whether or not to restart their own momentum;

c) A robust *set-valued* distributed coordination mechanism that synchronizes the reset times of the players to induce suitable system-wide acceleration properties, represented in the flow and jump sets of $\mathcal{H}_1$.

ii) To accommodate situations where players do not have access to full-information Oracles that provide evaluations of their pseudogradients, we introduce a new *distributed* momentum-based hybrid NSS algorithm for games with partial information, where players leverage communication with neighbors to estimate their actions on-the-fly in order to obtain gradient evaluations from the Oracle. The design of these dynamics follows similar multi-time scale ideas used for ODEs in the literature [107], but which are not directly applicable to systems of the form (5.1). Indeed, unlike existing results based on fast

consensus dynamics and "reduced" pseudogradient flows, our reduced dynamics are hybrid and set-valued, which prevents the direct application of standard singular perturbation tools for ODEs.

iii) We present payoff-based versions of all our hybrid NSS algorithms, suitable for *model-free* learning in non-cooperative games where players have access only to measurements of their cost. These dynamics exploit recent tools developed in the context of averaging-based hybrid extremum seeking control [111, 25], and their analyses is fundamentally different from other model-free non-hybrid algorithms studied in the literature, e.g. [35, 112, 113]. In particular, the dynamics considered in this chapter have set-valued jump maps that lead to non-unique solutions with non-trivial hybrid time domains having multiple simultaneous jumps in the standard continuous time domain, a behavior that is unavoidable in decentralized multi-agent HDS. We also show that these adaptive dynamics can approximately recover the acceleration properties of the model-based algorithms.

To our best knowledge, the algorithms of this section (model-based, partial-information, and model-free) are the first in the literature that implement dynamic momentum and distributed restarting techniques in N-player noncooperative games.

Problem Formulation

In this chapter, we consider noncooperative games with $N \in \mathbb{Z}_{\geq 2}$ players, where each player i can control its own action q_i, and has access to the actions q_j of neighboring players $j \in \mathcal{N}_i \coloneqq \{j \in \mathcal{V} : (i,j) \in \mathcal{E}\}$, who are characterized by an undirected, connected, and time-invariant graph $\mathbb{G} = \{\mathcal{E}, \mathcal{V}\}$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ is the set of players and $\mathcal{E}$ is the set of edges between players. We use $\mathcal{L}$ to denote the Laplacian matrix of the graph $\mathbb{G}$. The main goal of each player i is to minimize its own cost function $\phi_i : \mathbb{R}^n \to \mathbb{R}$ by controlling its own action q_i. We assume that the costs ϕ_i are *twice continuously differentiable*, and we use $q = (q_1, q_2, \ldots, q_N)$ to denote the overall vector of actions of the game. We also use denote as q_{-i} the vector of all actions with the action of player i

removed. We use $\mathcal{G}$ to denote the pseudogradient of the game, where

$$q \mapsto \mathcal{G}(q) \coloneqq \left(\frac{\partial \phi_1(q)}{\partial q_1}, \frac{\partial \phi_2(q)}{\partial q_2}, \ldots, \frac{\partial \phi_N(q)}{\partial q_N} \right) \in \mathbb{R}^{nN}. \tag{5.2}$$

Following standard assumptions in the literature of fast NES [58, 60] and accelerated optimization [19, 101, 114, 115, 116, 25, 4], in this chapter we will work with the following assumptions.

Assumption 5.1. The mapping $\mathcal{G}$ is ℓ-globally Lipschitz, i.e., there exists a constant $\ell > 0$ such that $|\mathcal{G}(q) - \mathcal{G}(q')| \leq \ell |q - q'|$, for all $q, q' \in \mathbb{R}^{nN}$.

Assumption 5.2. The mapping $\mathcal{G}$ is $1/\ell$-cocoercive, i.e., there exists ℓ such that $\big(\mathcal{G}(q) - \mathcal{G}(q')\big)^\top (q - q') \geq \frac{1}{\ell} |\mathcal{G}(q) - \mathcal{G}(q')|^2$ for all $q, q' \in \mathbb{R}^{nN}$. Moreover, the map $q \mapsto |\mathcal{G}(q)|^2$ is radially unbounded.

The first property of Assumption 5.2 implies Assumption 5.1 via direct application of the Cauchy–Schwarz inequality. However, the converse is not necessarily true in non-potential games [117]. We will also use the following definition to characterize the monotonicity properties of the games.

Definition 5.1. A game with pseudogradient $\mathcal{G}$ is said to be:

1. *monotone* if it satisfies $\big(\mathcal{G}(q) - \mathcal{G}(q')\big)^\top (q - q') \geq 0$, for all $q,\ q' \in \mathbb{R}^{nN}$.
2. *strictly monotone* if it satisfies $\big(\mathcal{G}(q) - \mathcal{G}(q')\big)^\top (q - q') > 0$, for all $q \neq q' \in \mathbb{R}^{nN}$.
3. κ-*strongly monotone* with $\kappa > 0$, if it satisfies $\big(\mathcal{G}(q) - \mathcal{G}(q')\big)^\top (q - q') \geq \kappa |q - q'|^2$, for all $q, q' \in \mathbb{R}^{nN}$.
4. κ-*strongly monotone quadratic* if it is a κ-strongly monotone game with $\mathcal{G}(q) = Aq + b$ for some $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.

Monotone, strictly and κ-strongly monotone games that are also potential games will be referred to as *monotone potential games*, *strongly monotone potential games* and *κ-strongly monotone potential games*, respectively. Their monotonicity properties are defined by $\mathcal{G}$.

Remark 5.1. Cocoercive mappings are monotone but not necessarily strongly monotone. However, games that are κ-strongly monotone and ℓ-Lipschitz are also κ/ℓ^2-cocoercive [117, Prop. 2.1].

Strict monotonicity of $\mathcal{G}$ implies that there is exactly one Nash equilibirium, if it exists. For κ-strongly monotone games and monotone potential games, existence is always guaranteed [118, Thm. 2.3.3]. In some cases, we will also work with the following assumption.

Assumption 5.3. The function $\phi_i : \mathbb{R} \to \mathbb{R}$ is radially unbounded in q_i for ever $q_{-i} \in \mathbb{R}^{n(N-1)}$ and all $i \in \mathcal{V}$.

Building upon the regularity properties established for the pseudogradient $\mathcal{G}$ and the individual cost functions ϕ_i of the game, we now turn our attention to analyzing the stability properties of the *hybrid momentum-based Nash-Set-Seeking* (HM-NSS) dynamics $\mathcal{H}_1$, introduced in Chapter 4 (Section 4.2). To facilitate our analysis, we will restate the definition of $\mathcal{H}_1$, highlighting its structure and key components. The system $\mathcal{H}_1$ is characterized by an overall state $x = (p, q, \tau) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N$ and is described by the following hybrid dynamic inclusion:

$$\mathcal{H}_1 : \begin{cases} x \in C_1 & \dot{x} = F_1(x) = \begin{pmatrix} 2\mathcal{T}^{-1}(p-q) \\ -2k\mathcal{T}\mathcal{G}(q) \\ \eta \mathbf{1}_N \end{pmatrix}, \\ x \in D_1 & x^+ \in G_1(x) \end{cases} \tag{5.3a}$$

where $0 < \eta \leq 1/2$, $\mathcal{T} \coloneqq \operatorname{diag}(\tau \otimes \mathbf{1}_n)$. The flow set C_1 is defined as:

$$C_1 \coloneqq \Big\{ x = (q, p, \tau) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N \ : \ \tau \in [T_0, T]^N \Big\}, \tag{5.3b}$$

and the jump set is

$$D_1 \coloneqq \Big\{ x = (q, p, \tau) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N \ : \ x \in C_1, \ \max_{i \in \mathcal{V}} \tau_i = T \Big\}. \tag{5.3c}$$

To ensure the robustness and well-posedness of the hybrid system, we construct the jump map $G_1(x)$ by taking the outer-semicontinuous hull of an intermediate set-valued mapping G^0. We derive this mapping $G^0 : \mathbb{R}^{2nN+N} \rightrightarrows \mathbb{R}^{2nN+N}$ from the individual resetting map $\mathcal{R}_i$ and the coordination mechanism $\mathcal{C}_j$, and design it to be non-empty only under specific conditions, namely when $\tau_i = T$ and $\tau_j \in [T_0, T)$ with $j \neq i$, for each $i \in \mathcal{V}$. When these conditions are met, G^0 is defined as follows:

$$\begin{aligned} G^0(x) \coloneqq \Big\{ & (v_1, v_2, v_3) \in \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \mathbb{R}^N : (v_{1,i}, v_{2,i}, v_{3,i}) = \mathcal{R}_i(x_i), \\ & v_{1,j} = q_j, v_{2,j} = p_j, v_{3,j} \in \mathcal{C}_j(\tau_j), \ \forall \ j \in \mathcal{N}_i, \\ & v_j = x_j, \forall \ j \notin \mathcal{N}_i \Big\}, \end{aligned} \tag{5.3d}$$

where $x \coloneqq (x_1, x_2, \cdots, x_N)$, $(q_i, p_i, \tau_i) = x_i \in \mathbb{R}^{2n+1}$, and $\mathcal{R}_i$ and $\mathcal{C}_j$ are defined as:

$$\mathcal{R}_i(x_i) = \begin{pmatrix} q_i \\ \alpha_i p_i + (1-\alpha_i) q_i \\ T_0 \end{pmatrix}, \tag{5.3e}$$

and

$$\mathcal{C}_j(\tau_j) \coloneqq \begin{cases} T & \text{if } \tau_j \in (T_0 + r_j, T] \\ \{T_0, T\} & \text{if } \tau_j = T_0 + r_j \\ T_0 & \text{if } \tau_j \in [T_0, T_0 + r_j) \end{cases}, \tag{5.3f}$$

with $\alpha_i \in \{0, 1\}$ for all $i \in \mathcal{V}$, and $r_j \in (0, \frac{T - T_0}{N})$ for all $j \in \mathcal{V}$.

The qualitative behavior of system $\mathcal{H}_1$ will depend on the choice of parameters (η, T, T_0), which characterize the frequency and the minimum and maximum values of the momentum coefficient τ. Different choices of (η, T, T_0) will lead to different *reset conditions* (RCs). These RCs will be defined in terms of the following *condition numbers* of the game, the reset mechanism, and the graph, respectively:

$$\sigma_\phi \coloneqq \frac{\ell}{\kappa}, \quad \sigma_r \coloneqq \frac{T}{T_0}, \quad \sigma_{\mathcal{L}} = \frac{\lambda_{\max}(\mathcal{L})}{\lambda_2(\mathcal{L})}, \tag{5.4}$$

where ℓ is given by Assumptions 5.1 or 5.2, κ is given in Definition 5.1, and $\lambda_2(\mathcal{L}), \lambda_{\max}(\mathcal{L})$ are the smallest positive and the largest eigenvalues, respectively, of the Laplacian $\mathcal{L}$.

5.1.1 Nash Equilibrium Seeking in Non-Potential Games

We study the stability and convergence properties of the dynamics $\mathcal{H}_1$ with respect to the compact set

$$\mathcal{A} \coloneqq \mathcal{A}_{pq} \times \mathcal{A}_{\text{sync}}, \tag{5.5a}$$

where

$$\mathcal{A}_{\text{sync}} \coloneqq \left(\mathbf{1}_N[T_0, T]^N\right) \cup \{T_0, T\}^N \tag{5.5b}$$

$$\mathcal{A}_{qp} \coloneqq \{(q,p) \in \mathbb{R}^{2n} : p = q,\ q \in \mathcal{A}_{\text{NE}}\}, \tag{5.5c}$$

and $\mathcal{A}_{\text{NE}}$ denotes the set of points $q^* \in \mathbb{R}^{nN}$ satisfying

$$\phi_i(q_i^*, q_{-i}^*) = \inf_{q_i \in \mathbb{R}^n} \phi_i(q_i, q_{-i}^*), \quad \forall\ i \in \mathcal{V} = \{1, 2, \cdots, N\}\,, \tag{5.5d}$$

i.e., the set of Nash equilibria of the game.

> **Remark 5.2.** Given a monotone game with pseudogradient $\mathcal{G}$, it follows that q^* is a NE if and only if $\mathcal{G}(q^*) = 0$ [66, Prop. 2.1]. Then, we can formally define $\mathcal{A}_{\text{NE}}$ as follows:
>
> $$\mathcal{A}_{\text{NE}} \coloneqq \{q \in \mathbb{R}^{nN} \ :\ \mathcal{G}(q^*) = 0\}. \tag{5.5e}$$

The first RC that we consider is given by

$$T^2 - T_0^2 > \frac{\rho_J}{2} \cdot (1 - \underline{\alpha})\,, \tag{RC$_1$}$$

where $\rho_J \in \mathbb{R}_{\geq 0}$ is a parameter to be determined and $\underline{\alpha} = \min_{i \in \mathcal{V}} \alpha_i$. This condition will regulate how frequently players reset their states. Finally, we also introduce the constant

$$\gamma(\rho_J) \coloneqq \left(1 - \frac{1}{\sigma_r^2} - \frac{\rho_J}{2T^2}\right). \tag{5.6}$$

where σ_r is defined in (5.4). This quantity will be instrumental to characterize the rates of convergence towards $\mathcal{A}$.

In contrast to the analysis presented in Chapter 4 (4.3), when a potential function does not exist the analysis of the HDS $\mathcal{H}_1$ is more challenging. In this case, we now

introduce the following state-dependent matrix parameterized by $(\rho_F, \delta) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0}$:

$$\mathcal{M}_\delta(q, \rho_F) \coloneqq I_N - \mathcal{S}_\delta\left(q, \rho_F\right) \mathcal{S}_\delta\left(q, \rho_F\right)^\top, \tag{5.7}$$

with $\mathcal{S}_\delta : \mathbb{R}^{nN} \times \mathbb{R}_{>0} \to \mathbb{R}^{nN \times nN}$ given by the scaled matrix

$$\mathcal{S}_\delta(q, \rho_F) \coloneqq \chi(\rho_F, \delta)^{\frac{1}{2}} \Big(\rho_F I_{nN} - \partial\mathcal{G}(q) \Big),$$

where $\partial\mathcal{G}$ is the Jacobian of $\mathcal{G}$, i.e. $\frac{\partial \mathcal{G}}{\partial q}$, and where the mapping $\chi : \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is given by

$$\chi(\rho_F, \delta) = \frac{T^2}{1 - \delta T^2} \cdot \frac{1}{\rho_F(1 - \eta) - \delta \rho_F^2},$$

which is defined for all arguments such that $\delta^2 T^2 < 1$ and $1 - \eta > \delta \rho_F$. We use the following definition to extend [119, Def. 4.1.2] to matrices of the form (5.7).

Definition 5.2. The matrix-valued function $q \mapsto \mathcal{S}_\delta(q, \rho_F)$ is said to be ρ_F-*Globally Contractive* $\big(\rho_F\text{-GC}\big)$ if $\mathcal{M}_\delta(q, \rho_F) \succ 0$ for all $q \notin \mathcal{A}_{\text{NE}}$.

Note that when $\mathcal{M}_\delta \succ 0$, the coefficient χ characterizes the *level of contraction* of $\mathcal{S}_\delta$. Indeed, $\mathcal{M}_\delta \succ 0$ if and only if

$$\frac{1}{\chi(\rho_F, \delta)} \geq \sigma_{\max}\Big(\rho_F I_n - \partial\mathcal{G}(q) \Big)^2, \tag{5.8}$$

where $\sigma_{\max}(\cdot)$ is the maximum singular value of its argument [120, Thm. 7.7.2]. Using the definition of χ, and inequality (5.8), it can be observed that in order to ensure that $\mathcal{S}_\delta$ is ρ_F-GC for some pair (δ, ρ_F), the resetting parameter T cannot be chosen arbitrarily large. Example 4 illustrates this point.

Example 5.1. Consider a κ-strongly monotone quadratic game with $\kappa = 6$, and

$$\mathcal{G}(q) = \begin{pmatrix} 6 & 1.5 \\ -1.5 & 6 \end{pmatrix} (q - q^*), \tag{5.9}$$

where $q^* = (2, -2)$. First, let $\delta = 0$, and note that for this game $\mathcal{M}_0(q, \rho_F) = \mathcal{D}\left(m_0(\rho_F)\mathbf{1}_2\right) \in \mathbb{R}^{2\times 2}$, where

$$m_0(\rho_F) = 1 - T^2 \frac{4(\rho_F - 12)\rho_F + 153}{4(1-\eta)\rho_F}.$$

Notice that $4(\rho_F - 12)\rho_F + 153 > 0$ for all $\rho_F \in \mathbb{R}_{>0}$, and recall that $\eta \leq \frac{1}{2}$ by assumption. Thus, for all $\rho_F > 0$ there exists $\overline{T} \in \mathbb{R}_{>0}$ such that $\mathcal{M}_0(q, \rho_F) \succ 0$ for all $T \in (0, \overline{T})$, and $\mathcal{M}_0(q, \rho_F) \preceq 0$ for $T \geq \overline{T}$. Similarly, when $\delta > 0$ we have that if $\mathcal{S}_\delta$ is ρ_F-GC, then $\mathcal{S}_0$ is also ρ_F-GC. Thus, we can conclude that for every ρ_F and $\delta \geq 0$ there exists $\overline{T}$ such that $\mathcal{S}_\delta$ is not ρ_F-GC for any $T \geq \overline{T}$.

Using the global contractivity property of Definition 5.2, we have the following result for non-potential games.

Theorem 5.1. Let $\mathcal{G}$ describe a strictly monotone game, and suppose that Assumptions 5.2 and 5.3 hold. Consider the HDS $\mathcal{H}_1$ under (RC_1) with $\rho_J \geq 0$ and with reset policy $\alpha \coloneqq (\alpha_1, \alpha_2, \ldots, \alpha_N) = \mathbf{1}_n$. If $\mathcal{S}_0$ is ℓ-GC then the set $\mathcal{A}$ is UGAS, for every $i \in \mathcal{V}$, and for all solutions x the following bound holds during flows

$$|\mathcal{G}(q(t,j))|^2 \leq \frac{\tilde{c}_j}{\tau_i^2(t,j)}, \tag{5.10}$$

for all $(t, j) \in \mathrm{dom}\,(x)$ such that

$$t + j \geq (T - T_0)/\eta + N =: T^*, \tag{5.11}$$

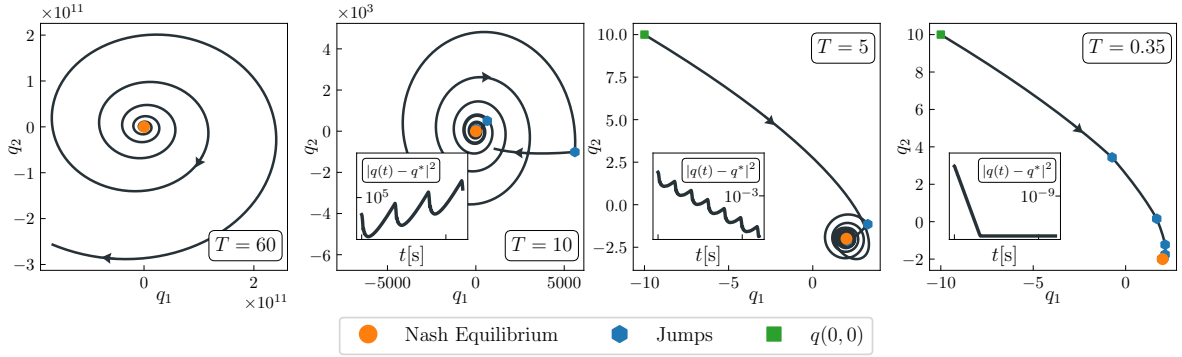


Figure 5.1. Phase plane plots showing the trajectories of the actions resulting from the HM-NSS dynamics in a non-potential 2-player *κ-strongly monotone quadratic game* with $\kappa = 6,\ \ell = 6.2$ and $\tau(0,0) = 0.1 \cdot \mathbf{1}_2$.

> and where $\{\tilde{c}_j\} \to 0^+$ is a sequence of positive and monotonically decreasing numbers parameterized by $x(0,0)$.

Unlike the results of Theorem 4.1, in non-potential games the ρ_F-global-contractivity of $\mathcal{S}_\delta$ plays a fundamental role in the stability analysis of $\mathcal{H}_1$. In particular, the ℓ-GC property of $\mathcal{S}_\delta$ will guarantee a suitable dissipativity property during flows via Lyapunov-based tools. While in Theorem 2 this is only a sufficient condition, the plots of Figure 5.1 indicate that keeping T "sufficiently small" is also a necessary condition to preserve stability in non-potential games. In this figure, we show the phase plane of solutions to $\mathcal{H}_1$ with different values of T, in a game $\mathcal{G}$ given by (5.9). It can be observed in the left plots that if T is not small, divergent trajectories emerge. As shown in the right plots, the instability is removed by implementing sufficiently frequent resets.

Next, we provide a sufficient reset condition that guarantees that $\mathcal{S}_0$ is ℓ-GC in any cocoercive strictly monotone game.

> **Lemma 5.1.** Suppose that Assumption 5.2 holds, and $\mathcal{G}$ describes a strictly monotone game. Let (η, T, ℓ) satisfy the scalar inequality:
>
> $$0 < T^2 < \frac{1-\eta}{2\ell}. \qquad (\mathrm{RC}_2)$$

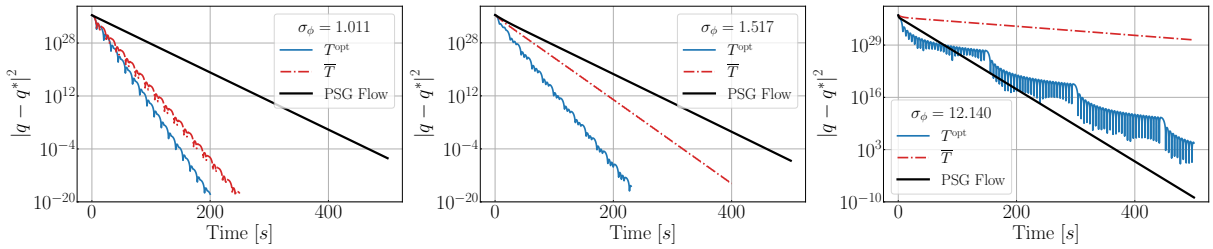


Figure 5.2. Distance to the Nash Equilibrium for trajectories resulting from the HM-NSS dynamics in 5-player κ-strongly monotone quadratic games with $\kappa = 0.099$, and different condition numbers σ_ϕ.

Then, $\mathcal{S}_0$ is ℓ-GC.

We now turn our attention to games that are κ-strongly monotone and ℓ-Lipschitz. For these games, we ask that the contractivity properties of $\mathcal{S}_\delta$ hold with $\delta > 0$, and that (RC_1) holds with a particular value of ρ_J. We recall that the condition numbers (σ_ϕ, σ_r) are defined in (5.4).

Theorem 5.2. Suppose that Assumption 5.1 holds and let $\mathcal{G}$ describe a κ-strongly monotone game. Consider the HDS $\mathcal{H}_1$ under (RC_1), and suppose that $\mathcal{S}_\delta$ is $(\sigma_\phi \ell)$-GC with $0 < \delta < (1 - \eta)/(\sigma_\phi \ell)$. Then, the following holds:

(i_4) If $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \{0, 1\}^n$ and $\rho_J = 0$, then $\mathcal{A}$ is UGES, and there exists $\lambda > 0$ such that for each compact set $K \subset C_1 \cup D_1$ there exists $M_0 > 0$ such that for all solutions x, with $x(0, 0) \in K_0$, the bound (4.18) holds.

(i_5) If $\alpha = \mathbf{0}_n$ and $\rho_J = \sigma_\phi^2 \kappa^{-1}$, then $\mathcal{A}$ is UGES and for each compact set $K_0 \subset C_1 \cup D_1$ there exists $M_0 > 0$ such that for all solutions x, with $x(0, 0) \in K_0$, and for all $(t, j) \in \text{dom}\,(x)$, the following bound holds:

$$|q(t, j) - q^*| \leq \sigma_r \sigma_\phi \left(1 - \gamma\left(\rho_J\right)\right)^{\frac{\alpha(j)}{2}} M_0,$$

where $\alpha(j) := \max\{0, \lfloor \frac{j-N}{N} \rfloor\}$, and $\gamma\left(\rho_J\right) \in (0, 1)$.

Before commenting on the implications of Theorem 5.2, we present a reset condition for κ-strongly monotone games that is analogous to the one of Lemma 5.1.

Lemma 5.2. Suppose that Assumption 5.1 holds and that $\mathcal{G}$ describes a κ-strongly monotone game. Let $(\eta, T, \sigma_\phi \ell)$ satisfy the scalar inequality:

$$0 < T^2 < \frac{1 - \eta - \delta\sigma_\phi \ell}{\sigma_\phi \ell - \kappa + \delta(1 - \eta - \delta\sigma_\phi \ell)}, \tag{RC$_3$}$$

with $0 \leq \delta < (1 - \eta)/(\sigma_\phi \ell)$. Then $\mathcal{S}_\delta$ is $(\sigma_\phi \ell)$-GC.

Similar to (RC$_2$), the reset condition (RC$_3$) imposes an upper bound on the lengths of the intervals of flow of the HDS $\mathcal{H}_1$, now modulated by the condition number σ_ϕ and the Lipschtiz constant ℓ.

Remark 5.3. When $\rho_J = \sigma_\phi^2 \kappa^{-1}$, the conjunction of (RC$_1$) and (RC$_3$) imposes upper *and* lower bounds for the reset times of the HDS $\mathcal{H}_1$ for all times $(t, j) \in \mathrm{dom}\,(x)$ such that $t + j \geq (T - T_0)/\eta + N$. This result is in contrast to potential games (and standard convex optimization problems) with periodic restarting where only a lower bound between resets is usually needed to achieve exponential convergence [31, 25]. Instead, Theorem 5.2 suggests that resets must occur in a particular frequency band: they should not occur too frequently (i.e., T should not be too small) such that (RC$_1$) holds and the distance $|q - q^*|$ shrinks by a constant quantity after each interval of flow; however, resets should also happen frequently enough (i.e., T should not be too large) such that $\mathcal{S}_\delta$ remains $(\sigma_\phi \ell)$-GC. Whenever a resetting time T is in such frequency band, i.e., whenever it satisfies (RC$_1$) and (RC$_3$), or (RC$_1$) and ρ_F-global contractivity of $\mathcal{S}_\delta$, we say that the T is *feasible*.

The next lemma provides a sufficient condition to guarantee feasibility of the reset

conditions of Theorem 5.2.

Lemma 5.3. For any $\kappa > 0$, $\eta \leq 1/2$ and σ_ϕ such that $\sigma_\phi^4 - \sigma_\phi^2 < 2(1-\eta)$, there exists (T, T_0) such that (RC_1) and (RC_3) simultaneously hold with $\rho_J = \sigma_\phi^2 \kappa^{-1}$, provided that δ is sufficiently small.

In Theorem 5.2, the restarting policy $\alpha = \mathbf{0}_n$ leads to exponential NSS with rate of convergence characterized by $(1 - \gamma(\sigma_\phi^2/\kappa))$. For this coefficient, one can choose a "quasi-optimal" restarting parameter T to induce an acceleration-like property in κ-strongly monotone games:

Lemma 5.4. Under the Assumptions of Theorem 5.2-(i_5), and for any $\nu > 0$, the choice $T = T^{\mathrm{opt}} := e\sigma_\phi \sqrt{\frac{1}{2\kappa} + \frac{T_0^2}{\sigma_\phi^2}}$ guarantees that $|q(t,j) - q^*| \leq \nu$ for all $t \geq t_{\nu^{\mathrm{opt}}}$, where

$$t_\nu^{\mathrm{opt}} = \frac{1}{\eta}\left(e\sigma_\phi \sqrt{\frac{1}{2\kappa} + \frac{T_0^2}{\sigma_\phi^2}} - T_0\right) \ln\left(\frac{\sigma_\phi \sigma_r M_0}{\nu}\right),$$

and M_0 is a constant that depends on $|q(0,0) - q^*|$. Moreover, the convergence is of order $\mathcal{O}(e^{-\sqrt{\kappa}/\sigma_\phi})$ as $T_0 \to 0^+$.

Remark 5.4. The result of Lemma 5.4 showcases the exponential bound induced by the HM-NSS dynamics: as $\sigma_\phi \to 1$, the convergence is of order $\mathcal{O}(e^{-\sqrt{\kappa}t})$, which is advantageous in games with low curvature and moderate condition number, see Figure 2.3. However, as σ_ϕ increases, the theoretical convergence bound deteriorates. In Figure 5.2, we confirm this fact by implementing the HM-NSS in different κ-strongly non-potential monotone quadratic games with low curvature ($\kappa \ll 1$) and different condition numbers $\sigma_\phi \in \{1.011,\ 1.517,\ 12.140\}$. For these games, we compare two resetting times: 1) $T = T^{\mathrm{opt}}$, which is only feasible for $\sigma_\ell = 1.011$, and 2) $T = \overline{T}$,

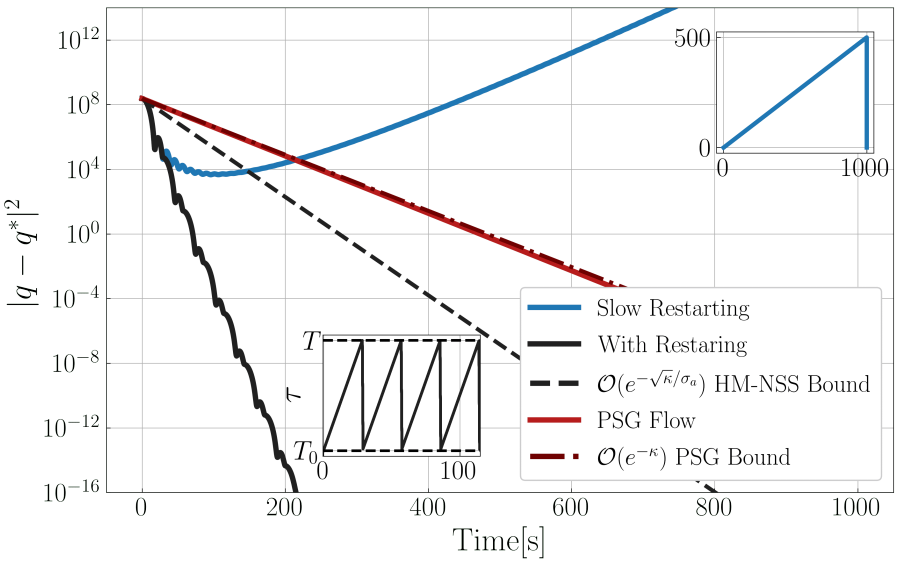


Figure 5.3. Convergence of trajectories of (2.1) in a the non-potential game of Example 2.3 with $\kappa = 0.02,\ \ell = 0.0214,\ n = 30,\ T_0 = 0.1,\ T = 3.74$. The black line shows the trajectory of a hybrid mechanisms implementing coordinated resets with $T = T^{\text{opt}}$

> where $\overline{T}$ is the biggest resetting time found to guarantee that $\mathcal{S}_\delta$ is $(\sigma_\phi \ell)$-GC, with $\delta > 0$ approaching to 0, and is feasible for all σ_ℓ. Experimentally, we observe that performance deteriorates slower when using $\overline{T}$ instead of T^{opt}, and acceleration can be achieved for a wider range of values of σ_ℓ. However, when $T = T^{\text{opt}}$ is not feasible we cannot use the results of Theorem 5.2 to guarantee stability certificates under the implementation of the HM-NSS dynamics. Whether or not a small condition number is a *necessary* condition to achieve acceleration in games remains an open question.

It is possible to find additional conditions on the game and the parameters of $\mathcal{H}_1$ to guarantee that T^{opt} satisfies (RC_1) and (RC_3), i.e., that is *feasible*. However, such conditions are rather involved and unintuitive. Yet, as we show in Figure 5.3, where we revisit the nonpotential game of Example 2.3, the quasi-optimal restarting T^{opt} can be verified to be feasible, both enabling the recovery of stability as well as ensuring suitable transient performance. We also note that, based on numerical experiments, our theoretical bound is conservative.

Remark 5.5. The stability results of Theorem 3-(i_3) are obtained by guaranteeing strong-decrease of a suitably chosen Lyapunov function $\tilde{V}$ during flows and jumps of the HM-NSS dynamics. In Figure 5.4 we show the value of $\tilde{V}$ for different trajectories resulting from the HM-NSS dynamics in the same κ-strongly monotone games studied in Remark 5.4, and using the resetting time T^{opt} of Lemma 5.4. As seen in the figure, the Lyapunov function does not exhibit strong decrease during flows $\Big($with sign $\big(\dot{\tilde{V}}\big)$ changing multiple times$\Big)$ for the non-feasible values of T^{opt}, which correspond to the condition numbers $\sigma_\phi = 1.157$ and $\sigma_\phi = 12.140$. However, the function does experience an overall decrease during the simulation time due to the strong decrease enforced through the jumps of the hybrid dynamics. Experimentally, we observe that to obtain better transient performance (with Ts that are outside the feasible set), one could allow the Lyapunov function to increase during flows, provided this increase is compensated via jumps. This indicates interesting opportunities to attain even better behavior via adaptive resetting policies that reset whenever the Lyapunov function increases. However, the development of those techniques in a distributed way is challenging and falls out of the scope of the current paper.

We finish this section with a result for quadratic games.

Proposition 5.1. Suppose that Assumption 5.1 holds and $\mathcal{G}$ describes a κ-strongly monotone quadratic game. Consider the hybrid dynamics $\mathcal{H}_1$ under the reset condition (RC_1) with $\rho_J = \kappa^{-1}$ and reset policy $\alpha = \mathbf{0}_n$. If $\mathcal{S}_\delta$ is κ-GC, then $\mathcal{A}$ is UGES, and for each compact set $K_0 \subset C_1 \cup D_1$ there exists $M_0 > 0$ such that all solutions x with $x(0,0) \in K_0$ satisfy

$$|q(t,j) - q^*| \leq \sigma_r \sqrt{\sigma_\phi} \left(1 - \gamma\left(\kappa^{-1}\right)\right)^{\frac{\alpha(j)}{2}} M_0,$$

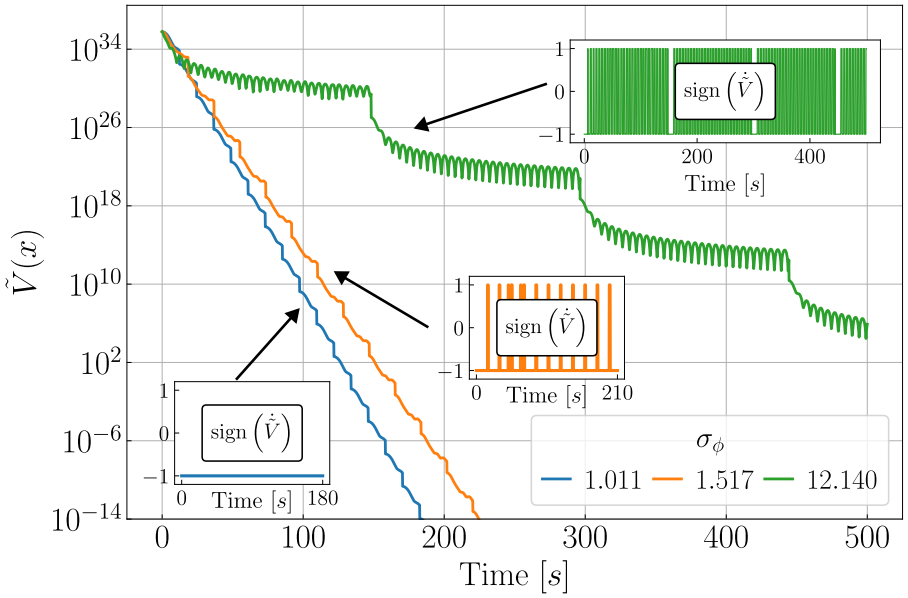


Figure 5.4. Lyapunov function $\tilde{V}(x)$ for trajectories resulting from the HM-NSS in 5-player κ-strongly monotone quadratic games with $\kappa = 0.009$, and different condition numbers σ_ϕ. In all the cases the dynamics are implemented with the resetting period $T = T^{\text{op}}$ presented in Lemma 5.4.

where $\alpha(j) \coloneqq \max\{0, \lfloor \frac{j-N}{N} \rfloor\}$, and $\gamma\left(\kappa^{-1}\right) \in (0,1)$.

It is not difficult to show that the assumptions of Proposition 5.1 can be satisfied when T_0 is sufficiently small and $\sigma_\phi < \sqrt{2}$. However, it is also not difficult to find games for which this condition is violated, yet (RC_1) holds and $\mathcal{S}_\delta$ is still κ-GC. In fact, for quadratic games, condition (5.7) simplifies to $0 \prec I_n - k_1(\kappa I_n - A)(\kappa I_n - A)^\top$, with $k_1 = \frac{T^2}{\kappa(1-\eta)}$.

Remark 5.6. For some games, a potential function P might be available but its gradient ∇P might only be monotone if a suitable vector of weights $\omega \in \mathbb{R}^n$ is chosen so that $\text{diag}(\omega)\nabla P$ is strictly or κ-strongly monotone. In such cases, the results of Theorem 4.1 are not applicable, but those of Theorems 5.1 and 5.2 hold provided that $\mathcal{G} \coloneqq \text{diag}(\omega)\nabla P$ satisfies the suitable assumptions on its continuity and contractivity of S_δ. Such weighted potential games have been recently studied in [121] in the context of congestion games.

5.1.2 Hybrid Momentum-Based NSS with Partial Information

In the previous section, we assumed that each player i had access to an individual Oracle able to generate measurements of the partial gradient $\dfrac{\partial \phi_i(\cdot)}{\partial q_i}$ at the *overall* state q of the game. In this section, we now relax this assumption by considering Oracles that provide *evaluations* of the partial derivatives. In this case, to carry out the gradient evaluations each player needs to estimate the actions of the other players by extending the hybrid dynamics $\mathcal{H}_1$. To simplify our exposition, in the following sections we assume that the actions q_i are scalars, i.e., that $n = 1$. However, all our results also hold for vectorial actions by using suitable Kronecker products.

Individual Multi-Time Scale Hybrid Dynamics

To achieve distributed NSS over graphs with partial information, we proceed to endow each player $i \in \mathcal{V}$ with an auxiliary state $\mathbf{e}^i$ that serves as an individual estimation of the actions of the other players:

$$\mathbf{e}^i \coloneqq (\mathbf{e}^i_1, \mathbf{e}^i_2, \ldots, q_i, \ldots, \mathbf{e}^i_{n-1}, \mathbf{e}^i_N) \in \mathbb{R}^N. \tag{5.12}$$

Since players do not need to estimate their own action, it is also convenient to introduce the auxiliary state $\mathbf{e}^i_{-i} \in \mathbb{R}^{N-1}$ which contains the same entries of $\mathbf{e}^i$ with the exception of q_i, which is removed. Using this notation, we now assume that each player i has access to individual *gradient Oracles* characterized by mappings of the form $(q_i, \mathbf{e}^i_{-i}) \mapsto \hat{\mathcal{G}}_i(q_i, \mathbf{e}^i_{-i})$, which satisfy $\hat{\mathcal{G}}_i(q_i, q_{-i}) = \frac{\partial \phi_i(q)}{\partial q_i}$. Following similar notation used in the literature [107], we define the matrices

$$\mathcal{Q}_i \coloneqq \begin{bmatrix} I_{(i-1)} & 0_{(i-1)\times 1} & 0_{(i-1)\times(N-i)} \\ 0_{(N-i)\times(i-1)} & 0_{(N-i)\times 1} & I_{(N-i)} \end{bmatrix},$$

$$\mathcal{P}_i \coloneqq \begin{bmatrix} 0_{1\times(i-1)} & 1 & 0_{1\times(N-i)} \end{bmatrix}.$$

By using these definitions, each player i now implements the following accelerated augmented continuous-time NSS dynamics:

$$\begin{pmatrix} \dot{q}_i \\ \dot{p}_i \\ \dot{\tau}_i \\ \dot{\mathbf{e}}^i_{-i} \end{pmatrix} = \begin{pmatrix} \frac{2}{\tau_i}(p_i - q_i) + \mathcal{P}_i \sum_{j\in\mathcal{N}_i}(\mathbf{e}^i - \mathbf{e}^j) \\ -2\tau_i \hat{\mathcal{G}}_i(q_i, \mathbf{e}^i_{-i}) \\ \eta \\ -\frac{1}{\varepsilon}\mathcal{Q}_i \sum_{j\in\mathcal{N}_i}(\mathbf{e}^i - \mathbf{e}^j) \end{pmatrix}, \tag{5.13}$$

where $\varepsilon > 0$ is a new tunable parameter. These dynamics are implemented whenever the state τ_i satisfies $\tau_i \in [T_0, T)$. The momentum-based dynamics (5.13) implement a dynamic consensus mechanism with state $\mathbf{e}^i_{-i}$. This mechanism uses a high gain $\frac{1}{\varepsilon}$ to induce a time-scale separation in the flows of the hybrid algorithm. In particular, if the states $\mathbf{e}^i$ were to instantaneously achieve their steady state value, the flows (5.13) would reduced to the flows (4.9c).

When players are uncoordinated, the individual resets are triggered by the condition $\tau_i = T$, and are given by

$$x_i^+ = \mathcal{R}_i(x_i), \quad \mathbf{e}^{i+}_{-i} = \mathbf{e}^i_{-i}, \tag{5.14}$$

where $\mathcal{R}_i$ is defined in (5.3e). However, lack of coordination between resets can induced the same issues discussed in Chapter 2 (Section 2.2). To avoid this, we will incorporate the hybrid coordinated restarting mechanism described in (5.3). Figure 5.5 shows a block-diagram representation of the multi-time scale hybrid dynamics of each player.

Well-Posed Coordinated Hybrid System with Partial Information

To write the coordinated HDS in vectorial form, we introduce the matrices $\mathcal{Q} \coloneqq D(\mathcal{Q}_i) \in \mathbb{R}^{(N^2-N)\times N^2}$ and $\mathcal{P} \coloneqq D(\mathcal{P}_i) \in \mathbb{R}^{N\times N^2}$, and note that $q = \mathcal{P}\mathbf{e} \in \mathbb{R}^{N^2-N}$. Additionally, we define the state $\hat{q} \coloneqq \mathcal{Q}\mathbf{e}$, such that using $\mathcal{P}\mathcal{P}^\top = I_N$, $\mathcal{Q}\mathcal{Q}^\top = I_{N^2-N}$, and $\mathcal{P}\mathcal{Q}^\top = 0$, we can write $\mathbf{e} = \psi(q, \hat{q}) \coloneqq \mathcal{P}^\top q + \mathcal{Q}^\top \hat{q}$, where $\mathbf{e} = \left(\mathbf{e}^1, \cdots, \mathbf{e}^N\right)$, and express the overall

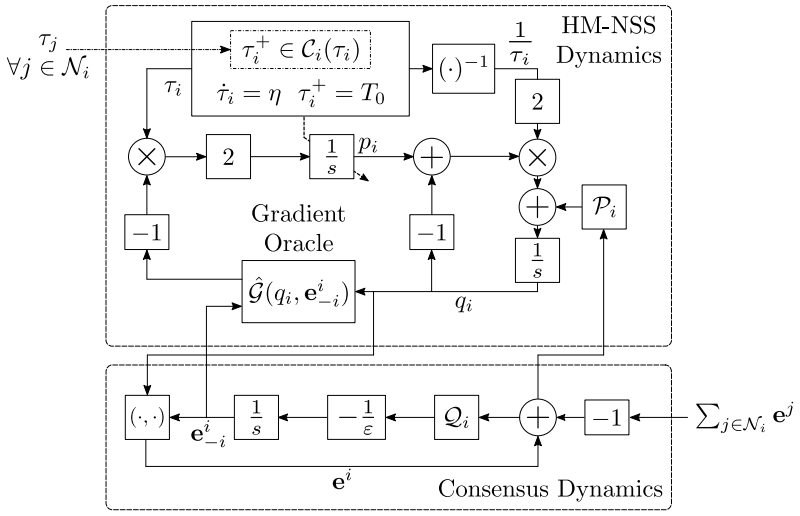


Figure 5.5. Scheme of Individual HM-NSS dynamics for games with partial information. Consensus dynamics are implemented to estimate the actions of other players.

hybrid NSS dynamics as a HDS with state $(x, \hat{q})$, where $x \coloneqq (q, p, \tau) \in \mathbb{R}^{3N}$, and data $\mathcal{H}_2 = (C_2, F_2, D_2, G_2)$. The flows are given by

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \\ \dot{\hat{q}} \end{pmatrix} = F_2(x, \hat{q}) \coloneqq \begin{pmatrix} 2\mathcal{T}^{-1}(p - q) - \mathcal{P}\mathbf{L}\psi(q, \hat{q}) \\ -2\mathcal{T}\hat{\mathcal{G}}(\psi(q, \hat{q})) \\ \eta \mathbf{1}_N \\ -\frac{1}{\varepsilon}\mathcal{Q}\mathbf{L}\psi(q, \hat{q}) \end{pmatrix}, \tag{5.15}$$

where $\mathbf{L} \coloneqq \mathcal{L} \otimes I_n$ denotes the communication matrix of the graph $\mathbb{G}$. The continuous-time dynamics in (5.15) are allowed to evolve whenever $(x, \hat{q})$ belongs to the flow set:

$$\begin{aligned} C_2 \coloneqq \Big\{ (x, \hat{q}) \in \mathbb{R}^{N^2+2N} : \; & q \in \mathbb{R}^N, p \in \mathbb{R}^N, \\ & \tau \in [T_0, T]^N, \; \hat{q} \in \mathbb{R}^{N^2-N} \Big\}. \end{aligned} \tag{5.16}$$

On the other hand, the jump set is defined as:

$$D_2 \coloneqq \Big\{ (x, \hat{q}) \in \mathbb{R}^{N^2+2N} : x \in C, \; \max_{i \in \mathcal{V}} \tau_i = T \Big\}, \tag{5.17}$$

and the discrete-time dynamics of the algorithm are given by:

$$(x^+, \hat{q}^+) \in G_2(x, \hat{q}) \coloneqq G_1(q, p, \tau) \times \{\hat{q}\}, \tag{5.18}$$

where G_1 is defined as in (5.3). Thus, the jump map affects only the momentum coefficients τ and the state p via the reset policy $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$. Similar to Lemma 4.1, the next lemma follows directly by construction of the HDS.

Lemma 5.5. For the HDS $\mathcal{H}_2 \coloneqq (C_2, F_2, D_2, G_2)$, all the properties of Lemma 4.1 still hold.

We will study the stability properties of the HDS $\mathcal{H}_2$ with respect to the following compact set:

$$\mathcal{A}_{\mathbb{G}} \coloneqq \mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_N \otimes q^*)\}, \tag{5.19}$$

where $\mathcal{A}$ was defined in (5.5). In this case, we will use the following restricted reverse-Lipschitz assumption, also used in [122] for optimization, and in [123] for NES with static inertia.

Assumption 5.4. There exists $\zeta > 0$ such that $|\mathcal{G}(q) - \mathcal{G}(q^*)| \geq \zeta |q - q^*|$, for all $q \in \mathbb{R}^N$.

Next, we have the following result, which leverages items (i_1)-(i_5) of Theorems 5.1-5.3.

Theorem 5.3. Let $\mathcal{G}$ describe a strictly monotone game. Suppose that Assumptions 5.2, 5.3 and 5.4 hold, and consider the HDS $\mathcal{H}_2$ under (RC_1). If $\mathcal{S}_\delta$ is $\ell-$GC with $0 < \delta < (1 - \eta)/\ell$, then under any of the conditions (i_1)-(i_5) the following holds:

a) For all $\varepsilon \in (0, \varepsilon_\delta^*)$, the set $\mathcal{A}_{\mathbb{G}}$ is UGAS. Here ε^* is given by

$$\varepsilon_\delta^* \coloneqq \frac{1}{2\sigma_{\mathcal{L}}\sqrt{N}} \left(1 + \sigma_r^2 \frac{\max\left\{ \frac{1}{T^2} + 4\frac{\ell}{T\lambda_{\max}(\mathcal{L})} \ , \ 2 + 2\frac{\ell}{T\lambda_{\max}(\mathcal{L})} \right\}}{\delta \min\{1, \zeta^2\}} \right)^{-1}. \quad (5.20)$$

b) For each $(\hat{t}, \hat{j}, \nu) \in \mathbb{R}^3_{>0}$ and each compact set $K_x \times K_{\hat{q}} \subset C_2 \cup D_2$, there exists ε^{**} such that for each $\varepsilon \in (0, \varepsilon^{**})$ and each solution of $\mathcal{H}_2$ with $x(0,0) \in K_x$ and $\hat{q}(0,0) \in K_{\hat{q}}$, there exists a solution $\tilde{x}$ of system $\mathcal{H}_1$ with $\tilde{x} \in K_x$ such that x and $\tilde{x}$ are $(\hat{t}, \hat{j}, \nu)$-close.

Item (a) of Theorem 5.3 establishes robust stability and convergence properties for the hybrid NSS dynamics $\mathcal{H}_2$. On the other hand, item (b) establishes that on compact sets of initial conditions and on compact time domains, the trajectories x will behave as the trajectories of the "full-information" system $\mathcal{H}_1$ as $\varepsilon \to 0^+$ in (5.15). In particular, by combining items (a) and (b), we recover the convergence bounds of Theorems4.1, 5.1, and 5.2, now in a semi-global practical sense as $\varepsilon \to 0^+$. This behavior is illustrated in Figure 5.6, which shows the trajectories q and $\hat{q}$ in a κ-strongly monotone game. As it can be observed, the solutions of $\mathcal{H}_2$ approximate the solutions of $\mathcal{H}_1$ as $\varepsilon \to 0^+$.

Remark 5.7. Assumption 5.4 always holds for *κ-strongly monotone games* with $\zeta = \kappa$. Hence, for these games one can compute an alternative expression of ε_δ^* by substituting Assumptions 5.2-5.4 in Theorem 5.3 by Assumption 5.1 when $\mathcal{S}_\delta$ is $(\sigma_\phi \ell)$-GC. Moreover, to guarantee that $\mathcal{S}_\delta$ is ℓ-GC, required in Theorem 5.3, a suitable upper bound for T can be obtained by mirroring the derivations of Lemma 5.2, which we omit here due to space limitations.

To our best knowledge, Theorem 5.3 is the first result in the literature that establishes robust convergence and stability properties for decentralized momentum-based

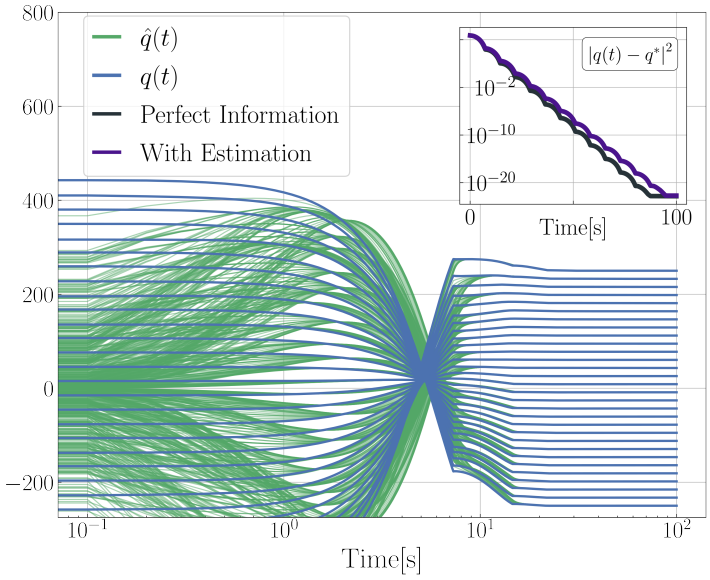


Figure 5.6. Trajectories of actions q and estimates $\hat{q}$ in a non-potential κ-strongly monotone quadratic game with $N = 30$, $\kappa = 0.01$, $\ell = 0.1$, $\tau_s(0,0) = 0.1 \cdot \mathbf{1}_N$, and $\varepsilon = 5 \times 10^{-3}$. The inset shows the distance to the NE.

NSS algorithms over graphs. Note that the stable incorporation of the multiple-time scale consensus mechanism is enabled by the use of resets, since otherwise no $\mathcal{KL}$ bound (or strong Lyapunov function) would exist for the slow dynamics (also called "reduced dynamics" in the singular perturbation literature [22, Ch. 11]) of the flows, which are precisely given by (5.1).

5.1.3 Model-Free NSS with Momentum

In the previous sections, we studied algorithms that made use of gradient Oracles with full or partial information. In this section, we dispense with this assumption by designing accelerated *model-free* hybrid NSS dynamics, suitable for applications where players have access only to real-time *measurements* of the signals that correspond to their cost functions ϕ_i (e.g., the difference between the individual cost and revenue in a market), which are generated by the game. Such payoff-based algorithms can be designed using tools recently developed in the context of hybrid equilibrium seeking control [111, 25].

To achieve model-free NSS, each player i generates an individual probing signal

$t \mapsto \mu_i(t)$, obtained as the solution of a dynamic oscillator with state $\mu_i := (\tilde{\mu}_i, \hat{\mu}_i) \in \mathbb{R}^2$, evolving on the unit circle $\mathbb{S}^1$ according to

$$\dot{\mu}_i = \frac{1}{\varepsilon_p}\Omega_i \mu_i, \quad \mu_i \in \mathbb{S}^1, \quad \Omega_i := 2\pi\varsigma_i \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \tag{5.21}$$

where ε_p and ς_i are positive tunable parameters. Note that $\mathbb{S}^1$ is forward invariant under the dynamics of μ_i. Using this probing signal, each player implements the flows:

$$\begin{pmatrix} \dot{q}_i \\ \dot{p}_i \\ \dot{\tau}_i \end{pmatrix} = \begin{pmatrix} \frac{2}{\tau_i}(p_i - q_i) \\ -\frac{4}{\varepsilon_a}\tau_i \phi_i(q + \varepsilon_a \tilde{\mu}) e_1^\top \mu_i \\ \eta \end{pmatrix}, \tag{5.22}$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_n) \in \mathbb{R}^{2N}$, and where $\tilde{\mu}$ is the vector that contains the odd components of μ. The dynamics (5.22) use real-time *measurements* of the cost ϕ_i, and are implemented whenever $\tau_i \in [T_0, T)$. Conversely, when $\tau_i = T$ and players are uncoordinated, they reset their states according to the dynamics $x_i^+ = \mathcal{R}_i(x_i), \;\; \mu_i^+ = \mu_i$, where $\mathcal{R}_i$ is defined as in (5.3). The constant $\varepsilon_a > 0$ is also a tunable parameter.

We impose the following assumption on the parameters ς_i of (5.21), which is standard in the averaging-based NES literature [35, 124, 113, 112].

Assumption 5.5. For all i, ς_i is a positive rational number, $\varsigma_i \neq \varsigma_j$, $\varsigma_i \neq 2\varsigma_j$, $\varsigma_i \neq 3\varsigma_j$, for all $i \neq j \in \mathcal{V}$.

As in the model-based case, an uncoordinated implementation of the model-free hybrid dynamics can be detrimental to the stability and/or transient performance of the algorithm. Thus, we incorporate the hybrid coordination mechanism described in (5.3) to coordinate the resets of the players. Using the set-valued coordination mechanism, the

overall discrete-time dynamics of the algorithm are modeled by the difference inclusion

$$(x^+, \mu^+) \in G_3(x, \mu) \coloneqq G_1(x) \times \{\mu\}, \tag{5.23}$$

where G_1 is given by (4.9a). This jump map will preserve the sequential nature of the resets needed to guarantee a well-posed HDS that satisfies (4.6) and (4.7). Using $\bar{\phi} \coloneqq (\phi_1, \phi_2, \ldots, \phi_N)$, the continuous-time dynamics of the model-free hybrid NSS algorithm can be written in vector form as:

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \\ \dot{\mu} \end{pmatrix} = F_3(x, \mu) \coloneqq \begin{pmatrix} 2\mathcal{T}^{-1}(p - q) \\ -\frac{4}{\varepsilon_a}\mathcal{T}\bar{\phi}(q + \varepsilon_a\tilde{\mu})\tilde{\mu} \\ \eta \\ \frac{1}{\varepsilon_p}D(\mathcal{R}_i)\mu \end{pmatrix}, \tag{5.24}$$

and the flow and jump sets are defined as:

$$C_3 \coloneqq C_1 \times \mathbb{T}^N \quad \text{and} \quad D_3 \coloneqq D_1 \times \mathbb{T}^N. \tag{5.25}$$

Figure 5.7 shows a scheme of the proposed model-free NSS dynamics.

Semi-Global Practical Stability Results

The data $\mathcal{H}_3 = (C_3, F_3, D_3, G_3)$ defines the third hybrid NSS dynamics considered in this chapter. The stability and convergence properties of $\mathcal{H}_3$ are given in the following theorem, which also leverages items (i_1)-(i_5) of Theorems 1-3.

Theorem 5.4. Let $\mathcal{G}$ describe a strictly monotone game, and consider the HDS $\mathcal{H}_3$ under (RC_1). Then, under any of the conditions (i_1)-(i_5) the following holds:

a) The set $\mathcal{A} \times \mathbb{T}^N$ is SGPAS as $(\varepsilon_p, \varepsilon_a) \to 0^+$.

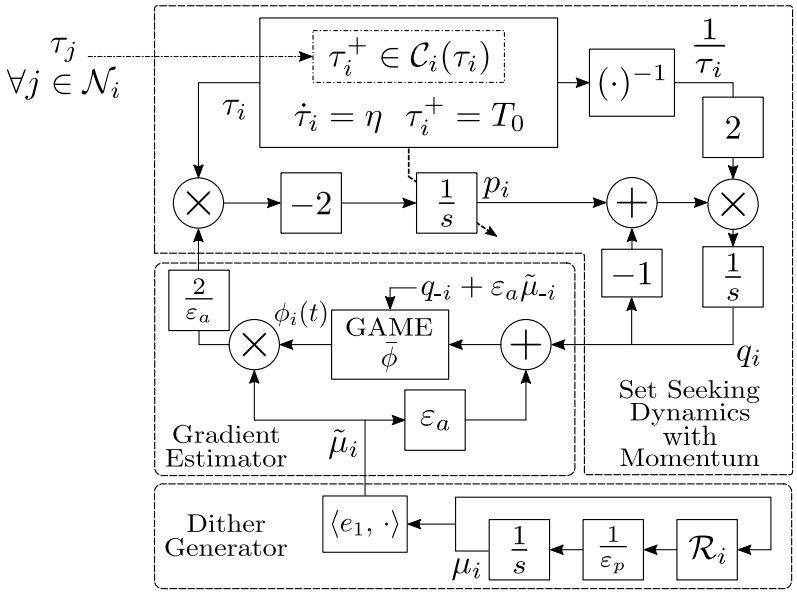


Figure 5.7. Scheme of Individual Model-Free HM-NSS dynamics with real-time measurements of the cost.

b) For each $(\hat{t}, \hat{j}, \nu) \in \mathbb{R}^3$ and each compact set $K_x \subset C_1 \cup D_1$, $\exists\ \varepsilon_a^* > 0$ s.t. $\forall\ \varepsilon_a \in (0, \varepsilon_a^*)\ \exists\ \varepsilon_p^* > 0$ s.t. $\forall\ \varepsilon_p \in (0, \varepsilon_p^*)$, and for each trajectory x of system $\mathcal{H}_3$ with $x(0,0) \in K_x$ there exists a solution $\tilde{x}$ of system $\mathcal{H}_1$ such that x and $\tilde{x}$ are $(\hat{t}, \hat{j}, \nu)$-close.

Similar to Theorem 5.3, the result of Theorem 5.4 establishes two key properties: First, for any desired precision $\nu > 0$, and any compact set of initial conditions K_x, every solution of the HDS $\mathcal{H}_3$ initialized in K_x will satisfy a bound of the form[1]

$$|x(t,j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t+j) + \frac{\nu}{2}, \tag{5.26}$$

with $\beta \in \mathcal{KL}$, provided the parameters ε_a and ε_p are selected sufficiently small. Moreover, item (b) implies that by selecting ε_a and ε_p sufficiently small, the trajectories x of $\mathcal{H}_3$ will recover all the fast convergence bounds established in Theorems 3.3, 5.2, modulo a small residual error and on compact sets.

[1]We note that $|\mu(t,j)|_{\mathbb{T}^n} = 0$ for all (t,j) in the domain of the solutions.

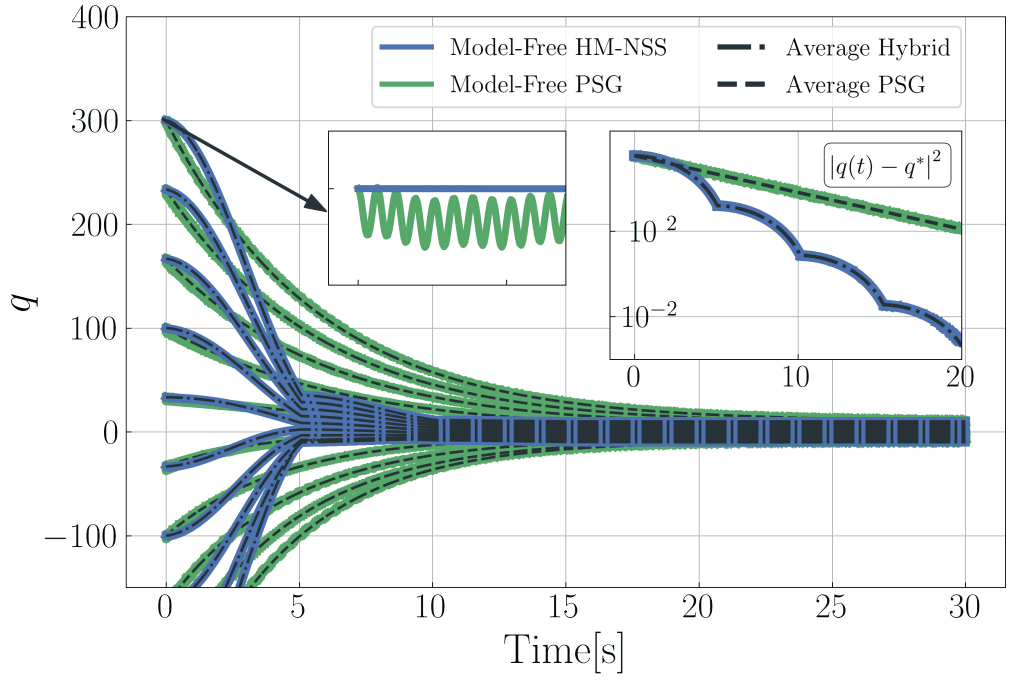


Figure 5.8. Trajectories of $\mathcal{H}_1$ and $\mathcal{H}_3$ in a non-potential κ-strongly monotone quadratic game with $\kappa = 0.197$, $\ell = 0.2$ and $n = 10$.

> **Remark 5.8.** The model-free dynamics $\mathcal{H}_3$ are based on averaging theory for (perturbed) hybrid systems [30, 25]. In particular, as $\varepsilon_a, \varepsilon_p \to 0^+$ the trajectories of $\mathcal{H}_3$ behave as their average hybrid dynamics (modulo a small perturbation), which are precisely given by $\mathcal{H}_1$. Both dynamics are set-valued, which differs from existing results in the literature of model-free Nash seeking [35]. Figure 5.8 shows a numerical experiment in a κ-strongly monotone quadratic game where a solution of $\mathcal{H}_3$ is compared to a solution of the model-free dynamics of [35].

We finish this section by commenting on the extensions of system $\mathcal{H}_3$ to applications where players could have access to an individual "Black-Box Oracle" that allows them to *evaluate* (as opposed to measure) their local cost ϕ_i at their current state q_i, using estimations of the actions of the other players (in a similar spirit as in Section 5.1.2) and without knowledge of the mathematical form of ϕ_i (e.g., using dynamic simulators). In this case, we can follow the same approach of Section 5.1.2, incorporating an auxiliary

estimation state $\hat{q}$. In this, the vectorial hybrid system $\mathcal{H}_4 = (C_4, F_4, D_4, G_4)$ will have flow map given by

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\tau} \\ \dot{\mu} \\ \dot{\hat{q}} \end{pmatrix} = F_4(\zeta) \coloneqq \begin{pmatrix} 2\mathcal{T}^{-1}(p-q) - \mathcal{P}\mathbf{L}\psi(q,\hat{q}) \\ -\frac{4}{\varepsilon_a}\mathcal{T}\bar{\phi}(\psi(q+\varepsilon_a\tilde{\mu},\hat{q}))\tilde{\mu} \\ \eta \\ \frac{1}{\varepsilon_p}D(\mathcal{R}_i)\mu \\ -\frac{1}{\varepsilon_c}\mathcal{Q}\mathbf{L}\psi(q,\hat{q}) \end{pmatrix}, \tag{5.27}$$

jump map $G_4(x,\mu,\hat{q}) \coloneqq G_1(x) \times \{\mu\} \times \{\hat{q}\}$, flow set $C_4 \coloneqq C_1 \times \mathbb{T}^n \times \mathbb{R}^{N^2-N}$ and jump set $D_4 \coloneqq D_1 \times \mathbb{T}^n \times \mathbb{R}^{N^2-N}$. For this hybrid system, a result like Theorem 5.4-(a) also holds, now with respect to the set $\mathcal{A} \times \mathbb{T}^n \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*)\}$ and as $(\varepsilon_p, \varepsilon_a, \varepsilon_c) \to 0^+$. Similarly, a result like Theorem 5.4-(b) holds by noting that the average hybrid dynamics of $\mathcal{H}_4$ are precisely given by the HDS $\mathcal{H}_2$ (modulo a small perturbation on the gradient), whose robust stability properties were already established in Section 5.1.2. Thus, we can follow exactly the same steps of the proof of Theorem 5.3 to obtain an equivalent result.

In this section we introduced a class of Nash set-seeking algorithms with dynamic momentum for the efficient solution of non-cooperative games with finitely many players. The algorithms are modeled by hybrid dynamical systems that incorporate continuous-time dynamics with momentum and discrete-time coordinated resets. We developed model-based algorithms, as well as algorithms suitable for games with partial information and model-free settings where players have access only to measurements of their cost. In each case, we established robust stability and accelerated convergence properties using multi-time scale techniques for hybrid dynamical systems.

In the next section, we study how dynamics of the form (5.1) can be applied in decentralized concurrent learning scenarios, where we consider a network of agents collaborating to learn a common model of a scalar signal, by exchanging information via a

graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$ using a consensus mechanism. In this context, the map $\mathcal{G}$ in (5.1) takes the form of a vector field $\mathbf{A}(t)q + \mathbf{L}q$, where $\mathbf{A}(t)$ is a symmetric matrix incorporating the agents' individual learning processes, and $\mathbf{L} = \mathcal{L} \otimes I_n$, with $\mathcal{L}$ being the Laplacian matrix of $\mathbb{G}$, encapsulates the network topology and governs inter-agent information flow. While $\mathbf{A}(t)$ is symmetric, $\mathbf{L}$ typically is not, especially for graphs with directed edges. This asymmetry precludes the existence of a potential function for the entire system, paralleling our earlier discussion of nonpotential games. As our analysis will reveal, this characteristic can induce instability when reset mechanisms are implemented infrequently. However, stability can be restored by increasing the frequency of these resets.

5.2 Decentralized Concurrent Learning with Coordinated Momentum and Restart

Concurrent Learning (CL) techniques have emerged as powerful data-driven tools for designing estimation and learning dynamics in a wide range of applications where persistence of excitation (PE) conditions are either impractical or infeasible [125, 45]. These techniques have demonstrated their utility in diverse fields, including parameter estimation in batteries [6], exoskeleton robotic systems [126, 127], mobile robots and aerial vehicles [128], extremum seeking algorithms [129], and reinforcement learning controllers [2, 41]. In these applications, extensive datasets containing historical *recorded* measurements of the relevant system signals are often available and can be leveraged for estimation purposes. When these datasets are "sufficiently rich", they can be seamlessly integrated into dynamic estimation algorithms, enabling (uniform) exponential convergence to the unknown parameters even in the absence of PE conditions.

However, relaxations of PE conditions can lead to suboptimal transient performance, particularly in terms of slow convergence rates that depend on the "level of richness" of the dataset used by the algorithm. Since datasets readily available in applications may

exhibit prohibitively small levels of richness, there is a growing need for the development of enhanced CL techniques that can accelerate the convergence rate of the estimation dynamics while maintaining the desirable stability and robustness properties. One promising direction to alleviate the slow convergence issue in decision-making algorithms is the incorporation of momentum with dynamic damping, see [6, 130, 131]. For single-agent gradient-based dynamics with momentum, the use of decreasing damping has been shown to play a crucial role in inducing favorable acceleration properties [132, 17, 18, 133]. However, it has also been shown that stability bounds in terms of $\mathcal{KL}$ functions may not exist for such systems unless the damping coefficients are persistently exciting [29, Thm. 2], a condition that precludes vanishing coefficients. Furthermore, it is well-known that, without proper tuning, the use of momentum can lead to undesirable oscillations [93]. To address potential instability issues and to eliminate oscillatory behaviors, restart mechanisms that reset the momentum have been developed for single-agent systems using adaptive [93, 134] and periodic policies (usually called "scheduled") with carefully selected restarting frequencies [17, 25, 135, 93]. Recent works have also investigated the development of similar momentum-based algorithms in multi-agent systems, including distributed continuous-time heavy-ball dynamics with constant damping [136], limiting equations of stochastic recursive algorithms as multi-agent flows with momentum [137], and decision-making algorithms with momentum for high-order multi-agent systems [138, 1, 139]. However, existing approaches have primarily focused on *undirected* network topologies. Additionally, the incorporation of momentum and restarting mechanisms in *decentralized concurrent learning algorithms* has remained unexplored. Such algorithms are essential when a network of agents seeks to collaboratively and efficiently learn a common model by sharing local estimates with neighboring agents, *without revealing their private data*. Applications of these algorithms span various domains, including source seeking in autonomous mobile robots [140], adaptive formation control of robotic teams [141], and cooperative adaptive control [142].

Motivated by the previous background, in this section we study the synthesis and analysis of decentralized concurrent learning dynamics with momentum and restart for general directed graphs. In particular, we consider a model that extends the centralized dynamics studied in [17, 18], and [25] to cases where each agent implements its own restart policy and shares information only with neighbors characterized by the topology of the communication graph. To assess the impact of the topology of the graph on the stability properties of the dynamics, we exploit analytical tools from graph theory [143] and hybrid dynamical system's theory [33]. Using these tools, this section presents the following primary contributions:

(1) We first introduce a class of multi-agent concurrent learning (CL) algorithms that incorporate momentum and a restarting mechanism coordinated by a centralized resetting state. We demonstrate that if: (a) the graph is strongly connected, (b) the overall data collected by the agents satisfies a "cooperative richness condition," and (c) the restarting frequency exceeds a certain threshold that encodes the "asymmetry" of the communication graph, then the resulting error estimation dynamics are input-to-state stable [144] with respect to measurement noise and model error approximations. Furthermore, the convergence is exponential with rates assignable via the restarting period. These results are presented in Theorem 5.5.

(2) Next, by leveraging the robustness properties of the dynamics, we interconnect the momentum-based concurrent learning algorithms with a fully decentralized coordinated restarting mechanism, enabling a complete decentralized implementation. The resulting dynamical systems are also globally stable and exhibit convergence rates consistent with Theorem 5.5, following an initial synchronization phase of the restarting times. These results are presented in Theorem 5.6.

(3) Finally, we present three applications of the proposed momentum-based CL algorithms with restart within the context of data-enabled control: (a) cooperative parameter estimation without persistently exciting regressors in networks where nodes

have private data with heterogeneous informativity properties; (b) data-enabled cooperative model-reference adaptive control; (c) data-enabled cooperative feedback-optimization. By employing (hybrid) Lyapunov-based techniques, we show that the resulting closed-loop systems exhibit favorable stability and convergence properties, which are also illustrated via numerical examples.

Problem Formulation

We consider a decentralized learning problem in a multi-agent system (MAS), where a group of $N \in \mathbb{Z}_{\geq 2}$ agents seeks to collaboratively estimate a common model characterized by a parameter $\theta^\star \in \mathbb{R}^n$. The agents share information with each other via a directed communication network modeled by a *strongly connected* digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} \coloneqq \{1, 2, \ldots, N\}$ is the set of nodes, and $\mathcal{E}$ is the set of edges. We denote by $(i, j) \in \mathcal{E}$ a directed edge from node i to node j, we call node i an *in-neighbor* of node j, and we call node j an *out-neighbor* of node i. We consider digraphs that do not have self-arcs. A weighted *Laplacian* matrix $\mathcal{L} = [l_{ij}] \in \mathbb{R}^{N \times N}$ associated with $\mathcal{G}$ satisfies the following: the off-diagonal entries are such that $l_{ij} < 0$ if (i, j) is an edge, and $l_{ij} = 0$ otherwise; the diagonal entries l_{ii} are determined such that every row of $\mathcal{L}$ sums to zero, and all its nonzero eigenvalues have positive real part [143, Lemma 6.5]. A digraph is *strongly connected* if for any two distinct nodes i and j, there is a path from i to j. The Laplacian matrix of a strongly connected digraph satisfies $\text{rank}(\mathcal{L}) = N - 1$ [143, Ch. 6]. We assume that each agent $i \in \mathcal{V}$ has access to both real-time and past recorded measurements of a signal of the form

$$\psi_i^\star(t, d_i(t)) = \phi_i(t)^\top \theta^\star + d_i(t), \tag{5.28}$$

where $d_i \in \mathbb{R}$ represents an unknown and possibly time-varying disturbance, and $\phi_i : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ represents a regressor function (or basis functions), which is assumed to be

continuous, uniformly bounded, and known to the i^{th} agent. These assumptions are typical in parameter estimation problems in the context of single-agent [41, 130, 6, 6, 45] and distributed CL [129, 142].

Model Description and Key Assumptions

The main idea behind Concurrent Learning (CL) is to use both real-time and past recorded measurements of $\psi_i^\star$ in (5.28) to recursively estimate the unknown parameter $\theta^\star$. This approach is particularly useful in situations where the regressors $t \mapsto \phi_i(t)$ are not persistently exciting [41], that is, when there are no constants $T, m > 0$ such that $\int_t^{t+T} \phi_i(s)\phi_i(s)^\top ds \geq mI_n$, for all $t > 0$ and all $i \in \mathcal{V}$. To address this limitation by leveraging each agent i's access to past recorded measurements of $\psi_i^\star$, this paper introduces a *decentralized momentum-based concurrent learning* (DMCL) algorithm to estimate $\theta^\star$. In this algorithm, each agent $i \in \mathcal{V}$ maintains its own individual estimate of $\theta_i^\star$, denoted as $\theta_i \in \mathbb{R}^n$, which is updated according to the following dynamics:

$$\dot{\theta}_i(t) = \frac{2}{\tau_i(t)}(p_i(t) - \theta_i(t)), \quad \dot{\tau}_i(t) \in [0, \omega], \quad \forall \ \ i \in \mathcal{V}, \tag{5.29}$$

where τ_i is a dynamic, non-decreasing coefficient, with rate of growth upper bounded by $\omega > 0$, and which satisfies

$$\tau_i(t) \in [T_0, T], \quad \forall\, t \in \mathbb{R}_{\geq 0}, \quad T > T_0 > 0,$$

where (T, T_0) are tunable parameters. The auxiliary state $p_i \in \mathbb{R}^n$ captures the incorporation of momentum, and it satisfies

$$\dot{p}_i(t) = -2\tau_i(t)\left(\Lambda_i\left(\theta_i(t), \nu_i(t), t, \upsilon_i\right) + k_c \sum_{j \in \mathcal{V}} a_{ji}\left(\theta_i(t) - \theta_j(t)\right)\right), \tag{5.30}$$

where $k_c > 0$ is a tunable gain, Λ_i is a suitable mapping described below, and a_{ji} is the ji^{th} entry of the adjacency matrix of the graph $\mathcal{G}$ modeling the flow of information betwen agent i and its neighbors. The key components of the DMCL dynamics are explained below:

a) In (5.30), the function Λ_i has the general form

$$\Lambda_i(\theta_i, \nu_i, t, \upsilon_i) = k_t \Psi_i(\theta_i, t, \upsilon_i) + k_r \Phi_i(\theta_i, \nu_i), \tag{5.31}$$

where $k_r > 0$ and $k_t \geq 0$ are tunable constants.

b) In (5.31), the function Ψ_i is given by

$$\Psi_i(\theta_i, t, \upsilon_i(t)) := \phi_i(t) \left(\hat{\psi}_i(\theta_i, t) - \psi_i^\star(t, \upsilon_i(t)) \right), \tag{5.32}$$

and it incorporates the *real-time* information available to the i^{th} agent, where $\psi_i^\star$ is given by (5.28), $\hat{\psi}_i(\theta_i, t) := \phi_i(t)^\top \theta_i$, and $\upsilon_i(t) := d_i(t)$ is the time-varying disturbance in (5.28).

c) The function Φ_i in (5.31) is given by

$$\Phi_i(\theta_i, \nu_i) := \sum_{k=1}^{\bar{k}_i} \phi_i(t_{i,k}) \Big(\hat{\psi}_i(\theta_i, t_{i,k}) - \psi_i^\star(t_{i,k}, \nu_{i,k}) \Big), \tag{5.33}$$

and it incorporates past recorded measurements of the signal $\psi_i^\star$ in (5.28) and the regressor ϕ_i, obtained at a sequence of times $\{t_{i,k}\}_{k=1}^{\bar{k}_i}$, where $\bar{k}_i \in \mathbb{Z}_{\geq 1}$ is the number of measurements recorded by agent i, and where $\nu_{i,k} := d_i(t_{i,k})$ captures the persistent disturbances occurring during data collection in (5.28), which are stored in the vector $\nu_i := (\nu_{i,1}, \nu_{i,2}, \ldots, \nu_{i,\bar{k}_i}) \in \mathbb{R}^{\bar{k}_i}$.

d) The last term in the dynamics of p_i captures the exchange of information between

agent i and its neighbors. Note that, in general, we have $a_{ij} \neq a_{ji}$ because the graph $\mathcal{G}$ can be directed.

To study the DMCL dynamics, the data matrix associated to the i^{th} agent is defined as follows:

$$\Delta_i \coloneqq \sum_{k_i=1}^{\bar{k}_i} \phi(t_{i,k})\phi(t_{i,k})^\top \in \mathbb{R}^{n\times n}. \tag{5.34}$$

Instead of assuming that every matrix Δ_i is positive definite, as in standard single-agent concurrent learning (CL) [41], we will assume a weaker "cooperative" richness condition on the overall data of the network [145, Def. 2].

Assumption 5.6. There exists a constant $\alpha > 0$, such that

$$\sum_{i=1}^{N} \Delta_i \succeq \alpha I_n. \tag{5.35}$$

Moreover, the graph $\mathcal{G}$ is strongly connected.

If (5.35) holds, the data $\{\Delta_i\}_{i\in\mathcal{V}}$ is said to be *cooperatively sufficiently rich* (CSR).

Remark 5.9. Assumption 5.6 allows for some agents to have uninformative data (e.g., $\phi_i(t_{i,k}) = 0$) provided other agent's data is sufficiently rich to satisfy (5.35), see also [146]. This is an important relaxation for large-scale MAS where, unlike standard centraCL [41], it might be unreasonable to assume that *every* agent's data satisfies $\Delta_i \succ 0$. Moreover, note that in the DMCL dynamics, agents do not share their data with other agents in the network. In fact, only the local estimates θ_i are shared with the neighboring agents. This prevents the direct solution of the estimation problem using "single-shot" techniques and instead necessitates recursive algorithms that converge to $\theta^\star$ while preserving the privacy of individual data.

Connections to Momentum-Based Dynamics in Potential Settings

The form of the DMCL dynamics is closely related to the accelerated gradient flows with momentum studied in [17, 147, 18, 1], which have the general form

$$\dot{x}_1(t) = \frac{2}{\tau_c(t)}(x_2(t) - x_1(t)), \tag{5.36a}$$

$$\dot{x}_2(t) = -2\tau_c(t)\nabla f(x_1(t)), \tag{5.36b}$$

and where f is a suitable convex cost function and $\tau_c : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is a time-varying coefficient. Indeed, using the vectors $\theta := (\theta_1, \theta_2, \ldots, \theta_N)$, $p := (p_1, p_2, \ldots, p_N)$, the parameter error coordinates $\tilde{\theta} := \theta - \mathbf{1}_N \otimes \theta^\star$, $\tilde{p} := p - \mathbf{1}_N \otimes \theta^\star$, and the Laplacian matrix of the graph $\mathcal{L}$, the DMCL dynamics with a centralized coefficient $\tau = \tau_1 = \ldots = \tau_N$ can be written as the following dynamical system:

$$\begin{pmatrix} \dot{\tilde{\theta}} \\ \dot{\tilde{p}} \end{pmatrix} = \hat{\mathbf{F}}(\tilde{\theta}, \tilde{p}, \tau, t), \tag{5.37}$$

where $\hat{\mathbf{F}}$ is given by

$$\hat{\mathbf{F}}(\tilde{\theta}, \tilde{p}, \tau, t) = \begin{pmatrix} \frac{2}{\tau}(\tilde{p} - \tilde{\theta}) \\ -2\tau\left(k_t\mathbf{A}(t) + k_r\mathbf{\Delta} + k_c\mathbf{L}\right)\tilde{\theta} + \mathbf{U}(t) \end{pmatrix}. \tag{5.38}$$

In (5.38), $\mathbf{L} \coloneqq \mathcal{L} \otimes I_n$, $\mathbf{A}$ and $\mathbf{\Delta}$ are the block-diagonal matrices

$$\mathbf{A}(t) \coloneqq \operatorname{diag}\left(\left\{\phi_1(t)\phi_1(t)^\top, \ldots, \phi_N(t)\phi_N(t)^\top\right\}\right),$$

$$\mathbf{\Delta} \coloneqq \operatorname{diag}\left(\left\{\sum_{k=1}^{\bar{k}_1} \phi_1(t_{1,k})\phi_1^\top(t_{1,k}), \ldots, \sum_{k=1}^{\bar{k}_N} \phi_N(t_{N,k})\phi_N^\top(t_{N,k})\right\}\right),$$

and $\mathbf{U}$ is given by

$$\mathbf{U}(t) := \begin{bmatrix} -2\tau k_t \phi_1(t) \upsilon_1(t) + k_c \sum_{k=1}^{\bar{k}_1} \phi_1(t_{1,k}) \nu_{1,k} \\ \vdots \\ -2\tau k_t \phi_N(t) \upsilon_N(t) + k_c \sum_{k=1}^{\bar{k}_N} \phi_N(t_{N,k}) \nu_{N,k} \end{bmatrix}. \tag{5.39}$$

However, while similar decentralized algorithms have been studied in [138, 1, 139], the DMCL dynamics do not describe a standard gradient flow with momentum due to the lack of symmetry on $\mathcal{L}$, i.e., the right-hand side of (5.38) cannot be expressed as the gradient of a potential function, a property that usually plays a crucial role in the stability properties of momentum-based dynamics. The following example highlights some of the challenges that can arise when momentum is used and the multi-agent system (MAS) has a communication topology characterized by a directed graph.

Example 5.2. Consider a multi-agent system with three agents, i.e., $\mathcal{V} = \{1, 2, 3\}$. We let $k_t = 0$ and $d_i = 0$, and for simplicity we assume that all agents use the same coefficient $\tau_c = \tau_1 = \tau_2 = \tau_3$, with $\tau(0) = T_0$, $\omega = 1/2$, $T_0 = 0.1$, $T = 200$. We consider regressors $\phi_i(t) = (1, 10e^{-it}, 100e^{-2it})$ with collected data satisfying Assumption 5.6, and the parameter $\theta^\star = (1, -2, 1)$. The DMCL dynamics are implemented using $\dot{\tau}_c = \omega$ until $\tau_c = T$, at which point $\dot{\tau}_c$ is set to zero. This selection satisfies (5.29) and keeps τ bounded in the set $[T_0, T]$. The left plot of Figure 5.9 shows the evolution in time (in logarithmic scale) of the estimation error $\tilde{\theta} = \theta - \mathbf{1}_N \otimes \theta^\star$ when the graph $\mathcal{G}$ is fully connected. As observed, the estimation error converges to zero, which is consistent with the stability results of [17, Thm. 3] and the fact that in this case, the DMCL dynamics describe an accelerated gradient system, similar to (5.36). Now, suppose that the communication graph is a directed cycle graph, as shown in the inset of the center plot of Figure 5.9. In this case, the same DMCL algorithm ceases to be a momentum-based gradient flow and it exhibits

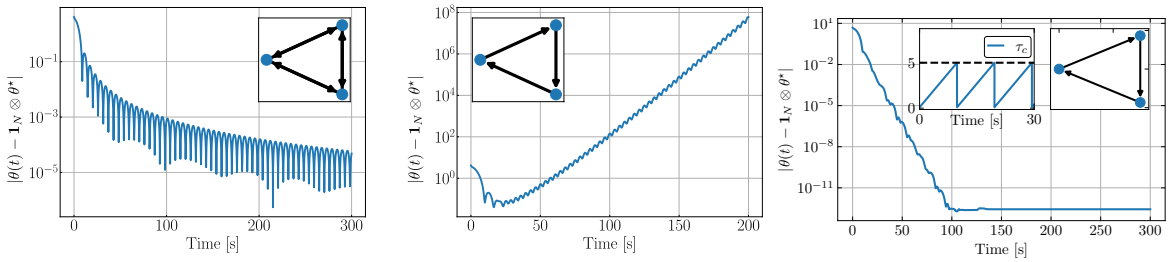


Figure 5.9. Solutions to DMCL without restart can exhibit stability in symmetric graphs (left) and instability in asymmetric graphs (center) when τ is sufficiently large. Stability in asymmetric graphs is recovered by employing a suitable coordinated restart mechanism (right).

> the instability shown in the plot. In particular, the asymmetric component of $\mathbf{L}$ now induces instability when τ becomes "sufficiently large", at approximately $t \approx 25s$, which corresponds to $\tau \approx 12.5$. The right plot, however, reveals a promising solution to the instability issue in asymmetric graphs. In particular, stability can be restored by implementing a well-designed coordinated restart mechanism that accounts for the graph's asymmetry. The details of this mechanism will be elaborated upon in the following sections.

DMCL with Coordinated Restart

To address the instability observed in Example 5.2, while simultaneously inducing suitable convergence rates achieved via momentum, we incorporate restart mechanisms into the algorithm (5.29)-(5.30). These mechanisms persistently reset the momentum $\dot{\theta}_i$ and the dynamic coefficients τ_i whenever they exceed a carefully selected upper bound T. The resets are performed according to the following discrete-time updates:

$$\theta_i^+ = \theta_i, \quad p_i^+ = p_i + \eta_i(\theta_i - p_i), \quad \tau_i^+ = T_0, \quad \forall\, i \in \mathcal{V}, \tag{5.40}$$

where $\eta_i \in \{0, 1\}$ is a pre-defined parameter indicating the restart policy of agent i. Similar resets have been shown to reduce oscillations in single-agent momentum-based algorithms [17, 93] and enhance their stability properties in the presence of persistent disturbances [25].

Note that the policy $\eta_i = 1$ implies $p_i^+ = \theta_i$, which in turn implies $\dot{\theta}_i^+ = 0$ in (5.29), that is, the momentum state of agent i is reset to zero. On the other hand, the choice $\eta_i = 0$ only resets the coefficients τ while keeping the momentum states constant during resets, thus emulating a heavy-ball system with a "persistently exciting" damping coefficient [29]. For multi-agent systems with undirected graphs, similar restart mechanisms of the form (5.40) have been studied in [139, 1]. However, the effectiveness of restarting in the context of multi-agent systems with *directed* graphs has remained largely unexplored, and, as suggested by Example 5.2, the extension is non-trivial.

Remark 5.10. The behavior observed in the center plot of Figure 5.9 clearly shows that a "slow" restart frequency (i.e., allowing T to be arbitrarily large) does not achieve stable parameter estimation, as the trajectories of the system initially approach the true parameter, but eventually diverge around $t \approx 25s$. Conversely, a very "fast" restart frequency might reduce the effectiveness of using momentum with dynamic damping, as it would keep T and p_i approximately constant. Alternatively, the right plot of Figure 5.9 demonstrates the emerging behavior of the DMCL algorithm when restarts are implemented by each node of the network under a "suitable" frequency and in a coordinated manner. In this case, the dynamics exhibit fast linear convergence to the true parameter, with a convergence rate tunable via the parameter T. While similar phenomena have been recently observed in game-theoretic problems [1], the use of momentum and restart in decentralized CL problems, and its dependence on the system's data, graph topology, and perturbed models (5.28), have remained largely unexplored.

The previous observations motivate the main research problem that we study in this section:

Problem 5.1. Characterize the restart mechanisms that: a) robustly stabilize the DMCL algorithm in directed networks; b) achieve ISS with respect to the disturbances d_i in (5.28); c) induce network-wide acceleration properties in the MAS.

To tackle Problem 1, we first consider a centralized restart mechanism that makes use of a common state $\tau_c \in \mathbb{R}_{>0}$ that satisfies $\dot{\tau}_c \in [0, \omega]$. This "centralized" restarting state will initially simplify the analysis, and it will be removed in the subsequent subsections to encompass decentralized implementations. For the purpose of analysis, we also use an auxiliary state $s \in \mathbb{R}_{\geq 0}$ with dynamics $\dot{s} = 1$ to model any explicit dependence on time t.

Centralized Restart: Hybrid Systems Model

When using a common coefficient $\tau_c \in \mathbb{R}_{>0}$ to coordinate the restart of the DMCL algorithm, the resulting dynamical system can be modeled by the following differential inclusion, in vectorial form, with overall state $y_c \coloneqq (\theta, p, \tau_c, s)$:

$$\dot{y}_c \in \mathbf{F}_c(y_c, u) \coloneqq \begin{pmatrix} \frac{2}{\tau_c}(p - \theta) \\ -2\tau_c \mathbf{\Lambda}(\theta, s, u) \\ [0, \omega] \\ 1 \end{pmatrix}. \tag{5.41}$$

In (5.41), the state y_c evolves in the flow set $\mathbf{C}_c \times \mathbb{R}_{\geq 0}$, with input $u \coloneqq (\upsilon, \nu) \in \mathbb{R}^{N+\bar{k}}$, where the vectors υ and ν are defined as

$$\upsilon \coloneqq (\upsilon_1, \upsilon_2, \ldots, \upsilon_N) \in \mathbb{R}^N, \quad \nu \coloneqq (\nu_1, \nu_2, \ldots, \nu_N) \in \mathbb{R}^{\bar{k}},$$

where $\bar{k} \coloneqq \sum_{i \in \mathcal{V}} \overline{k}_i$. The function $\mathbf{\Lambda}$ is given by

$$\mathbf{\Lambda}(\theta, s, u) \coloneqq k_t \mathbf{\Psi}(\theta, s, \upsilon) + k_r \mathbf{\Phi}(\theta, \nu) + k_c \mathbf{L}\theta, \tag{5.42}$$

and the functions $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are defined as

$$\mathbf{\Psi}(\theta, s, \upsilon) \coloneqq (\Psi_1(\theta_1, s, \upsilon_1), \dots, \Psi_N(\theta_N, s, \upsilon_N)) \tag{5.43a}$$

$$\mathbf{\Phi}(\theta, \nu) \coloneqq (\Phi_1(\theta_1, \nu_1), \dots, \Phi_N(\theta_N, \nu_N)), \tag{5.43b}$$

where the functions Φ_i, Ψ_i were defined in (5.32)-(5.33) for all $i \in \mathcal{V}$. Since the vectors θ and p are both allowed to evolve in $\mathbb{R}^{nN}$, while $\tau_c \in [T_0, T]$, the set $\mathbf{C}_c$ is defined as:

$$\mathbf{C}_c \coloneqq \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times [T_0, T], \tag{5.44}$$

Therefore, the overall flows of the system have the form (1.3a), and are given by

$$(y_c, u) \in (\mathbf{C}_c \times \mathbb{R}_{\geq 0}) \times \mathbb{R}^{N+\bar{k}}, \qquad \dot{y}_c \in \mathbf{F}_c(y_c, u).$$

To incorporate the restarts (5.40) into the DMCL algorithm, each time the condition $\tau_c = T$ is satisfied, the state τ_c is allowed to be reset to T_0, while the states (θ_i, p_i) are updated as in (5.40). Therefore, using

$$\mathbf{R}_\eta \coloneqq \operatorname{diag}(\eta) \otimes I_n, \tag{5.45}$$

with $\eta = (\eta_1, \eta_2, \dots, \eta_N)$, the discrete-time updates of the state y_c of the hybrid system can be written in vectorial form as

$$y_c^+ = (\theta^+, p^+, \tau_c^+, s^+) = \hat{\mathbf{G}}_c(y_c) = (\theta, p + \mathbf{R}_\eta(\theta - p), T_0, s) \tag{5.46}$$

which are executed whenever $(\theta, p, \tau_c, s) \in \mathbf{D}_c \times \mathbb{R}_{\geq 0}$, where

$$\mathbf{D}_c \coloneqq \mathbb{R}^{nN} \times \mathbb{R}^{nN} \times \{T\}. \tag{5.47}$$

Therefore, the overall discrete-time dynamics of the system (which do not depend on the input u) with state y_c, have the form (1.3b), and are given by

$$(y_c, u) \in (\mathbf{D}_c \times \mathbb{R}_{\geq 0}) \times \mathbb{R}^{N+\bar{k}},\ y_c^+ = \mathbf{G}_c(y_c) \coloneqq \hat{\mathbf{G}}_c(y_c) \times \{s\}. \tag{5.48}$$

By combining (5.41) and (5.48), the DMCL algorithm with centralized restart can be viewed as a HDS of the form (1.1), with data

$$\mathcal{H}_c \coloneqq (\mathbf{C}_c \times \mathbb{R}_{\geq 0}, \mathbf{F}_c, \mathbf{D}_c \times \mathbb{R}_{\geq 0}, \mathbf{G}_c, u). \tag{5.49}$$

Note that in this centralized HDS the jump set (5.47) only imposes conditions on the state τ_c. Namely, a restart is enabled whenever $\tau_c = T$. If, at this time, a restart is not executed, solutions can only continue evolving by flowing using $\dot{\tau}_c = 0$, i.e., keeping $\tau_c = T$ constant for all time until a reset is executed. If $\dot{\tau}(t) = \text{constant} \in (0, \omega]$ for all time, then the HDS would model a DMCL algorithm with *scheduled* periodic restart, where the time between two consecutive restarts is $(T - T_0)(\text{constant})^{-1}$. However, the differential inclusion in (5.41) also allows us to consider scenarios where $\dot{\tau}$ is not constant but rather is any absolutely continuous function (between restarts) satisfying $\dot{\tau} \in [0, \omega]$, which includes functions that remain constant for arbitrarily long periods of time.

Before presenting our first main result, we introduce two technical propositions that play important roles in our results. All the proofs are presented in Appendix C.2.

Proposition 5.2. Suppose that Assumption 5.6 holds. Then, there exists a unit vector $q \in \mathbb{R}^N$ such that:

a) The entries q_i of q satisfy:

$$\overline{\sigma}_{\mathbf{Q}} \coloneqq \max_{i \in \mathcal{V}} q_i \geq \min_{i \in \mathcal{V}} q_i \coloneqq \underline{\sigma}_{\mathbf{Q}} > 0. \tag{5.50}$$

b) $q^\top \mathcal{L} = 0$ and $\mathcal{Q}\mathcal{L} + \mathcal{L}^\top \mathcal{Q} \succeq 0$ with $\mathcal{Q} \coloneqq \operatorname{diag}(q)$.

c) The function $\mathbf{\Lambda}$ in (5.42) with $k_t = 0$ and $\nu = 0$ can be decomposed as follows:

$$k_r \mathbf{\Phi}(\theta, 0) + k_c \mathbf{L}\theta = \mathbf{Q}^{-1} \left(\mathbf{\Sigma} + \mathbf{\Omega} \right) \tilde{\theta}, \tag{5.51}$$

where $\mathbf{Q} \coloneqq \mathcal{Q} \otimes I_n$, $\tilde{\theta} \coloneqq \theta - \mathbf{1}_N \otimes \theta^\star$,

$$\mathbf{\Sigma} \coloneqq k_r \mathbf{Q}\mathbf{\Delta} + \frac{k_c}{2} \left(\mathbf{Q}\mathbf{L} + \mathbf{L}^\top \mathbf{Q} \right) \tag{5.52a}$$

$$\mathbf{\Omega} \coloneqq \frac{k_c}{2} \left(\mathbf{Q}\mathbf{L} - \mathbf{L}^\top \mathbf{Q} \right), \tag{5.52b}$$

and $\mathbf{\Delta} \coloneqq \mathbf{diag}\left(\{\Delta_1, \Delta_2, \ldots, \Delta_N\} \right)$, where Δ_i is given by (5.34).

d) There exists a class-$\mathcal{K}_\infty$ function $\chi(\cdot)$ such that

$$\left[\mathbf{\Omega} + k_t \tilde{\mathbf{A}}(t) \right] \left[\mathbf{\Omega} + k_t \tilde{\mathbf{A}}(t) \right]^\top \preceq (\overline{\sigma}_{\mathbf{\Omega}}^2 + \chi(k_t)^2) I_{Nn}, \tag{5.53}$$

$\forall\, t \geq 0$, where $\tilde{\mathbf{A}}(t) \coloneqq \mathbf{Q}\mathbf{A}(t)$ and $\overline{\sigma}_{\mathbf{\Omega}}$ is the largest singular value of $\mathbf{\Omega}$.

Remark 5.11. By construction, if the Laplacian $\mathcal{L}$ is symmetric, then $\overline{\sigma}_{\mathbf{\Omega}}^2 = 0$. However, if $\mathcal{L}$ is asymmetric, then in general we have $\overline{\sigma}_{\mathbf{\Omega}}^2 \neq 0$. For the purpose of illustration, Figure 5.10 presents four examples of different graphs $\mathcal{G}$ and their corresponding numerical values of $\overline{\sigma}_{\mathbf{\Omega}}^2$.

Proposition 5.3. Suppose that Assumption 5.6 holds; then, there exist $\overline{\sigma}_{\mathbf{\Sigma}} \geq \underline{\sigma}_{\mathbf{\Sigma}} > 0$ such that

$$\overline{\sigma}_{\mathbf{\Sigma}} I_{Nn} \succeq \mathbf{\Sigma} \succeq \underline{\sigma}_{\mathbf{\Sigma}} I_{Nn}, \tag{5.54}$$

where $\mathbf{\Sigma}$ is given by (5.52a).

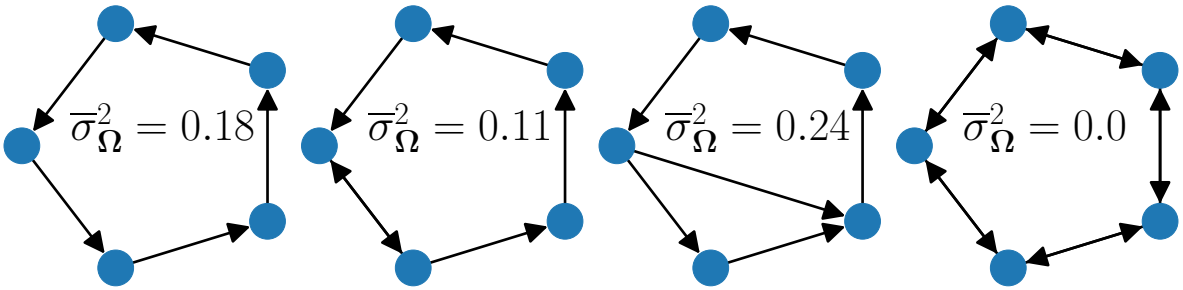


Figure 5.10. Parameter $\overline{\sigma}^2_{\mathbf{\Omega}}$ for strongly connected graphs with binary adjacency matrices and varying degrees of symmetry.

5.2.1 Input-to-State Stability of the Centralized Model $\mathcal{H}_c$

With Propositions 5.2-5.3 at hand, we are now ready to present the first main result of this paper, which provides conditions to stabilize the DMCL algorithm using a coordinating centralized state τ_c. In particular, we study the stability properties of $\mathcal{H}_c$ with respect to the closed set $\mathcal{A}_c \coloneqq \mathcal{A}_{\theta p} \times [T_0, T] \times \mathbb{R}_{\geq 0}$, where

$$\mathcal{A}_{\theta p} \coloneqq \{\mathbf{1}_N \otimes \theta^\star\} \times \{\mathbf{1}_N \otimes \theta^\star\}, \tag{5.55}$$

which precisely describes the situation where all agent's estimates θ_i are equal to the true parameter θ^*.

Theorem 5.5. Suppose that Assumption 5.6 holds, and let the constants ($\underline{\sigma}_{\mathbf{Q}}$, $\overline{\sigma}_{\mathbf{Q}}$, $\overline{\sigma}^2_{\mathbf{\Omega}}$, $\underline{\sigma}_{\mathbf{\Sigma}}$) be given by Proposition 5.2. If the restart parameters (ω, T_0, T) satisfy $\omega \in (0, 1)$ and

$$\left(\frac{1}{2}\frac{\overline{\sigma}_{\mathbf{Q}}}{\underline{\sigma}_{\mathbf{\Sigma}}} + T_0^2\right)^{\frac{1}{2}} =: \underline{\mathbf{T}} < T < \overline{\mathbf{T}} \coloneqq \left(\frac{\underline{\sigma}_{\mathbf{Q}}(1-\omega)\underline{\sigma}_{\mathbf{\Sigma}}}{\overline{\sigma}^2_{\mathbf{\Omega}} + \chi(k_t)^2}\right)^{\frac{1}{2}}, \tag{5.56}$$

then the following hold:

a) For any restart policy $\eta \in \{0, 1\}^N$ the HDS $\mathcal{H}_c$ renders the set $\mathcal{A}_c$ ISS with respect to the input u.

b) If $\eta_i = 1$ for all $i \in \mathcal{V}$, and $\dot{\tau}_c \coloneqq \omega$, then, for every initial condition $y_0 := y_c(0,0) \in (\mathbf{C}_c \cup \mathbf{D}_c) \times \mathbb{R}_{\geq 0}$, every solution-input pair (y_c, u) of $\mathcal{H}_c$, and every $(t_j, j) \in \mathrm{dom}(y_c)$ with $t_j \coloneqq \min\{t : (t,j) \in \mathrm{dom}(y_c)\}$, the sampled sequence of estimates $\theta(t_j, j)$ satisfies

$$|\theta(t_j, j) - \mathbf{1}_N \otimes \theta^\star|^2 \leq k_1 \cdot \mu^j |y_0|^2_{\mathcal{A}_c} + k_2 |u|^2_{(t_j,j)}, \tag{5.57}$$

where $k_1, k_2 > 0$, and $\mu(T) \coloneqq (\underline{\mathbf{T}}/T)^2 \in (0,1)$.

The main result of Theorem 5.5 reveals the impact of the asymmetry of $\mathcal{G}$ on the resetting parameter T. In particular, the following observations are in order:

(1) When $\mathcal{L}$ is symmetric (i.e., $\overline{\sigma}^2_{\mathbf{\Omega}} = 0$) and the DMCL dynamics do not use real-time data (i.e., $k_t = 0$), condition (5.56) reduces to $\underline{\mathbf{T}} < T < \infty$, which can always be satisfied using *any* positive constant T, recovering the results of [25, Thm. 2] in the context of standard optimization.

(2) In general, the more "informative" is the collective data in the overall system (i.e., the larger is α in (5.35)), the larger the parameter $\underline{\sigma}_{\mathbf{\Sigma}}$ will be, thus providing more flexibility to increase the upper bound T.

(3) The ISS result implies that the trajectories of the algorithm will converge to a neighborhood of the true parameter $\theta^\star$, where the size of the neighborhood shrinks as the disturbances d_i shrink in (5.28). When $d_i = 0$, the result establishes asymptotic convergence to the true parameter.

(4) In item (b), the assignment $\dot{\tau}_c := \omega$ induces periodic resets in the system, where the time between consecutive resets is $(T - T_0)\omega^{-1}$. Moreover, the policy choice $\eta_i = 1$ implies that all agents reset their momentum. In this case, the rate of convergence between reset times is explicitly characterized by a contraction coefficient $\mu(T)$, which can be tuned to improve performance and reduce oscillations. In particular, following similar steps

as in the centralized case [25], the "optimal" value of T that minimizes the contraction coefficient $\mu(T)$ over a given window of time can be computed as $T^* = e\left(\frac{\overline{\sigma}_{\mathbf{Q}}}{2\underline{\sigma}_{\mathbf{\Sigma}}} + T_0^2\right)^{\frac{1}{2}}$. (5) Lastly, when $u \equiv 0$, the convergence bound (5.57) characterizes the "accelerated" convergence properties of $\mathcal{H}_c$ towards the true model $\theta^\star$.

> **Remark 5.12.** The upper bound in (5.56) reveals an interesting trade-off between the choice of T and ω. Specifically, larger values of ω in (5.41) (indicating a more aggressive decreasing damping during flows) leads to more conservative values of T (indicating more frequent resets) to maintain stability.

Next, the following corollary leverages the expression of T^* to obtain convergence bounds that parallel those obtained for centralized single-agent systems [25].

> **Corollary 5.1.** Suppose that all the assumptions of Theorem 5.5 hold with $T = T^*$, $\dot{\tau}_c = \omega$ and $u \equiv 0$; then, (5.57) holds with $k_2 = 0$, and for each $\varepsilon > 0$ we have $|\theta(t_j, j) - \mathbf{1}_N \otimes \theta^\star|^2 \leq \varepsilon$ for all $t_j > t_j^*$, where $t_j^* \coloneqq \frac{1}{2\omega}\left(T^* - T_0\right) \log\left(\frac{1}{\varepsilon}\frac{\overline{c}}{\underline{c}}|y_0|^2_{\mathcal{A}}\right)$.

The bound in Corollary 5.1 implies that, as $T_0 \to 0^+$, the convergence of θ_i towards θ^* is of order $\mathcal{O}\left(e^{-\sqrt{\underline{\sigma}_{\mathbf{\Sigma}}/\overline{\sigma}_{\mathbf{Q}}}}\right)$, for all $i \in \mathcal{V}$. We complete this section with a corollary for the case $\eta = 0$, which guarantees the ISS properties of $\mathcal{H}_c$, but not convergence bounds of the form (5.57).

> **Corollary 5.2.** Suppose that Assumption 5.6 holds, $\eta_i = 0$ for all $i \in \mathcal{V}$, $\omega \in (0, 1)$, and $T_0 < T < \overline{\mathbf{T}}$, with $\overline{\mathbf{T}}$ as defined in (5.56). Then, the HDS $\mathcal{H}_c$ renders the set $\mathcal{A}_c$ ISS.

The resetting bounds of Theorem 5.5 and Corollary 5.2 only provide sufficient conditions for ISS (with exponential convergence rates). It remains an open question how to obtain tight bounds on (T_0, T) that are also *necessary* for stability. We do not further

pursue these questions in this chapter.

5.2.2 Decentralized Momentum-Based Concurrent Learning

Since a central coordinator with state τ_c might not exist in large-scale networks, in this section, we study decentralized restart strategies based on each agent $i \in \mathcal{V}$ implementing an individual dynamic coefficient τ_i with initial conditions $\tau_i(0,0) \in [T_0, T]$, which might not be synchronized *a priori*, namely, $\tau_i(0,0) \neq \tau_j(0,0)$, for some $i, j \in \mathcal{V}$. To simplify our discussion, we assume that $\eta_i = 1$ and $\dot{\tau}_i = \omega \in (0,1)$ for all $i \in \mathcal{V}$, and that $k_t = 0$, which allows us to remove the auxiliary state variable s and its associated dynamics. However, all our results can be extended to the case when time-varying regressors are included.

When each agent implements its own coefficient τ_i, the continuous-time DMCL dynamics (5.41) become

$$(x,u) \in \mathbf{C} \times \mathbb{R}^{N+\bar{k}},\ \dot{x} = \mathbf{F}(x,u) \coloneqq \begin{pmatrix} 2\mathcal{T}^{-1}(p-\theta) \\ -2\mathcal{T}(k_r \mathbf{\Phi}(\theta,u) + k_c \mathbf{L}\theta) \\ \omega \mathbf{1}_N \end{pmatrix}, \tag{5.58}$$

where the main state is now $x = (\theta, p, \tau) \in \mathbb{R}^{Nn} \times \mathbb{R}^{Nn} \times \mathbb{R}^N$, $\mathcal{T} \coloneqq \operatorname{diag}(\tau \otimes \mathbf{1}_n)$, $\tau = (\tau_1, \tau_2, \ldots, \tau_N)$, $\mathbf{\Phi}$ is given by (5.43b), and the flow set is now given by

$$\mathbf{C} \coloneqq \mathbb{R}^N \times \mathbb{R}^N \times [T_0, T]^N. \tag{5.59}$$

In this case, restarts of the form (5.40) with $\eta_i = 1$ occur whenever at least one of the agents satisfies the condition $\tau_i = T$. This behavior can be modeled by the following jump set:

$$\mathbf{D} = \left\{ x \in \mathbf{C} : \max_{i \in \mathcal{V}} \tau_i = T \right\}. \tag{5.60}$$

However, note that this approach would lead to uncoordinated restarts of the individual dynamics of the agents across the system. For example, for any time window $[T_0, T]$, one can select N equidistant initial conditions $\tau_i(0,0) \in [T_0, T]$, where $i \in \{1, 2, \ldots, N\}$, which result in solutions experiencing N restarts during this time window, each restart separated by intervals of flow of length $\frac{T-T_0}{N}$. Therefore, as $N \to \infty$, asynchronous restarts would occur more often, hindering the advantages of incorporating momentum into the flows of the algorithm to accelerate the overall system.

To address this issue, and inspired by the synchronization algorithms of [148], we integrate the restart dynamics (5.40) of each agent with a decentralized coordination mechanism for the states τ_i. Specifically, each agent $i \in \mathcal{V}$ performs individual restarts of the form (5.40) when $\tau_i = T$. However, the agents also implement the following *additional* discrete-time updates whenever their neighbors $j \in \mathcal{N}_i$ satisfy the condition $\tau_j = T$:

$$\tau_i^+ \in \mathcal{R}_i(\tau_i) \coloneqq \begin{cases} T_0 & \text{if } \tau_i \in [T_0, r_i) \\ \{T_0, T\} & \text{if } \tau_i = r_i \\ T & \text{if } \tau_i \in (r_i, T] \end{cases}, \tag{5.61}$$

where $r_i > 0$ is a tunable parameter that partitions the interval $[T_0, T]$ of each agent. Note that in (5.61), the update rule is set-valued whenever $\tau_i = r_i$, and in this case, the parameter τ_i can be updated either as $\tau_i^+ = T_0$ or $\tau_i^+ = T$. By studying this set-valued rule, we can establish suitable robustness properties concerning potential disturbances that might slightly perturb τ_i near the point $r_i \in [T_0, T]$. This ensures that such perturbations will not significantly alter the system's behavior.

To incorporate the additional discrete-time updates (5.61) into the overall jump map of the system, we consider the following set-valued mapping:

$$G_d(x) \coloneqq \Big\{(\hat{\theta}, \hat{p}, \hat{\tau}) \in \mathbb{R}^{(2n+1)N} : \hat{\theta} = \theta,\ \hat{p}_i = p_i,\ \hat{\tau}_i = T_0,$$
$$\tau_j \in \mathcal{R}_j(\tau_j),\ \hat{p}_j = p_j,\ \forall j \in \mathcal{N}_i,$$

$$\hat{p}_k = p_k,\ \hat{\tau}_k = \tau_k\ \forall k \neq i \neq j\Big\},$$

which is defined to be non-empty if and only if $\tau_i = T$ and $\tau_j \in [0, T)$. In words, the mapping $G_d(x)$ captures the resets of the *individual states* $(\theta_i, p_i, \tau_i) \in \mathbb{R}^{2n+1}$ of agent i via (5.40), and also the updates of its neighbors $j \in \mathcal{N}_i$ via (5.61). The overall jump-map of the multi-agent hybrid system can then be defined using the outer-semicontinuous hull of G_d[2], which is denoted as $\overline{G}_d$, leading to the overall discrete-time dynamics

$$(x, u) \in \mathbf{D} \times \mathbb{R}^{N+\bar{k}}, \quad x^+ \in \mathbf{G}(x) \coloneqq \overline{G}_d(x). \tag{5.62}$$

Note that system (5.62) preserves the sparsity property of the graph $\mathcal{G}$.

The decentralized continuous-time dynamics (5.58) and the decentralized discrete-time dynamics (5.62) comprise the overall DMCL algorithm with restarts studied in this chapter. This algorithm is fully modeled by the HDS

$$\mathcal{H} \coloneqq (\mathbf{C}, \mathbf{F}, \mathbf{D}, \mathbf{G}). \tag{5.63}$$

The following theorem provides a decentralized version of Theorem 5.5. In this case, stability of τ is studied with respect to the "synchronized" set $\mathcal{A}_{\text{sync}} \coloneqq ([T_0, T] \cdot \mathbf{1}_N) \cup \{T_0, T\}^N$, and the stability properties of the overall state x are studied with respect to the compact set

$$\mathcal{A} \coloneqq \mathcal{A}_{\theta p} \times \mathcal{A}_{\text{sync}}. \tag{5.64}$$

For simplicity, we state the result for the case $u = 0$, but we also comment on the robustness properties of the dynamics.

[2]The outer-semicontinuous hull of a set-valued mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the unique set-valued mapping $\bar{G} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfying $\text{graph}(\bar{G}) = \text{cl}(\text{graph}(G))$, where $\text{cl}(\cdot)$ stands for the closure.

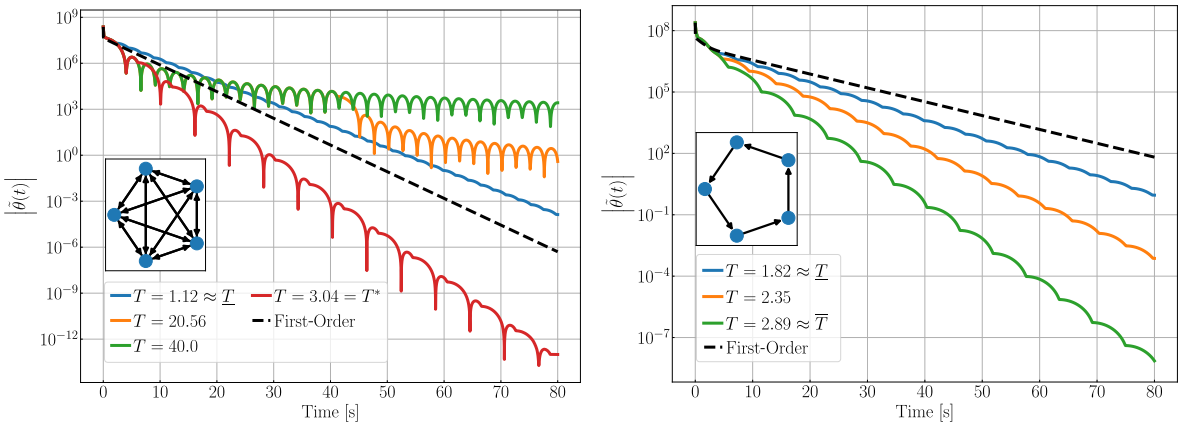


Figure 5.11. Left: Trajectories of $\mathcal{H}$ when $\mathcal{G}$ is fully connected. Right: Trajectories of $\mathcal{H}$ when $\mathcal{G}$ is a cycle. Here, $\tilde{\theta} = \theta - \mathbf{1}_N \otimes \theta^\star$

Theorem 5.6. Consider the HDS $\mathcal{H}$ given by (5.63), and suppose that Assumption 5.6 holds and that:

a) The parameters (T_0, T) satisfy (5.56).

b) The constants $\{r_i\}_{i\in\mathcal{V}}$ satisfy $T_0 < r_i < T_0 + \frac{(T-T_0)}{N-1}$

Then, the set $\mathcal{A} := \mathcal{A}_{\theta p} \times \mathcal{A}_{\text{sync}}$ is UGES for $\mathcal{H}$, and there exists a time $t^* \in \left[0, 2\frac{T-T_0}{\omega}\right)$ such that for every solution x of $\mathcal{H}$ and every $(t, j) \in \text{dom}(y)$ such that $t+j \geq t^*+2N$, the bound (5.57) holds.

Remark 5.13. (Nominal Robustness) Since the hybrid system $\mathcal{H}$ is nominally well-posed in the sense of [33, Ch. 6], the UGES properties of the DMCL algorithm are preserved, in a semi-global practical sense, under arbitrarily small additive perturbations on states and dynamics. This property is crucial for the use of $\mathcal{H}$ in practical applications where dynamic disturbances are unavoidable, such as those in (5.28).

Remark 5.14. (Strong Robustness via ISS) The techniques employed to proof Theorem 5.6 can be further utilized to obtain ISS of $\mathcal{H}$ provided that u originates from a dynamical system evolving in a compact set. We omit this extension due to space constraints.

Remark 5.15. Since system $\mathcal{H}$ has no finite-escape times due to the global Lipschitz property of $\mathbf{F}$ in $\mathbf{C}$, it follows that the stability results of Corollary 5.2 also extend to $\mathcal{H}$ with $\eta_i = 0$ for all i, recovering the convergence result of Theorem 5.5 after an initial finite synchronization phase.

To the best of the author's knowledge, Theorems 5.5-5.6 and the respective corollaries, are the first stability results for momentum-based CL algorithms implemented in multi-agent systems with general directed graphs. We note that in the literature of centralized CL, other accelerated algorithms have been studied using finite-time and fixed-time stability tools in [6, 149, 150]. However, as shown in the comparison presented in [6], when the "level of richness" of the data (i.e., α in Assumption 1) is "low", momentum-based methods can achieve competitive transient performance compared to other first-order non-smooth techniques. For *decentralized* problems defined over networks, we are not aware of finite-time or fixed-time CL algorithms that are stable under Assumption 5.6. A natural progression for future research involves developing such algorithms and comparing them with the DMCL algorithms proposed in this chapter.

5.2.3 Applications in Estimation, Control, and Model-free Feedback Optimization

In this section, we apply the DMCL algorithm with restart in three different applications.

Hybrid Cooperative Identification Over Digraphs

First, we validate Theorem 5.6 in an cooperative estimation problem defined in a multi-agent system with $N = 5$, $n = 3$, and $\psi_i(s) = (10e^{-is} - 1)^2$, for all $i \in \mathcal{V}$. To implement the DMCL algorithm with coordinated restarts, we parameterize $\psi_i(\cdot)$ using the regressor $\phi_i(s) \coloneqq (1, 10e^{-is}, 100e^{-2is})$ and $\theta^\star = (1, -2, 1)$. To satisfy Assumption 5.6 with $\alpha = 5.5$, each agent records five measurements of ψ_i. We implement the hybrid system $\mathcal{H}$ and plot the resulting trajectories of the estimation error in the left plot of Figure 5.11, using $k_r = 80, k_c = 0.08$, and a fully connected graph. We also show with dashed lines the trajectory obtained when using the first-order decentralized CL dynamics of [129]. Since the graph is symmetric, in this case T can be selected arbitrarily large to tune the convergence rate of the dynamics (see inequality (5.56)). The simulations start from a non-synchronized initial condition $\tau(0, 0) \neq \tau_0 \mathbf{1}_5$ and rapidly achieve synchronization. Trajectories related to different choices of T are also shown to illustrate the impact of the restart period on the convergence rate. Next, we let $\mathcal{G}$ be a cycle digraph, for which $\overline{\sigma}^2_{\mathbf{\Omega}} = 0.18$. The resetting parameter T is selected to satisfy inequality (5.56), and the resulting trajectories are shown in the right plot of Figure 5.11. In this case, the best transient performance is obtained as T approaches the upper bound $\overline{\mathbf{T}}$.

Data-Enabled Hybrid Cooperative MRAC

A key advantage of the robust stability results presented in Theorems 1 and 2, is that the DMCL dynamics can be interconnected with other systems for the solution of feedback control problems. To illustrate this application, we consider a multi-agent dynamical system, where each agent has individual dynamics of the form:

$$\dot{\chi}_i = A_i \chi_i + B_i u_i + B_i \tilde{\psi}_i(\theta^\star, \chi_i), \quad \chi_i \in \mathbb{R}^n, \ u_i \in \mathbb{R}^m, \tag{5.65}$$

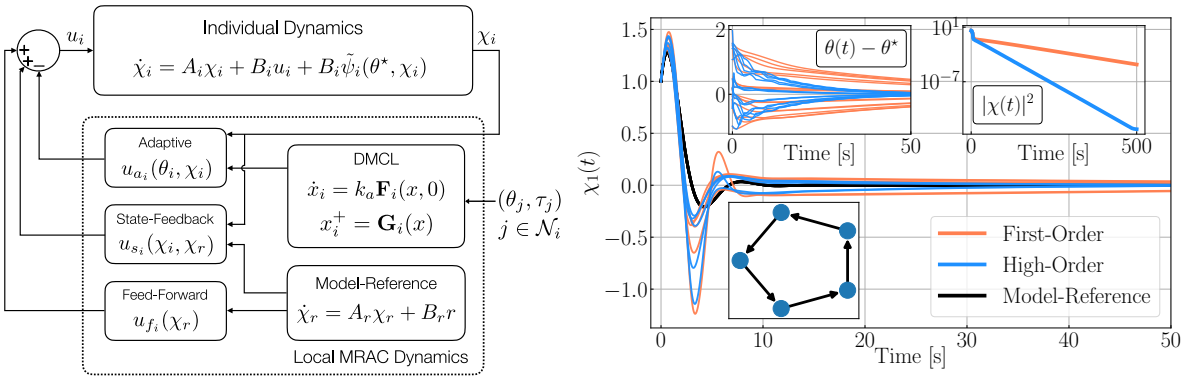


Figure 5.12. Left: Scheme of the i^{th} agent's dynamics in the Cooperative MRAC. $\mathbf{F}_i$ and $\mathbf{G}_i$ represent the components of the overall DMCL flow-map and jump-map, respectively, corresponding to the state $x_i = (\theta_i, p_i, \tau_i) \in \mathbb{R}^{2n+1}$. Right: Trajectories resulting from the Cooperative MRAC when $N = 5$.

where $\tilde{\psi}_i(\theta^\star, \chi) = \phi_i(\chi)^\top \theta^\star$ models structured uncertainty parameterized by a common vector $\theta^\star$, and an individual regressor ϕ_i that is known by each agent i. The agent's goal is to be able to asymptotically track a common bounded reference r despite the uncertainty in their model.

Two-Time Scale Hybrid Dynamics: To solve the tracking problem we use a two-time scale approach. First, we introduce a reference model $\dot{\chi}_r = A_r\chi_r + B_r r$, where A_r is assumed to be Hurwitz. Following the ideas of [41], each agent implements a model-reference adaptive control (MRAC) law that incorporates three elements: (1) an adaptive component $u_{a_i}(\theta_i, \chi_i) = \phi_i(\chi_i)^\top \theta_i$, where θ_i is the individual estimate of $\theta^\star$; (2) a state-feedback component $u_{s_i}(\chi_i, \chi_r) = -K(\chi_i - \chi_r)$; and (3) a feed-forward term u_{f_i} designed such that $B_i u_{f_i}(\chi_r) = (A_r - A_i)\chi_r + B_r r$; see Figure 5.12 for an illustration of the control law. Using $u_i(\theta_i, \chi_i, \chi_r) = u_{s_i}(\chi_i, \chi_r) + u_{f_i}(\chi_r) - u_{a_i}(\theta_i, \chi_i)$, and the error coordinates $e_i = \chi_i - \chi_r$, the error dynamics for agent i become:

$$\dot{e}_i = A_{m_i} e_i + B_i \left(\tilde{\psi}_i(\theta^\star, e_i + \chi_r) - u_{a_i}(\theta_i, e_i + \chi_r) \right), \tag{5.66}$$

where $A_{m_i} \coloneqq A_i - B_i K$, for all i. We make the assumption that system (5.66) has no finite

escape times from all initial conditions, and that Assumption 5.6 holds. To cooperatively estimate θ, we interconnect (5.66) with the DMCL algorith with restart given by (5.63), with flow map

$$x \in \mathbf{C}, \quad \dot{x} = k_a \mathbf{F}(x, 0), \tag{5.67}$$

where the pair $(\mathbf{C}, \mathbf{F})$ is given by (5.58), and where $k_a > 0$ is a tunable parameter.

To study the stability of the interconnected system, we first assume the existence of a centralized timer τ_c that coordinates the resets, with dynamics $\dot{\tau}_c = \omega \in (0, 1)$. We interpret the closed-loop system as a two-time scale hybrid dynamical system with the DMCL algorithm having continuous-time dynamics operating in a faster time scale compared to (5.66). Since A_m is Hurwitz, for each $Q \succ 0$ there exists $P \succ 0$ such that $A_m^\top P + P A_m = -Q$, i.e., system (5.66) is UGES when $\theta_i = \theta$. Similarly, by Theorem 5.5, the momentum-based hybrid dynamics $\mathcal{H}_c$ render the set $\mathcal{A}_c$ UGES via a Lyapunov function V. We can then study the interconnection of both systems using the Lyapunov function $V_1 = 0.5\tilde{V}(e) + 0.5V(x)$, with $\tilde{V}(e) = e^\top P e$, and V constructed as in Section C.2. Indeed, from the proof of Theorem 5.5 in Section C.2, the change of V_1 after a jump satisfies $\Delta V_1 := V_1(e^+, x^+) - V_1(e, x) = \Delta V(x) \leq 0$ because $e^+ = e$. On the other hand, during flows of the closed-loop system, the time derivative of V_1 satisfies

$$\begin{aligned} \dot{V}_1 &= -e^\top Q e - 0.5kV(y_c) + e^\top Q \phi(\chi(t))^\top \theta \\ &\leq -\lambda_{\min}(Q)|e|^2 - k_a |y_c|^2_{\mathcal{A}_c} + k\overline{\phi}|e||y|_{\mathcal{A}}, \end{aligned}$$

where we used the quadratic lower bounds of V, and the boundedness of the regressors to obtain $k\overline{\phi} > 0$. From here, the result follows by completing squares and taking k_a sufficiently large such that $\dot{V} < 0$ using standard arguments for two-time scale systems [22, Ch. 11.5]. Since $\Delta V \leq 0$, the jumps are periodic, and V_1 has quadratic upper and lower bounds, we obtain UGES of the set $\mathcal{A}_1 = \{0\} \times \mathcal{A}_c$, where $\mathcal{A}_c$ is defined right before (5.55).

The stability properties for the decentralized case follow now by leveraging the absence of finite escape times, and by using the reduction principle as in the proof of Theorem 5.6.

Numerical Example

To illustrate the previous result, we consider a multi-agent system with $N = 5$ agents, where the communication graph $\mathcal{G}$ is a directed cycle graph, see the inset in Figure 5.12. We consider open-loop unstable individual dynamics characterized by matrices $A_i = E_{12} \in \mathbb{R}^{2\times 2}, \quad B_i = \left(0, \frac{2i-1}{2i}\right)$, and the parameterized uncertainty $\tilde{\psi}_i(\chi_i) = \phi_i(\chi_i)^\top \theta$, with $\theta = (-1, 1, 0.5)$ and $\phi_i(\chi_i) = (\sin(\chi_{1,i}),\ |\chi_{2,i}|\chi_{2,i},\ e^{\chi_{2,i}\chi_{1,i}})$, for all $i \in \mathcal{V}$. For the MRAC controllers, we consider a second order reference model with natural frequency and damping ratio equal to 1, a state-feedback gain $K = (1, 1)$, and a feed-forward term $u_{f_i}(\chi_r) = -\frac{2i}{2i-1}\left(\mathbf{1}_2^\top \chi_r - r\right)$, for all $i \in \mathcal{V}$. Each agent records two measurements of $\tilde{\psi}_i$ and χ_i at times $t_{i,k} \in \{0, 1.5\}$. The corresponding data matrices Δ_i are not individually rich, which precludes the direct application of standard CL techniques [41] or "one-shot" methods. However, the collective data satisfies the CSR condition in Assumption 5.6 with $\alpha = 0.9$. To regulate the state χ_i to zero, we choose $r = 0,\ k_r = 1,\ k_t = 0,\ k_c = 0.1, k_a = 3$, $T_0 = 0.1$, and $T = 5$. We let each agent implement an MRAC controller interconnected with the hybrid dynamics $\mathcal{H}$ and show the resulting trajectories in Figure 5.12. As observed, the DMCL algorithm with restart yields better transient performance compared to traditional first-order cooperative approaches without momentum [129]. Note that these results are obtained using decentralized *recorded* (i.e., batch) data, as opposed to real-time PE data. The latter might require potentially extreme transient excursions of some states whenever the parameter estimation is accelerated, which is a well-known challenge in real-time adaptive control, see [151].

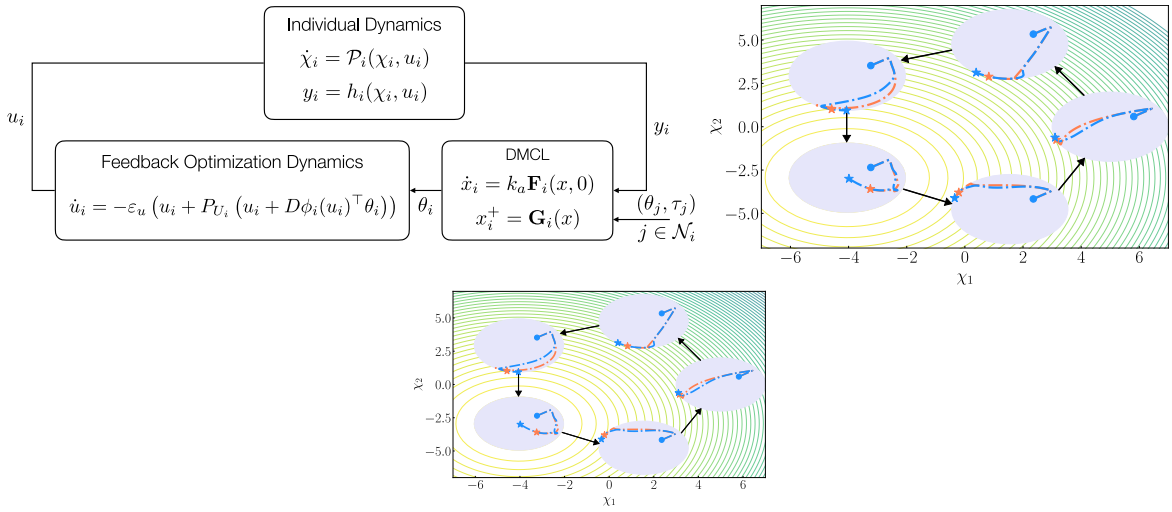


Figure 5.13. Left: Scheme of the i^{th} agent's dynamics in the data-enabled hybrid cooperative feedback optimization dynamics. Right: Trajectories of the vehicles. The arrows represent the edges of $\mathcal{G}$. The final positions of the vehicles are represented by stars.

Data-Enabled Hybrid Cooperative Feedback Optimization

Consider a multi-agent system with dynamics

$$\dot{\chi}_i = \mathcal{P}_i(\chi_i, u_i), \quad y_i = h_i(\chi_i, u_i), \tag{5.68}$$

where $\chi_i \in \mathbb{R}^n$ is the state, $u_i \in U_i \subset \mathbb{R}$ is the input, and $y_i \in \mathbb{R}$ is the output. The set U_i is assumed to be compact and convex for all $i \in \mathcal{V}$. We consider the setting where agents seek to cooperatively find, in real-time and in a model-free manner, an optimal input u^* that maximizes their individual outputs at *steady state*. This scenario describes a classic data-enabled model-free feedback optimization or extremum-seeking problem [129]. To guarantee that this problem is well-posed, the function $\mathcal{P} := \mathcal{P}_1 \times \mathcal{P}_2 \times \ldots \times \mathcal{P}_N$ is assumed to be globally Lipschitz in both arguments, and we also assume there exists a smooth function $u \mapsto m(u) = m_1(u_1) \times m_2(u_2) \times \ldots \times m_N(u_N)$, such that for each fixed $u \in U := U_1 \times U_2 \times \ldots \times U_N \subset \mathbb{R}^N$, the system $\dot{\chi} = \mathcal{P}(\chi, u)$ renders the equilibrium point $\chi^\star = m(u)$ UGES, uniformly on u. Since the function $m(\cdot)$ describes the steady-state

input-to-state mapping of (5.68), the optimization problem that each agent i seeks to solve can be written as

$$\max_{u_i \in U_i} \; J_i(u_i) \coloneqq h_i(m_i(u_i), u_i), \tag{5.69}$$

where the *response maps* J_i are assumed to be unknown, continuously differentiable, strongly concave, common across the network; and parametrizable as $J_i(u_i) = \phi_i(u_i)^\top \theta^\star$, for all $u_i \in U_i$, where ϕ_i is a known continuous and bounded regressor. Functions that satisfy these conditions are common in source seeking problems, where a group of mobile robots seeks to cooperatively find the maximizer of a *common* potential field using intensity measurements, see [129]. In the more general case, we note that, by the universal approximation property of smooth functions, the above assumption on J always holds on compact sets, modulo a small residual error that is also bounded on compact sets. In this case (i.e., non-zero approximation error), our result still holds but now in a "semi-global practical" sense, provided that the bound on the residual approximation error is sufficiently small, a property that can always be achieved by increasing the complexity (i.e., number of basis functions, etc) of the approximator.

Three-Time Scale Hybrid Dynamics: To solve the model-free feedback optimization problem using recorded data that is distributed among the agents, we use a three-time scale approach. Let $u^\star = (u_1^\star, u_2^\star, \ldots, u_N^\star)$ be the vector whose entries are the solutions to the N optimization problems defined in (5.69). To steer u towards u^*, we consider the following feedback optimization dynamics for each agent i:

$$\dot{u}_i = -\varepsilon_u u_i + \varepsilon_u P_{U_i}\left(u_i + D\phi_i(u_i)^\top \theta_i\right), \quad \forall\, i \in \mathcal{V}, \tag{5.70}$$

where $D\phi_i(u_i)$ is the Jacobian matrix of $\phi_i(u_i)$, the function $P_{U_i}(\cdot)$ is the Euclidean projection on the set U_i, $\varepsilon_u > 0$ is a tunable parameter, and θ_i is the individual estimate of $\theta^\star$, which will be recursively updated using the DMCL algorithm with restart, modeled by

the hybrid system $\mathcal{H}$; refer to Figure 5.13 for an illustration of the overall control scheme.

To study the stability properties of the closed-loop system, we modeled the overall dynamics as a three-time scale system, where the plant dynamics (5.68) operate at a faster time scale, the DMCL dynamics with restart operate in a medium time scale, and the optimization dynamics (5.70) operate at the slowest time scale. Such time scale separation can be induced by an appropriate tuning of the gains ε_u in (5.70) and k_a in (5.67). By the stability assumptions on the plant dynamics (5.68), and by using a standard converse Lyapunov theorem [22, Thm. 4.14], there exists a Lyapunov function $V_1 : \mathbb{R}^{nN} \to \mathbb{R}$, and constants $c_i > 0$, for $i \in \{1, 2, 3, 4\}$, such that $c_1|\chi - m(u)|^2 \leq V_1(\chi) \leq c_2|\chi - m(u)|^2$, $\langle \nabla V_1(\chi), P(\chi, u)\rangle \leq -c_3 V_1(\chi)$, and $|\nabla V_1(\chi)| \leq c_4|\chi - m(u)|$ for all $\chi \in \mathbb{R}^n$ and $u \in U$. Similarly, by the proof of Theorem 5.5, and since the HDS $\mathcal{H}$ satisfies the hybrid basic conditions [33, Ch.6], there exists a *quadratic* Lyapunov function V that decreases exponentially fast during flows and jumps of $\mathcal{H}$, provided that the data matrices $\{\Delta_i\}_{i\in\mathcal{V}}$ are CSR. Additionally, since the static-map (5.69) is strongly concave, the optimization dynamics (5.70) with $\theta_i = \theta^\star$ reduced to a projected gradient flow that renders UGES the point $u_i^\star$ via the quadratic Lyapunov function $V_2 = \frac{1}{2}|u_i - u_i^*|^2$, which satisfies $\dot{V}_2 \leq -\gamma_2 V_2$ [152, Thm. 3]. Using these individual quadratic-type Lyapunov functions, and the global Lipschitz properties of the vector fields (5.68), (5.67), and (5.70), we can now use the Lyapunov function $\hat{V} = V + V_1 + V_2$ to establish exponential stability of the closed-loop system by following, recursively, the exact same steps of [22, Ch. 11.5], and using sufficiently small gains ε_u and k_a.

Numerical Example: Consider a multi-vehicle system with $N = 5$ vehicles, seeking to collaboratively locate the source of a potential field that is only accessible via intensity measurements. The vehicles share information via a communication graph $\mathcal{G}$ characterized again by a cycle. We assume the plant dynamics (5.68) have the form $\mathcal{P}_i = A_i\chi_i + B_i u_i$ with matrices $A_i = -iI_2$, $B_i = iI_2$, and quadratic output $y_i = \chi_i^\top Q_i \chi_i + w_i^\top \chi_i + d_i$ where $Q_i = -I_2$, $w_i = (-8.1, -5.88)$, and $d_i = -25$ for all $i \in \mathcal{V}$.

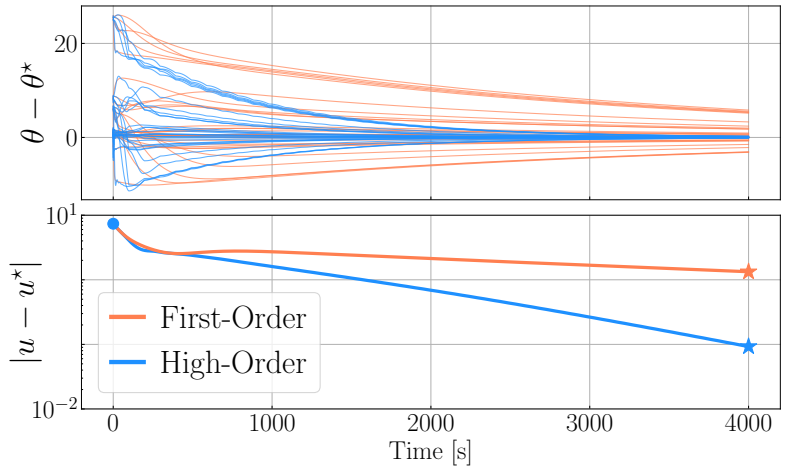


Figure 5.14. Evolution in time of parameter (top) and control (bottom) errors.

The sets U_i are given by $U_i = \xi_i + 2\mathbb{B}$ where $\xi_i = R(2\pi i/N)(1,0)$, with $R(\alpha)$ being the standard 2×2 matrix that rotates a vector by an angle α. In this case, the steady-state input-to-output map (5.69) reduces to $J_i(u_i) = -|u_i|^2 + w_i^\top u_i + d_i$, and each agent uses the vector of basis functions $\phi_i(u_i) = \big(u_{i,1}^2, u_{i,1}, u_{i,2}^2, u_{i,2}, u_{i,1}u_{i,2}, 1\big)$, where the parameter $\theta^\star = (-1, -8.09, -1, -5.88, 0, -25)$ is assumed to be unknown. To implement the DMCL dynamics with restart, each agent has access to only two points of data $\{u_{i,k}, y_{i,k}\}_{k=1}^2$. In this way, while the individual data is not persistently exciting for each agent, the collective data satisfies Assumption 5.6 with $\alpha = 0.75$. Using these data and the parameters $k_r = 1$, $k_t = 0$, $k_c = 0.1$, $k_a = 0.1$, $\varepsilon_u = 0.01$, $T_0 = 0.1$, and $T = 5$, we simulate the closed-loop system comprised of the plant dynamics, the optimization dynamics in (5.70), and the hybrid dynamics $\mathcal{H}$. Figure 5.13 shows the resulting trajectories of the vehicles, converging to the maximizers of J_i in U_i. Figure 5.14 shows the evolution in time of the parameter estimation error and the control signals. It can be observed that, given the low richness of the collected data (small α), the proposed decentralized concurrent learning algorithm with momentum achieves faster convergence compared to the first-order cooperative estimation approach of [129]. In this section, we explored decentralized concurrent learning dynamics with momentum and coordinated resetting for multi-agent systems over directed

graphs. The proposed approach utilizes intermittent coordinated resets to enable collective convergence to a common parameter estimate, even with asymmetric information flow. Using Lyapunov theory for hybrid systems, we established input-to-state stability properties for the momentum-based dynamics, subject to a cooperative richness condition on the data matrices and a topology-dependent lower bound on the resetting frequency. We demonstrated the effectiveness of the proposed dynamics in cooperative adaptive control, showcasing their advantages in accelerated convergence and enhanced transient behavior compared to first-order adaptation algorithms. Future research directions will investigate state-dependent resets and stability results for multi-agent systems with cooperative persistently exciting regressors using averaging theory for hybrid systems following the ideas of [153, Ex. 6.3].

Acknowledgements

Section 5.1, in full, is a reprint of the material as it appears in "Momentum-Based Nash Set-Seeking Over Networks via Multitime Scale Hybrid Dynamic Inclusions" [1]. Section 5.2, in full, in full, is a reprint of the material as it appears in "Decentralized Concurrent Learning with Momentum and Restart" [5]. The dissertation author is the first author of these papers.

Part II

Non-Lipschitz Methods

Chapter 6

Data-Driven Hybrid Dynamics and Nonsmooth ODEs for Accelerated Concurrent Learning

We study efficient algorithms for online parameter estimation problems which can be cast as uncertain linear parametric models of the form

$$y(t) = \phi(t)^\top \theta^*, \tag{6.1}$$

where $y : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is a measurable signal, $\phi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is a uniformly bounded vector-valued regressor function, and $\theta^* \in \mathbb{R}^n$ is an unknown parameter that we want to estimate. This problem plays an important role in different areas, such as adaptive control [154], model-free optimization of dynamical systems [155], and reinforcement learning [156], to name just a few. To achieve online parameter estimation with convergence and robustness certificates, different feedback-based algorithms have been proposed during the last three decades; see [157, 158, 159]. It is well-known that most of the adaptive estimation dynamics that achieve uniform convergence to the true parameter θ^* require a persistence of excitation (PE) condition in the regressor ϕ, of the form

$$\int_t^{t+T} \phi(s)\phi(s)^\top ds \succ \upsilon I, \quad \forall\, t \geq t_0, \text{ where } \; T, \upsilon > 0. \tag{6.2}$$

Indeed, in several adaptive estimation dynamics the PE condition has been shown to be *sufficient* and *necessary* to achieve (uniform) exponential convergence. This includes, the so-called gradient method [160]:

$$\dot{\hat{\theta}} = -\sigma\phi(t)\left(\phi(t)^\top\hat{\theta} - y(t)\right), \qquad \sigma \in \mathbb{R}_{>0}, \tag{6.3}$$

which has been widely used in academic and industrial applications. To relax the PE condition, the works [161], [162], and [163] introduced a class of concurrent learning (CL) adaptive dynamics that incorporate a sequence of recorded data $\{\phi(t_k)\}_{k=1}^N$ that is "sufficiently rich", resulting in a data-driven ordinary differential equation (DD-ODE) of the form

$$\dot{\hat{\theta}} = -\sigma\phi(t)\left(\phi(t)^\top\hat{\theta} - y(t)\right) - \rho\sum_{k=1}^{N}\phi(t_k)\left(\phi(t_k)^\top\hat{\theta} - y(t_k)\right), \tag{6.4}$$

where $\sigma \in \mathbb{R}_{\geq 0}$ and $\rho \in \mathbb{R}_{>0}$ are tunable gains. These types of algorithms have been extended in several directions to develop PE-free adaptive dynamics in the context of model-reference adaptive control [161], reinforcement learning [164], extremum seeking control [155], and general networked estimation problems [146], to name just a few examples. However, by removing (or relaxing) the PE condition, these types of data-driven algorithms can also suffer from poor transient performance in terms of slow rates of convergence, especially when the matrix of recorded data is ill-conditioned. This behavior stems from the fact that systems of the form (6.3) or (6.4) can be cast as *time-varying* gradient flows for which the Hessian matrix might be degenerate whenever the PE condition is relaxed. Indeed, the slow learning rates that may emerge in CL have limited its application in practical engineering problems that require *fast* adaptation and/or estimation.

Motivated by this background, in this report we introduce a novel class of concurrent learning algorithms able to achieve acceleration and/or fixed-time convergence properties. The dynamics make use of different types of regularization mechanisms that have been

explored during the last years to design optimization algorithms and feedback controllers with high-transient performance, but which have never been studied in the context of CL. Given that the proposed dynamics are non-smooth, they are modeled either as non-Lipschitz ODEs [22] or as hybrid dynamical systems [33]. For these systems, we exploit hybrid Lyapunov-based methods to establish suitable stability and convergence properties. Additionally, taking inspiration from machine-learning setups, we study a couple of numerical examples to illustrate the performance of our algorithms. As evidenced by the numerical experiments, the proposed fast algorithms significantly outperform the standard CL dynamics in terms of transient performance and steady state error.

Accelerated Adaptive Concurrent Learning Dynamics

To describe the dynamics considered in this chapter, for the estimation of θ^* in (6.1), let the mappings $A : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^n$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ be defined as

$$A(s, \hat{\theta}) \coloneqq \frac{\phi(s)}{\left(1 + \phi(s)^\top \phi(s)\right)^2} (\phi(s)^\top \hat{\theta} - y(s)), \tag{6.5a}$$

$$\text{and} \quad B(\hat{\theta}) \coloneqq \sum_{k=1}^{N} \frac{\phi(t_k)}{\left(1 + \phi(t_k)^\top \phi(t_k)\right)^2} \Big(\phi(t_k)^\top \hat{\theta} - y(t_k)\Big). \tag{6.5b}$$

Using (6.5) and (6.1), the DD-ODE (6.4) can be written as a time-invariant dynamical system of the form

$$(\hat{\theta}, s) \in \mathbb{R}^n \times \mathbb{R}_{\geq 0}, \qquad \dot{\hat{\theta}} = -\sigma A(s, \hat{\theta}) - \rho B(\hat{\theta}), \quad \dot{s} = 1. \tag{6.6}$$

Taking system (6.6) as a benchmark, we will construct four different data-driven CL dynamics that will achieve (uniform) global asymptotic convergence, exponential convergence, finite-time convergence, and fixed-time convergence, respectively, to the true parameter $\mathcal{A}_0 \coloneqq \{\theta^*\}$. The convergence of these dynamics will depend on the "richness" properties of the available recorded data, a notion that is captured by a finite-time version of persistence

of excitation; see [47].

Assumption 6.1. Let $\{\phi(t_k)\}_{k=1}^N$ be a sequence of recorded data. Then, the matrix

$$\mathcal{D} \coloneqq [\phi(t_1),\ \phi(t_2),\ \cdots,\ \phi(t_N)] \in \mathbb{R}^{n\times N}$$

satisfies $\text{rank}(\mathcal{D}) = n$.

Sequences of data satisfying Assumption 6.1 are said to be *sufficiently-rich* (SR). The following lemma provides an equivalent (and instrumental) characterization of SR data.

Lemma 6.1. Let $\{\phi(t_k)\}_{k=1}^N$ be a sequence of recorded data, and let $P \coloneqq \sum_{k=1}^N \frac{\phi(t_k)\phi(t_k)^\top}{\left(1+\phi(t_k)^\top\phi(t_k)\right)^2}$. Then $\{\phi(t_k)\}_{k=1}^N$ is SR if and only if there exists $\gamma \in \mathbb{R}_{>0}$ such that $P \succeq \gamma I_n$.

We call the constant γ the *level of richness* of the data $\{\phi(t_k)\}_{k=1}^N$. Additionally, for ease of presentation, we introduce the mappings $\Psi : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ and $\Omega : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times n}$

$$\Psi(s) \coloneqq \frac{\phi(s)}{1+\phi(s)^\top\phi(s)},$$
$$\Omega(s) \coloneqq \sigma\Psi(s)\Psi(s)^\top + \rho P,$$

such that $\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta}) = \Omega(s)\tilde{\theta}$, where

$$\tilde{\theta} \coloneqq \hat{\theta} - \theta^*.$$

The following Lemma characterizing bounds for the matrix Ω, will be instrumental for the main stability and convergence proofs of the algorithms. All proofs are presented in Appendix D.

Lemma 6.2. If the data is SR and the regressor vectors are uniformly bounded, then there exist $\kappa \in \mathbb{R}_{>0}$ and $K \in \mathbb{R}_{>0}$ such that

$$\kappa I_n \preceq \Omega(s) \preceq K I_n \qquad \forall s \in \mathbb{R}_{>0}$$

6.1 Data-Driven Accelerated Hybrid Dynamics with Periodic Restarting

The first dynamical system that we consider is inspired by Nesterov's ODEs studied in the context of accelerated optimization; see [19] and [18]. Such algorithms can induce suitable acceleration properties by incorporating dynamic momentum, emulating in continuous time the acceleration properties of Nesterov's accelerated optimization algorithm; see [165]. However, unlike the results of [19] and [18], in the setting of CL we are also interested in establishing suitable robustness properties that are relevant in applications where noisy measurements are unavoidable. Such robustness properties can be obtained by endowing the dynamics with discrete-time restarting mechanisms that persistently reset the momentum coefficient/state of the dynamics. The combination of continuous-time and discrete-time dynamics leads to a hybrid regularization of the Nesterov's discrete-time algorithm which, for time-invariant problems, has been modeled as a HDS in [29], [166], and [4]. Based on this setting, the *hybrid accelerated concurrent learning* (HACL) dynamics that we consider in this chapter are modeled by a HDS with state $x \coloneqq (\hat{\theta}, p, \tau)$, where $\hat{\theta}$ is the estimation state, $p \in \mathbb{R}^n$ is the momentum state, and $\tau \in \mathbb{R}_{>0}$ is a resetting state. The dynamics are given by

$$C \coloneqq \left\{ x \in \mathbb{R}^{2n+2} : \tau \in [T_0, T] \right\}, \qquad \begin{pmatrix} \dot{\hat{\theta}} \\ \dot{p} \\ \dot{\tau} \\ \dot{s} \end{pmatrix} = F(x, s) \coloneqq \begin{pmatrix} \frac{2}{\tau}\left(p - \hat{\theta}\right) \\ -2k\tau\left(\sigma A(s, \hat{\theta}) + \rho B(\hat{\theta})\right) \\ \frac{1}{2} \\ 1 \end{pmatrix}, \tag{6.7a}$$

$$D \coloneqq \Big\{ x \in \mathbb{R}^{2n+2} : \tau = T \Big\}, \quad \begin{pmatrix} \hat{\theta}^+ \\ p^+ \\ \tau^+ \\ s^+ \end{pmatrix} = G(x) \coloneqq \begin{pmatrix} \hat{\theta} \\ (1-q)p + q\hat{\theta} \\ T_0 \\ s \end{pmatrix}, \tag{6.7b}$$

where $k \in \mathbb{R}_{>0}$ is a tunable gain, $\infty > T > T_0 > 0$ are tunable parameters that describe how frequently the algorithm resets, and $q \in \{0, 1\}$ is a Boolean variable that characterizes the resetting policy of the algorithm. In particular, when $q = 0$ the HACL only resets the coefficient τ, whereas when $q = 1$ the algorithm also resets the momentum state p. By construction, the discrete-time updates of the system are periodic and separated by intervals of flow of duration $2(T - T_0)$. To guarantee suitable convergence properties, we will impose the following "data-driven" condition on the parameters (T_0, T) and the gains (k, ρ).

Assumption 6.2. The tunable parameters (T_0, T, k, ρ) satisfy $T_0^2 + \frac{1}{2k\gamma\rho} < T^2 < \frac{8\rho\gamma}{k\sigma^2}$ where $\gamma \in \mathbb{R}_{>0}$ is given by Lemma 6.1.

Remark 6.1. Note that, for all $(T_0, T) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$, there exists $(\rho^*, \sigma^*) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ such that, for all $\rho \geq \rho^*$ and $\sigma \leq \sigma^*$, the condition on the tunable parameters of Assumption 6.2 holds.

The following theorem, which is the first main result of this chapter, characterizes the convergence properties of the HACL dynamics.

Theorem 6.1. Suppose that Assumptions 6.1 and 6.2 hold. Then, every maximal solution of system (6.7) has an unbounded time domain, and the closed set $\mathcal{A} \coloneqq \mathcal{A}_0 \times \mathcal{A}_0 \times [T_0, T] \times \mathbb{R}_{\geq 0}$ is UGAS. Moreover, for each compact set of initial conditions $K \subset C \cup D$, the following convergence properties hold for all $(t, j) \in \mathrm{dom}(x, s)$:

(a) If $q = 0$, then for each $j \in \mathbb{Z}_{\geq 0}$ there exists a monotonically decreasing and

convergent to 0 sequence $\beta_j \in \mathbb{R}_{>0}$, such that each trajectory of the system satisfies the bound

$$|\tilde{\theta}(t,j)|^2 \leq \frac{\beta_j}{k\rho\tau(t,j)^2}. \tag{6.8}$$

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

(b) If $q = 1$, the set $\mathcal{A}$ is UGES, and each trajectory of the system satisfies the bound

$$|\tilde{\theta}(t,j)| \leq k_0 \tilde{\gamma}^j |\tilde{\theta}(0,0)|, \tag{6.9}$$

for all $(t,j) \in \mathrm{dom}(x)$, and where $\tilde{\gamma} = \sqrt{\frac{1}{k\rho T^2}\left(\frac{1}{2\gamma} + k\rho T_0^2\right)} \in (0,1), \quad k_0 = \frac{T}{T_0}$.

The result of Theorem 6.1 establishes two main convergence properties: item (a) establishes that the estimation error decreases at a rate of approximately $O(1/\tau^2)$ during flows, where τ increases linearly with time. On the other hand, item (b) establishes exponential convergence with a convergence rate adjustable via the values of (T_0, T, k, ρ). In this case, information-rich data sets ($\gamma \gg 1$) lead to faster rates of convergence. Optimal restarting periods, similar to those studied in [31] and [25], can also be derived for system (6.7).

6.2 Data-Driven Accelerated Hybrid Dynamics with Adaptive Restarting

The HACL dynamics (6.7) implement a periodic restarting mechanism that is coordinated by the state τ. In this subsection, we now consider an alternative approach based on *adaptive restarting*, where the momentum state is reset whenever a certain state-dependent condition is satisfied. Such type of mechanisms have been studied in the optimization literature; see [19], [31], [167]. However, in the context of CL, these types of mechanisms have remained mostly unexplored. To study this case, we introduce the

function

$$H(\hat{\theta}, p) = \frac{|\tilde{\theta}|_P^2}{2} + \frac{1}{2}|p|^2,$$

where $|\cdot|_P : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a data-induced norm defined as $|u|_P^2 \coloneqq u^\top P u$ for all $u \in \mathbb{R}^n$, with P as defined in Lemma 6.1. We then consider the *Hybrid Hamiltonian Concurrent Learning* (HHCL) algorithm, with $x = (\hat{\theta}, p, \tau) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_{\geq 0}$, and dynamics given by

$$x \in C \coloneqq C_0 \times [0,\ T], \quad \begin{pmatrix} \dot{\hat{\theta}} \\ \dot{p} \\ \dot{\tau} \end{pmatrix} = \begin{pmatrix} 0 & k\rho I_p & 0 \\ -k\rho I_p & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial H}{\partial \hat{\theta}} \\ \frac{\partial H}{\partial p} \\ 1 \end{pmatrix}, \tag{6.10a}$$

$$x \in D \coloneqq (C_0 \times \{T\}) \cup (D_0 \times [0, T]), \quad \begin{pmatrix} \hat{\theta}^+ \\ p^+ \\ \tau^+ \end{pmatrix} = \begin{pmatrix} I_p & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{\theta} \\ p \\ \tau \end{pmatrix}, \tag{6.10b}$$

where $T \coloneqq \frac{n\pi}{2k\rho\sqrt{\gamma}}$, γ is the level of richness of the data, and where

$$C_0 \coloneqq \left\{ (\hat{\theta}, p) \ : \ \langle B(\hat{\theta}), p \rangle \leq 0 \right\},$$
$$D_0 \coloneqq \left\{ (\hat{\theta}, p) \ : \ \langle B(\hat{\theta}), p \rangle = 0 \ \& \ |p|^2 \geq \left| B(\hat{\theta}) \right|^2 / \overline{\lambda} \right\},$$

with $\overline{\lambda} \geq \lambda_{\max}(P)$. Given that $\frac{\partial H}{\partial \hat{\theta}} = B(\hat{\theta})$ and $\frac{\partial H}{\partial p} = p$, the construction of the sets C_0 and D_0 indicate that system (6.10) is allowed to flow whenever there is no increase in the *potential* energy of the data-induced Hamiltonian function H.

Remark 6.2. The role of the timer τ in system (6.10) is to guarantee the existence of an initial reset after an interval of flow of duration $T > 0$. Once this reset has occurred, the update $p^+ = 0$ will guarantee that the next reset of the system will happen before $\tau = T$, i.e., due to the condition $|p|^2 \geq \left| B(\hat{\theta}) \right|^2 / \overline{\lambda}$ for all $x \in D$.

Such types of bounds on the reset times have also been used in [167] for standard optimization problems. However, in the context of CL, their application is new. In particular, note that the parameter T in the jump set D is now data-dependent.

The following theorem is the second main result of this chapter.

Theorem 6.2. Suppose that Assumption 6.1 holds. Then, system (6.10) renders the set $\mathcal{A}_H = \mathcal{A}_0 \times \{0\} \times [0, T]$ UGES, and every solution has an unbounded time-domain satisfies the bound

$$|\tilde{\theta}(t,j)| \leq \sqrt{\frac{2c_0}{\gamma}} \min\left\{1, e^{-\frac{\alpha}{2}(t-\tilde{T})}\right\} |\tilde{\theta}(0,0)| \tag{6.11}$$

for all $(t,j) \in \mathrm{dom}(x)$, where $\alpha = \frac{1}{T}\ln\left(1+\frac{\gamma}{\lambda}\right)$, $\tilde{T} = 2T$ and $c_0 > 0$.

6.3 Finite-Time and Fixed-Time CL Dynamics

While the hybrid CL dynamics (6.7) and (6.10) can induce sublinear and linear convergence rates, the convergence properties of the algorithms are still of *asymptotic* nature, i.e., $\theta(t) \to \theta^*$ only as $t \to \infty$. In this subsection, we consider a different class of learning dynamics able to achieve exact convergence to the true parameter θ^* in a *finite* amount of time. Moreover, in some cases this finite time can be upper bounded by a constant independent of the initial conditions of the estimate $\hat{\theta}$, which leads to *fixed-time* convergence guarantees.

In particular, to achieve finite time convergence we consider the *Finite-Time Concurrent Learning* (FTCL) dynamics modeled by the following non-smooth DD-ODE with $s \in \mathbb{R}_{\geq 0}$:

$$\hat{\theta} \in C \coloneqq \mathbb{R}^n, \qquad \dot{\hat{\theta}} = -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^{\frac{1}{2}}}, \qquad \dot{s} = 1, \tag{6.12}$$

where $(k, \sigma, \rho) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ are tunable gains, where the pair (A, B) is defined in (6.5), and where we let $\dot{\hat{\theta}} := 0$ at $\hat{\theta} = \theta^*$. In order to provide global convergence guarantees, we first proceed to prove that the vector field describing the FTCL dynamics only vanishes when the estimate is equal to the true parameter vector.

Lemma 6.3. If the data is SR, then:

a) $B(\hat{\theta}) = 0 \iff \hat{\theta} = \theta^*$.

b) For any $\sigma \in \mathbb{R}_{\geq 0}$, for the FTCL dynamics we have that $\dot{\hat{\theta}} = 0 \iff \hat{\theta} = \theta^*$.

Now, we show that the defined vector field is everywhere continuous, and hence, that the existence of solutions for the FTCL dynamics is guaranteed. To do so, we first introduce the following Lemma.

Lemma 6.4. If the data is SR, then:

a) $\lim_{\hat{\theta} \to \theta^*} \frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0$ and $\lim_{\hat{\theta} \to \theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \quad \forall a \in (0, 1)$.

b) The vector field defining the FTCL dynamics is everywhere continuous.

For the FTCL dynamics we establish the following stability result, which is the third main contribution of the paper.

Theorem 6.3. Suppose that Assumption 6.1 holds. Then, system (6.12) renders the set $\mathcal{A}_0 \times \mathbb{R}_{\geq 0}$ UGFTS, every solution has an unbounded time domain, and the settling time function satisfies

$$T(\hat{\theta}(0)) \leq \frac{2}{k\gamma\rho} \lambda_{\max}^{1/2}(P) \sqrt{|\hat{\theta}(0) - \theta^*|}.$$

The result of Theorem 6.3 guarantees that $\hat{\theta}(t) = \theta^*$ for all $t \geq T(\hat{\theta}(0))$, where $T(\hat{\theta}(0))$ depends on the initial conditions of the estimate $\hat{\theta}(0)$, as well as the level of

richness γ of the data. To remove the dependence on $\hat{\theta}(0)$ we can further consider a class of *Fixed-Time Concurrent Learning* (FXCL) dynamics, modeled by the following nonsmooth DD-ODE:

$$\hat{\theta} \in C \coloneqq \mathbb{R}^n, \quad \dot{\hat{\theta}} = -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^a} - k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^{-a}}, \quad \dot{s} = 1, \tag{6.13}$$

where $(k, \sigma, \rho) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{>0}$ are tunable gains, $a \in (0,1)$ is a tunable exponent, and where we let $\dot{\hat{\theta}} \coloneqq 0$ at $\hat{\theta} = \theta^*$. Uniqueness of the equilibrium point θ^* for the FXCL dynamics follows a similar argument to the one presented in the proof of Lemma 6.4. Moreover, we note that system (6.13) is not Lipschitz continuous, however, as we next prove, it is everywhere continuous in $\hat{\theta}$, and thus, existence of solutions is guaranteed.

Lemma 6.5. If the data is SR, the vector field defining the FXCL dynamics is everywhere continuous.

The following theorem is the fourth main result of this chapter.

Theorem 6.4. Suppose that Assumption 6.1 holds. Then, system (6.13) renders UGFXS the set $\mathcal{A}_0 \times \mathbb{R}_{\geq 0}$ with $T^* = \frac{\pi}{2a\gamma\rho k}\sqrt{\frac{\lambda_{\max}^a(P)}{\gamma^a}}$, and every solution has an unbounded time-domain.

Remark 6.3. Note that the fixed-time T^* is independent of the initial estimate $\hat{\theta}(0)$, but dependent on the level of richness of the data of the regressor, i.e., dependend on $\gamma > 0$.

The results of Theorems 6.1-6.4 establish new convergence bounds for CL algorithms that explicitly show the dependence on the richness of the data, i.e., the constant γ. In particular, while the standard CL dynamics of [161] achieve exponential convergence with rate of convergence proportional to the level or richness of the matrix P (cf. Lemma 6.1), under suitable tuning of the restarting parameters the data-driven hybrid dynamics

introduced in this chapter can achieve rates of convergence proportional to the squared root of the level of richness of the matrix P, see [24], [167] and [25]. This acceleration property is induced by the addition of momentum to the dynamics, and the design of the flow set and the jump set. Similarly, for the non-smooth DD-ODES (6.12) and (6.13), our results establish finite and fixed-time convergence bounds that are similar to those obtained in [168] for adaptive model-free optimization, but which are new in the context of CL, with an explicit characterization of the convergence time in terms of the level of richness of the data. Finally, note that for applications where γ is small the bound (6.8) establishes a desirable "semi-acceleration" property for estimation problems that lead to convex optimization problems that are not necessarily strongly convex.

Numerical Experiments

In this section, we study numerical examples to illustrate the practical advantages of the proposed algorithms. First, inspired by [161], we consider the problem of estimating the weights of a Radial Basis Function (RBF) Neural Network with one neuron and one bias term:

$$y(t) = 40e^{-\left|\omega(t)-\frac{\pi}{2}\right|^2} + 20.$$

Consequently, for the different CL dynamics we choose the regressor vector

$$\phi(\omega) = \begin{pmatrix} 1 \\ e^{-\left|\omega-\frac{\pi}{2}\right|^2} \end{pmatrix},$$

such that the parameter vector to be estimated is given by $\theta^* = [20, 40]^\top$. In order to gather data we consider the following probing signal

$$\omega(t) = 3\pi \sin(0.1t) - \pi,$$

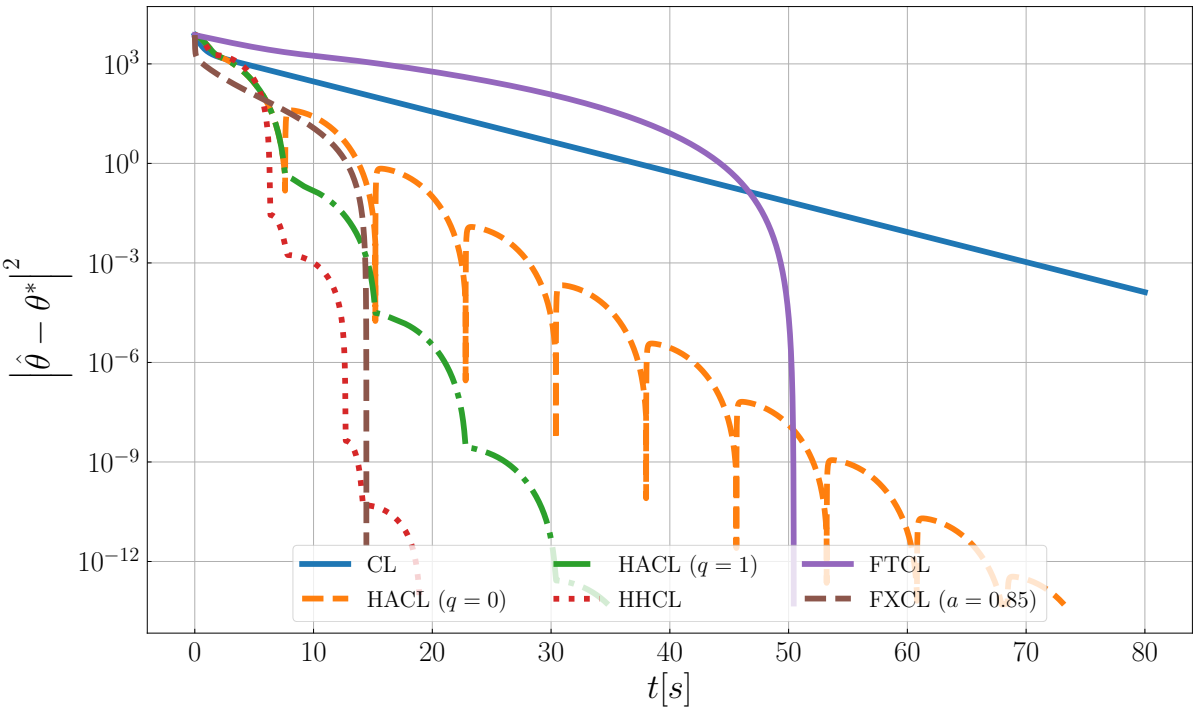


Figure 6.1. Comparison between the proposed algorithms and the DD-ODE (6.6), for the estimation of the weights of a RBF with one neuron and one bias term. All the algorithms are tested with $\sigma = 0,\ \rho = 1,\ k = 1$ and taking the initial estimate $\hat{\theta}(0,0)$ equal to $(-30,-30)$.

and use it to construct a sequence of recorded data $\{\phi(t_k)\}_{k=1}^{N}$ of length $N = 5$. We verify that the selected data is SR by computing the matrix P, as defined in Lemma 6.1, and finding that it has level of richness $\gamma = 0.1044$. Using this data, we compare the proposed algorithms and show the results in Figure 6.1. As depicted in the figure, the introduced algorithms significantly outperform the standard CL dynamics in terms of asymptotic convergence to zero steady-state estimation error. Specifically, we find that both the HHCL and the FXCL trajectories converge to the true parameter in fixed-time, and note that, in spite of rendering the set $\mathcal{A} = \{\theta^*\}$ uniformly-globally-finite-time-stable, the FTCL dynamics can suffer from worse transient performance than the standard CL. Now, we consider the problem of estimating the unknown coefficients of a cost function with known functional form. This setup takes inspiration from settings that arise in model-based machine-learning, where a training phase is carried out offline by sampling information

from a model of the environment that has been learnt beforehand. In particular, we study a cost function that depends on a state $z \in \mathbb{R}^2$ as follows:

$$f(z) = z^\top U z + v z + w,$$

where

$$U = \begin{pmatrix} 6 & 2.5 \\ 2.5 & 4 \end{pmatrix}, \quad v = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad w = 1.$$

Hence, we take as the regressor vector

$$\phi(z) = [z_1^2, z_1 z_2, z_2^2, z_1, z_2, 1]^\top,$$

such that the true parameter vector is given by

$$\theta^* = [6, 5, 4, 3, 2, 1].$$

We take $N = 150$ measurements from the cost function f by making use of the probing signal

$$z(t) = 4 \begin{pmatrix} \sin(0.1t) \\ \cos(0.1t) \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and obtain a matrix P from the recorded data with level of richness $\gamma = 1.011$. Under this setup, we compare our algorithms and show the results in Figure 6.2. As shown in the figure, the FXCL dynamics still outperforms the standard CL dynamics. However, and in contrast to the previous numerical example, where $\gamma < 1$, we find that the difference in performance between the standard CL dynamics and the momentum-based algorithms,

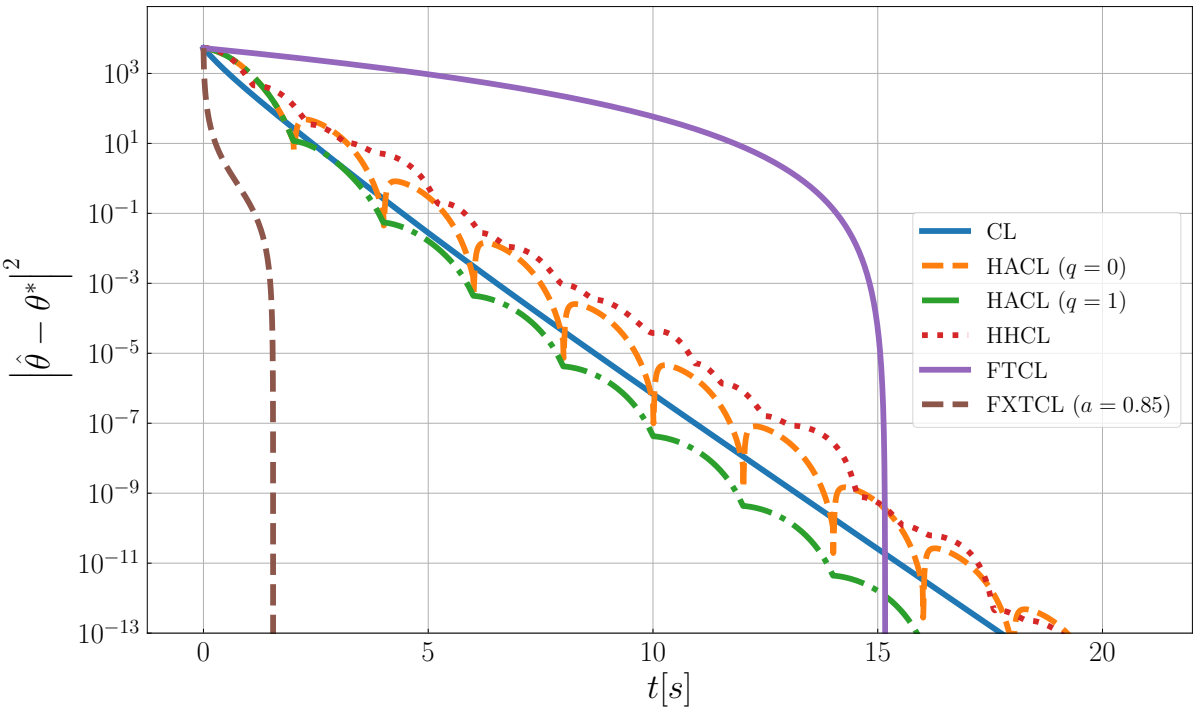


Figure 6.2. Comparison between the proposed algorithms and the DD-ODE (6.6), for the estimation of a cost function. All the algorithms are tested with $\sigma = 0,\ \rho = 1,\ k = 1$ and taking the initial estimate $\hat{\theta}(0,0)$ equal to $) - 30, 30, -30, 30, -30, 30)$.

HACL and HHCL, is reduced. Indeed, since the convergence under the standard CL dynamics depends proportionally on the level of richness γ, while the hybrid dynamics achieve rates proportional to the square root of γ, the benefit of using momentum for faster-convergence is not as evident when γ takes values above 1. However, as suggested by the number of data points used for this case, attaining such *high* level of richness in the recorded data often requires more measurements or a more thorough excitation of the measurable signal, which in practice could be too prohibitive for the implementation of the algorithms.

In this chapter, we introduced a new class of concurrent learning algorithms with acceleration and finite/fixed-time convergence properties. The algorithms are suitable for identification and parameter estimation problems that arise in the context of adaptive control, model-free optimization, and reinforcement learning. The proposed algorithms

are modeled as non-smooth ODEs or hybrid dynamical systems, for which suitable stability, convergence, and robustness properties can be established via Lyapunov-based tools and invariance principles. We illustrated the advantages of the methods via two numerical examples and showed the benefits of using the newly proposed algorithms by contrasting their performance with the standard concurrent learning dynamics. The usage of momentum for accelerated parameter estimation is found to be highly beneficial when the recorded data has low levels of richness; scenario that can arise when only a few data points are available for the algorithms.

Acknowledgements

Chapter 6, in full, is a reprint of the material as it appears in "Accelerated Concurrent Learning Algorithms via Data-Driven Hybrid Dynamics and Nonsmooth ODEs" [6]. The dissertation author is the first author of this paper.

Chapter 7

High-Performance Optimal Incentive-Seeking in Transactive Control for Traffic Congestion

The increase of population density in urban and sub-urban areas has triggered a significant growth of traffic congestion throughout the world, greatly affecting the commute of the public, as well as the expedited delivery of goods. For example, only in 2019, and solely in New York City, the economic losses induced by congestion climbed to $11 billion USD. Moreover, commute times have significantly increased during the last years, forcing drivers to spend, on average, 41 hours per year in congested traffic during morning (6 am to 9 am) and afternoon (3 pm to 6pm) peak travel times [169]. This problem is only expected to worsen during the next years, to the point that by the end of 2022 traffic congestion will cost $74 billion USD to the economy of the United States. To tackle this challenge, cities throughout the world are developing and implementing automated control and optimization algorithms that can guarantee an optimal operation of the transportation infrastructure at all times. Examples include smart traffic light systems [170], dynamic pricing [171, 90], ride-sharing services [172], etc. Among these mechanisms, dynamic pricing has emerged as a promising technology to minimize congestion in dense cities such as London [80, 81], Milan [82], and New York [83]. The goal of dynamic pricing is to induce "optimal" traffic flows that optimize a particular performance measure in the network by adaptively adjusting tolls or incentives [79] based on the current state of the roads. To guarantee that the transportation system continuously operates at its optimal point, pricing

algorithms must react *quickly* to changes in the traffic demand, weather conditions, road accidents, etc. This adaptability requirement has motivated the development of different recursive algorithms for optimal tolling computation, e.g., [84, 85, 86, 87]. Nevertheless, most existing pricing approaches are implemented based on (quasi) static lookup tables instead of real-time feedback traffic measurements, and therefore, under the presence of unexpected accidents or events in the system, are susceptible to generate sub-optimal or even "perverse" tolls that could exacerbate the very problems they were intended to solve [171]. Other recent approaches have relied on socio-technical models that aim to capture decision-making maps of drivers from recorded data; see [173]. In [174], [175] and [176], the authors studied PID controllers to manage the operation of lanes in highway systems. In [177] Hamilton-Jacobi-Bellman equations were solved for the optimal control of high-occupancy toll lanes, and adaptive algorithms based on linear parametrizations and welfare gradient dynamics were studied in [111]. A class of model-based saddle-flow dynamics were also recently studied in [178] in the context of ramp metering control. For a recent review of transactive control for dynamic pricing see [179].

In this chapter, we depart from the traditional model-based approaches studied in the setting of transactive control, and instead, we introduce a new class of *model-free* optimal incentive seeking controllers that can *rapidly* learn optimal incentives (e.g., tolls) using only output measurements from the transportation systems, guaranteeing closed-loop stability at all times. Specifically, motivated by recent advances in non-smooth and hybrid extremum seeking control [25, 180], we introduce three incentive-seeking controllers (ISC) for model-free optimal price seeking in dynamic pricing: a smooth ISC that emulates the performance of a gradient-flow in the slowest time scale; a non-smooth ISC that emulates the behavior of fixed-time gradient flows in the slowest time scale; and a hybrid ISC that leverages momentum to improve transient performance in the slowest time scale. Each of the three controllers are interconnected with the dynamics of the highway, which incorporate socio-technical dynamics as well as traffic flows. Even though the controllers

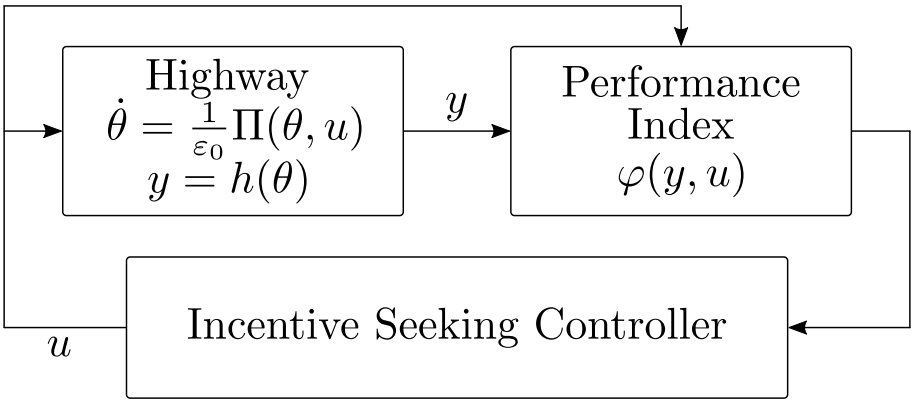


Figure 7.1. A closed-loop interconnection between an Incentive Seeking Controlller (ISC) and a highway system. The ISC will be designed to minimize in rel time a performance function defined by an external supervisor.

are agnostic to the traffic model, we establish practical asymptotic stability results for the resulting closed-loop system, and we numerically show that the non-smooth and hybrid ISCs can significantly outperform the smooth ISC in terms of transient performance under enough time scale separation in the closed-loop system.

Problem Statement

Consider a general highway network system modeled by a dynamical system of the form

$$\dot{\theta} = \frac{1}{\varepsilon_0}\Pi(\theta, u), \qquad y = h(\theta), \tag{7.1}$$

where ε_0 determines the time scale of the dynamics, $\theta \in \mathbb{R}^n$ is the state of the highway, which can include the density of the cars per unit of length in a given lane, $u \in \mathbb{R}^m$ denotes exogenous incentives which can influence the behavior of highway users (e.g., tolls), and $y \in \mathbb{R}^p$ represents measurements that can be obtained from the highway state via the output map $h : \mathbb{R}^n \to \mathbb{R}^p$. Assume that an external supervisor or social planner provides a performance index $\varphi : \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$, which depends on the inputs and outputs of (7.1). Our goal is to design feedback mechanisms able to find in real-time the optimal incentives that minimize the function $\varphi(y, u)$ at steady state. In particular, we consider closed-loop

systems with the structure shown in Figure 7.1, where the Incentive Seeking Controller (ISC) uses only real-time output *measurements* of the performance index. The controller should be designed so that it can find the "optimal" incentives while preserving closed-loop stability at all times. In the following sections, we formalize each of the components illustrated in the scheme of Figure 7.1.

Traffic in Highway Networks: Socio-Tehcnical Models

The performance of transportation systems is not solely dependent on their physical infrastructure, but also on their user behavior [78]. Indeed, in much of the literature that studies the modeling of dynamics in highway networks, the overall structure consists of a socio-technical model that combines a driver behavioral model and a traffic flow model; see [179, 181, 111]. In this work, we follow a similar approach and we assume that the socio-technical and traffic flow models can be lumped together leading to highway network dynamics described by ODEs of the form (7.1). Additionally, we make use of the following regularity assumption.

Assumption 7.1. The map $\Pi(\cdot,\cdot)$ in (7.1) is locally Lipschitz. Moreover, there exists a compact set $\tilde{\Lambda}_\theta \coloneqq \lambda_\theta \mathbb{B} \subset \mathbb{R}^n$ with $\lambda_\theta \in \mathbb{R}_{>0}$, a closed set $\tilde{\Lambda}_u = \hat{\Lambda}_u + \mathbb{B}$ where $\hat{\Lambda}_u \subset \mathbb{R}^m$, and a steady-state map $\ell : \mathbb{R}^m \to \mathbb{R}^n$ that is continuous and locally bounded relative to $\tilde{\Lambda}_u$, such that for each $\eta > 0$ the compact set $\mathbb{M}_\eta \coloneqq \left\{(\theta, u) \ : \ \theta = \ell(u),\ u \in \tilde{\Lambda}_u \cap \eta\mathbb{B}, \theta \in \tilde{\Lambda}_\theta\right\}$ is UAS for the HDS $\mathcal{H}_{ol} \coloneqq (\tilde{\Lambda}_\theta \times (\tilde{\Lambda}_u \cap \eta\mathbb{B}), \varepsilon_0^{-1}\Pi \times \{0\}, \emptyset, \emptyset)$ with state (θ, u).

In words, Assumption 1 guarantees that the highway dynamics are well-posed and stable with respect to external incentives u, and that the steady-state value of the traffic state is parameterized by u via the map ℓ. This assumption is standard (see [111], [87], and [84]), and it is reasonable for many socio-technical models where external incentives u determine the steady state equilibrium of the system.

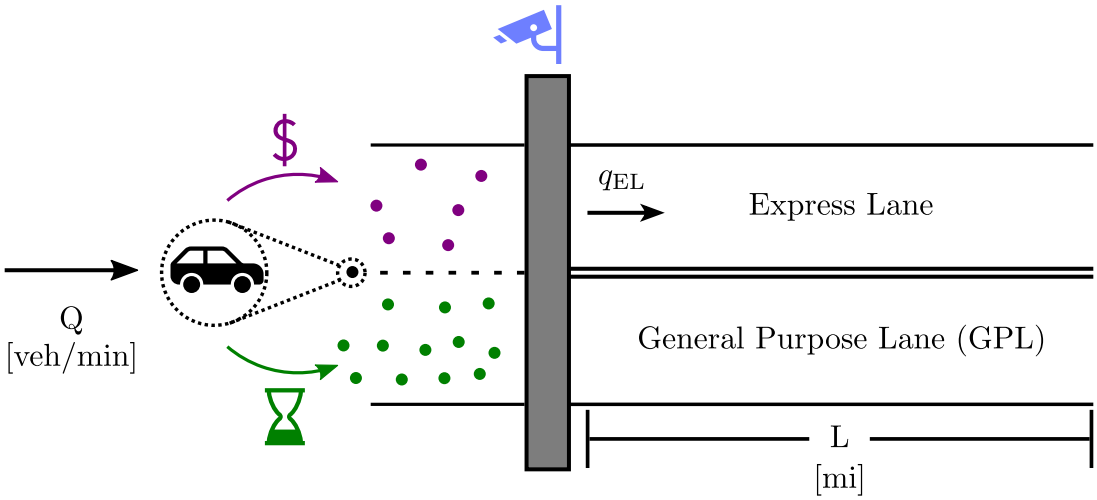


Figure 7.2. Scheme of segment with parallel lanes: the GP lane, and the Express lane.

Socio-Technical Model

To illustrate the advantages of the proposed ISC dynamics, we consider socio-technical models with a similar structure to the one described in [173]. In particular, we study the socio-technical model of a highway segment where drivers can choose between two parallel lanes: the general-purpose (GP) lane, which is uncharged, and the Express lane. Some of the motivations for choosing the Express lane include a faster travel time compared to the GP lane, as well as an expected reduced congestion. Prices (i.e., tolls) or subsidies can be assigned for the utilization of the Express lane depending on the traffic conditions. The model we consider focuses on the description of the average traffic density in the Express lane ρ, and the input flow of vehicles to the Express lane q_{EL}. Figure 7.2 shows a scheme representing the segment with the two parallel lanes: This model is divided into three main components:

a) *The driver behavioral model:* Each driver makes decisions based on the perceived costs of choosing either of the lanes. Some of the possible elements that can be included in these costs are travel times in the lanes, congestion, road conditions, among other quantities estimated by the drivers. To capture these scenarios, we model the

costs by locally Lipschtiz functions $c_{\mathrm{EL}}(q_{\mathrm{EL}}, \rho, u)$ and $c_{\mathrm{GP}}(q_{\mathrm{EL}}, \rho)$ associated with the Express lane and the GP lane respectively, where the input $u \in \mathbb{R}_{\geq 0}$ represents the tolls used to incentivize or to discourage the highway users from using the Express Lane.

Naturally, the response of the individual drivers to the costs extends to the macroscopic level, concurrently affecting the input flow of vehicles to the Express Lane q_{EL}. In general, the drivers require a minimum transient time to adjust to changes in the marginal cost, which for instance could be induced by changes in the tolls. To account for this dynamic response, we model the dynamics of the macroscopic driver behavior as an ODE of the form

$$\dot{q}_{\mathrm{EL}} = \Phi(q_{\mathrm{EL}}, \rho, u), \quad q_{\mathrm{EL}} \in [0, Q], \tag{7.2}$$

where $\Phi : [0, Q] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a locally Lipschitz function that implicitly depends on the marginal cost of choosing the EL. With (7.2), we are able to capture a variety of social dynamics including, among others, evolutionary population dynamics whose stability properties have been recently studied in the literature, c.f. [84].

b) *Equilibrium model:* The equilibrium model describes the resulting average velocity in the Express lane as a function of the average traffic density ρ. In this chapter, we use a mollified version of the average velocity model presented in [173] and described by

$$\overline{v}(\rho) = \frac{v_{\mathrm{free}} - v_{\mathrm{jam}}}{1 + \exp\left(\frac{4}{\rho_{\mathrm{jam}} - \rho_{\mathrm{critical}}}\left(\rho - \frac{\rho_{\mathrm{jam}} + \rho_{\mathrm{critical}}}{2}\right)\right)} + v_{\mathrm{jam}}, \tag{7.3}$$

where $v_{\mathrm{free}}, v_{\mathrm{jam}}$ are constants that represent the top speed and the jam vehicle speed in the Express lane, ρ_{critical} denotes the average density below which the speed of the vehicles is expected to be close to v_{free}, and where ρ_{jam} is the average vehicle density above which a traffic jam occurs in the Express lane. Consequently, these constants

satisfy the relations: $v_{\text{free}} > v_{\text{jam}}$ and $\rho_{\text{critical}} < \rho_{\text{jam}}$.

c) *The traffic flow model:* This model represents the dynamics of the average traffic density ρ of the Express lane, measured in vehicles per unit of length, as a function of the incoming rate of flow q_{EL} and the average velocity of the Express lane $\overline{v}$. It is given by

$$\dot{\rho} = \frac{1}{L}\Big(q_{\text{EL}} - \overline{v}(\rho)\rho\Big), \tag{7.4}$$

where $L \in \mathbb{R}_{>0}$ represents the length of the highway segment under study.

By putting together the driver behavioral model and the traffic flow model, the dynamics of the average density in the express lane can be written in compact form as:

$$\dot{\theta} = \frac{1}{\varepsilon_0}\Pi(\theta, u) \coloneqq \begin{pmatrix} k_m \Phi\left(q_{\text{EL}}, \rho, u\right) \\ k_\rho\Big(q_{\text{EL}} - \overline{v}(\rho)\rho\Big)/L \end{pmatrix}, \quad y = h(\theta) \tag{7.5}$$

where $\theta \coloneqq (q_{\text{EL}}, \rho)$, and $h(\theta) \coloneqq \rho$. The ratio between the constants k_m and k_ρ in (7.5), dictates how fast the driver decisions occur in comparison with the overall traffic flow evolution described by ρ. In some cases, depending on the particular properties of the highway segment and the population of drivers, it might be the case that $k_m/k_\rho \gg 1$. For such scenarios, the relation between the driver response and the associated macroscopic behavior, captured by q_{EL}, can be simplified as a static map that depends on the marginal cost of choosing the Express lane:

$$q_{\text{EL}}(\rho, u) = \lambda(\tilde{c}_{\text{EL}}(\rho, u) - \tilde{c}_{\text{GP}}(\rho))Q, \tag{7.6}$$

where $\lambda : \mathbb{R} \to [0, 1]$ is a locally Lipschitz function that represents the traffic entering into the Express lane as a fraction of the total incoming traffic Q, which we measure in number of vehicles per amount of time, and where $\tilde{c}_{\text{EL}}$ and $\tilde{c}_{\text{GP}}$ are locally Lipschitz costs.

When using relations of the form (7.6), the socio-technical dynamics of (7.5) is simplified as follows:

$$\dot{\rho} = \frac{k_\rho}{L} \left(q_{\mathrm{EL}}(\rho, u) - \overline{v}(\rho)\rho \right), \quad y = \rho. \tag{7.7}$$

Note that (7.5) and (7.7) are particular cases of the ODE in (7.1). For specific realizations of socio-technical models using static and dynamic formulations of the form (7.5) and (7.7), we refer the reader to Section 7.3.

Performance Indices

Depending on the objectives of the social planner, different performance indices can be considered for the purpose of real-time optimization. We will consider families of performance indices that satisfy the following assumption:

> **Assumption 7.2.** Suppose that Assumption 7.1 holds, and let $\tilde{\varphi}(u) := \varphi(h(\ell(u)), u)$. The function $\tilde{\varphi} : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable, strictly convex in $\tilde{\Lambda}_u$, and its gradient is Lipschitz in $\tilde{\Lambda}_u$.

Sometimes, we will also use the following assumption:

> **Assumption 7.3.** There exists $\kappa > 0$ such that $\tilde{\varphi}(\cdot)$ is κ-strongly convex in $\tilde{\Lambda}_u$ and its gradient is Lipschitz in $\tilde{\Lambda}_u$.

The above assumptions will guarantee enough regularity in the incentive-seeking problem, e.g., continuity of the cost and its gradient, the existence of finite optimal incentives, and sufficient monotonicity in the response map of the system. A particular example of a performance index satisfying Assumptions 7.2 and 7.3, and that penalizes the deviation of the current vehicle-density ρ from a desired operation point ρ_{ref} provided by the external supervisor, is given by

$$\varphi_{\mathrm{ref}}(\theta, u) = |\rho - \rho_{\mathrm{ref}}|^2. \tag{7.8}$$

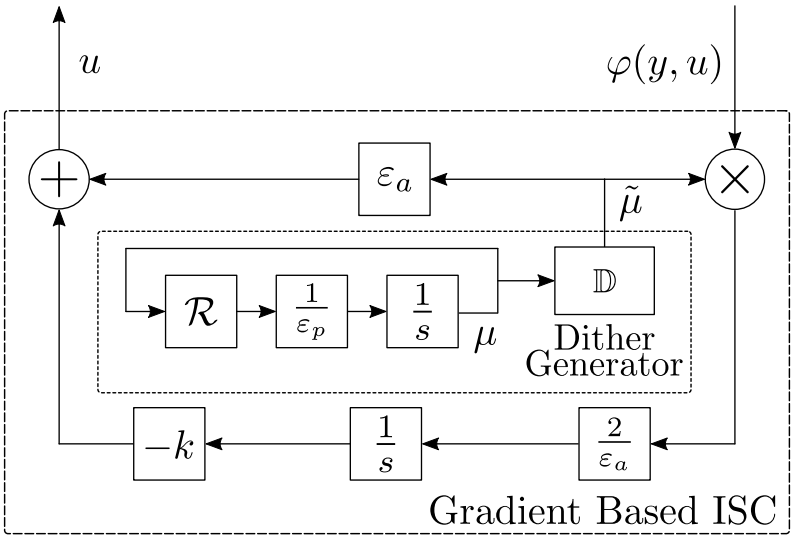


Figure 7.3. Gradient Based Incentive Seeking Control

Performance indices of the form (7.8) can be used by social planners who seek to improve traffic conditions, irrespective of the toll values needed to achieve such end. Among others, performance indices that explicitely depend on the toll prices u can also be considered for situations in which the Express lane manager seeks profit maximization. In most of the cases, we will only require that Assumptions 7.2 or 7.3 hold in a neighborhood of the minimizer of $\tilde{\varphi}$.

Incentive Seeking Feedback Schemes

In this section, we introduce three different ISC algorithms able to guarantee convergence and stability of the set $\mathcal{A}_0 := \{(\theta, u) \in \tilde{\Lambda}_\theta \times \tilde{\Lambda}_u : \theta = \ell(u), u = \arg\min_u \tilde{\varphi}(u)\}$, where Λ_u and $\tilde{\Lambda}_\theta$ are compact sets, ℓ is given in Assumption 7.1, and $\tilde{\varphi}$ is as in Assumption 7.2. The ISCs make use of small exploration signals injected into the transportation dynamics for the purpose of real-time learning. These signals are generated by dynamic oscillators of the form:

$$\dot{\mu} = \frac{1}{\varepsilon_p}\Omega\mu, \qquad \mu \in \mathbb{T}^m, \tag{7.9}$$

where $\varepsilon_p \in \mathbb{R}_{>0}$ is a tunable parameter, and the matrix $\Omega \in \mathbb{R}^{2m\times 2m}$ is a block diagonal matrix with blocks given by $\Omega_i = 2\pi \begin{pmatrix} 0 & \omega_i \\ -\omega_i & 0 \end{pmatrix}$, with $\omega_i \in \mathbb{R}_{>0}$ and for $i = \{1, \cdots, n\}$. We use $\omega \coloneqq (\omega_1, \cdots, \omega_n)$ to denote the vector of frequencies of the signals, and we consider ISCs that generate incentives u of the form

$$u = \hat{u} + \varepsilon_a \hat{\mu}, \tag{7.10}$$

where $\hat{\mu} = (\mu_1, \mu_3, \cdots, \mu_{2n-1})$ represents the odd components of μ, and $\hat{u}$ is the nominal incentive generated by each particular algorithm. We will impose the following assumption on ω.

Assumption 7.4. The dithering frequencies ω_i satisfy: 1) $\omega_i > 0$ is a rational number for all i, and 2) there are no repeated dither frequencies, i.e., $i \neq j \implies \omega_i \neq \omega_j$ and $\omega_i \neq 2\omega_j$.

Assumption (7.4) guarantees orthogonality conditions for the dither signals used by the ISCs to update the incentives u. These conditions will enable real-time learning in the closed-loop system via averaging theory.

7.1 Gradient Based Incentive Seeking Control

We first consider a smooth ISC, denoted GISC, presented in Figure 7.3, which generates the nominal incentive $\hat{u}$ via the following differential equation:

$$\begin{pmatrix} \dot{\hat{u}} \\ \dot{\mu} \end{pmatrix} = F_1(x_1) \coloneqq \begin{pmatrix} -k\varphi(y,u)M(\mu) \\ \frac{1}{\varepsilon_p}\Omega\mu \end{pmatrix}, \; x_1 \in \mathbb{R}^m \times \mathbb{T}^m, \tag{7.11}$$

where $x_1 \coloneqq (\hat{u}, \mu)$, and $M(\mu) = \frac{2}{\varepsilon_a}\hat{\mu}$. The controller (7.11) is based on smooth extremum-seeking controllers [182], which aim to emulate gradient flows whenever the highway

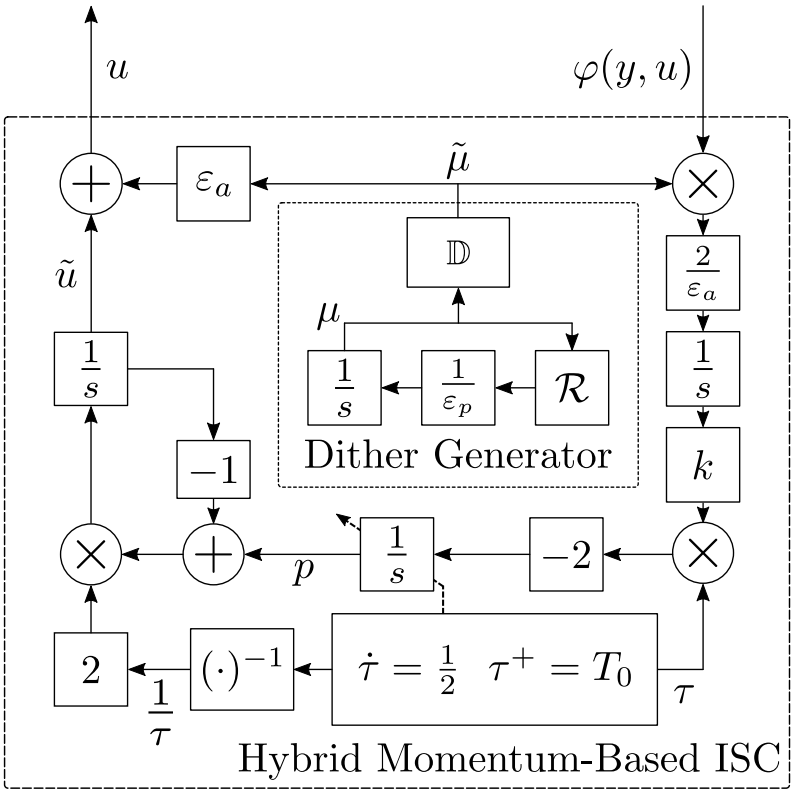


Figure 7.4. Hybrid Momentum Based Incentive Seeking Control

dynamics (7.1) are neglected. The controller makes use of direct measurements of the perfomance index $\varphi(y,u)$, and therefore it is agnostic to the dynamics of the transportation system. In the context of traffic congestion, related dynamics have been studied in [111] for adaptive pricing in affine congestion games, [183] for highways with bottlenecks, and in [184] via simulations for congestion lanes. The following theorem shows that (7.11) is a suitable controller to learn optimal incentives in transportation systems with socio-technical dynamics in the loop. All proofs are presented in Appendix E.

Theorem 7.1. Suppose that Assumptions 7.1, 7.2 (or 7.3) and 7.4 hold. Then, the closed-loop system corresponding to Figure 7.1 with ISC given by (7.11), renders PAS the set $\mathcal{A}_1 \coloneqq \mathcal{A}_0 \times \mathbb{T}^m$ as $(\varepsilon_0, \varepsilon_p, \varepsilon_a) \to 0^+$.

While the ISC (7.11) can achieve optimal incentive seeking, as $(\varepsilon_0, \varepsilon_p, \varepsilon_a) \to 0^+$ the rate of convergence achieved by this controller emulates the convergence rate of a

gradient descent flow, which is either of order $\mathcal{O}(1/t)$ or $\mathcal{O}(e^{-\kappa t})$, where κ is given by Assumption 7.3 (note that Assumption 7.3 implies Assumption 7.2). In Section 5, we will show that the steady-state performance function related to the socio-technical model of the traffic network can have drastically different "slopes" near the optimal point, including cases where $\kappa \ll 1$. These "flat" regions can drastically deteriorate the performance of controllers that seek to emulate traditional gradient flows, e.g., system (7.11). To achieve better transient performance in this scenario, we now consider a class of hybrid dynamics that use momentum.

7.2 Hybrid Momentum-Based Incentive Seeking Control

To achieve better transient performance compared to (7.11), we now consider the hybrid ISC shown in Figure 7.4, which has continuous-time and discrete-time dynamics given by:

$$\begin{pmatrix} \dot{\hat{u}} \\ \dot{p} \\ \dot{\tau} \\ \dot{\mu} \end{pmatrix} = F_2(x_2) \coloneqq \begin{pmatrix} \frac{2}{\tau}\left(p - \hat{u}\right) \\ -2k\tau\varphi(y,u)M(\mu) \\ \frac{1}{2} \\ \frac{1}{\varepsilon_p}\Omega\mu \end{pmatrix}, \tag{7.12a}$$

$$x_2 \in C_2 \coloneqq \Big\{ x_2 \in \mathbb{R}^{2m+1} \times \mathbb{T}^m \ : \ \tau \in [T_0, T] \Big\}, \tag{7.12b}$$

$$\begin{pmatrix} \hat{u}^+ \\ p^+ \\ \tau^+ \\ \mu^+ \end{pmatrix} = G_2(x_2) \coloneqq \begin{pmatrix} \hat{u} \\ \sigma p + (1-\sigma)q \\ T_0 \\ \mu \end{pmatrix}, \tag{7.12c}$$

$$x_2 \in D_2 \coloneqq \Big\{ x \in \mathbb{R}^{2m+1} \times \mathbb{T}^m \ : \ \tau = T \Big\}, \tag{7.12d}$$

where $x_2 \coloneqq (\hat{u}, p, \tau, \mu)$, $k \in \mathbb{R}_{>0}$ is a tunable gain. This controller resets the states p and τ via (7.12c) every time the timer τ satisfies $\tau = T$. The constants $0 < T_0 < T$ are tunable parameters that characterize the frequency of the resets. The parameter $\sigma \in \{0, 1\}$ describes the resetting policy for the state p. Namely, when $\sigma = 1$, we have that $p^+ = p$, while $\sigma = 0$ leads to $p^+ = q$.

In contrast to (7.11), as $(\varepsilon_p, \varepsilon_a, \varepsilon_0) \to 0^+$ the hybrid ISC (7.12) will emulate the behavior of a regularized version of Nesterov's accelerated ODE with momentum [19], given by $\ddot{u} + \frac{3}{t}\dot{u} + \nabla\tilde{\varphi}(u) = 0$, which achieves rates of convergence of order $\mathcal{O}(1/t^2)$ in convex functions, or $\mathcal{O}(e^{-\sqrt{\kappa}t})$ with suitable resets corresponding to $\sigma = 0$ in (7.12c). These resets are similar in spirit to "restarting" techniques used in the literature of machine learning [31]. In the context of model-free feedback control, the resets guarantee enough regularity and robustness in the controller so that it can be interconnected with a dynamical plant in the loop [25]. Thus, the hybrid controller is also able to achieve incentive seeking.

Theorem 7.2. Suppose that Assumptions 7.1, 7.2 (or 7.3) and 7.4 hold. are satisfied. Then, the closed-loop system corresponding to Figure 7.1 with ISC given by (7.12), renders PAS the set $\mathcal{A}_2 \coloneqq \{(\theta, \hat{u}, p, \tau) \ : \ (\theta, \hat{u}) \in \mathcal{A}_0, \ p = \hat{u}, \ \tau \in [T_0, T]\} \times \mathbb{T}^m$ as $(\varepsilon_0, \varepsilon_p, \varepsilon_a) \to 0^+$.

The key advantage of the ISC (7.12) is the incorporation of *dynamic momentum* via the states (p, τ), as well as periodic resets with frequency dependent on the pair (T_0, T). Note that "optimal" restarting frequencies can be used as in [25] to avoid oscillations in the control action induced by the presence of momentum. It is well-known that momentum-based optimization algorithms can significantly improve the transient performance in problems where the cost $\tilde{\varphi}$ exhibits shallow convexity properties (e.g., $\kappa \ll 1$). As shown later in Section 7.3, this will be the case under certain operation conditions of the highway networks.

On the other hand, when the steady state performance function $\tilde{\varphi}$ is strongly convex

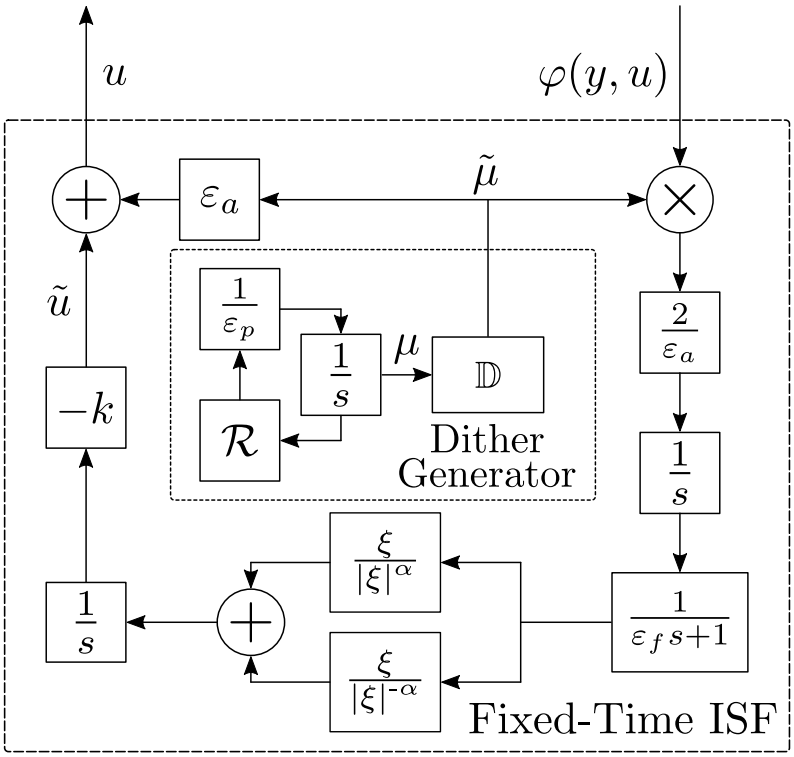


Figure 7.5. Fixed-Time Incentive Seeking Control

and its curvature is not necessarily small, one might wonder if it is possible to achieve better transient performance using non-smooth re-scaled gradient-based dynamics. We investigate this scenario in Section 4.3.

7.3 Fixed-Time Incentive Seeking Control

We now consider the *fixed-time ISC* (FxISC) presented in Figure 7.5 and described by the following dynamics:

$$\begin{pmatrix} \dot{\hat{u}} \\ \dot{\xi} \\ \dot{\mu} \end{pmatrix} = F_3(x_3) \coloneqq \begin{pmatrix} -k\left(\frac{\xi}{|\xi|^{\alpha}} + \frac{\xi}{|\xi|^{-\alpha}}\right) \\ \frac{1}{\varepsilon_f}\left(-\xi + \varphi(y,u)M(\mu)\right) \\ \frac{1}{\varepsilon_p}\Omega\mu \end{pmatrix} \tag{7.13}$$

where $x_3 \coloneqq (\hat{u}, \xi, \mu)$, and $\alpha \in (0,1)$ is a tunable exponent, and where the right hand side of $\dot{u}$ is defined to be zero whenever $\xi = 0$. In this controller, we have incorporated

a low-pass filter with state ξ and gain ε_f^{-1}, and the nominal incentive $\hat{u}$ is generated by a combination of sub-linear and super-linear feedback parametrized by the constant α. Note that the vector field (7.13) is continuous but not Lipschitz continuous at $\xi = 0$. The controller is designed to emulate the performance of fixed-time gradient flows [185, 180] as $(\varepsilon_0, \varepsilon_a, \varepsilon_p, \varepsilon_f) \to 0^+$. This non-smooth ISC also achieves optimal incentive seeking, but it requires (regional) strong convexity of $\tilde{\varphi}$.

Theorem 7.3. Suppose that Assumptions 7.1, 7.3 and 7.4 hold. Then, the closed-loop system corresponding to Figure 7.1 with ISC given by (7.13), renders PAS the set $\mathcal{A}_3 \coloneqq \mathcal{A}_0 \times \{0\} \times \mathbb{T}^m$ as $(\varepsilon_0, \varepsilon_p, \varepsilon_a, \varepsilon_f) \to 0^+$.

In contrast to (7.11) and (7.12), as $(\varepsilon_0, \varepsilon_p, \varepsilon_a, \varepsilon_f) \to 0^+$, the nonsmooth ISC (7.13) emulates the behavior of gradient flows able to converge to the optimal incentive before a fixed time $T^* = \frac{\pi}{2k\alpha\kappa}$, where (α, k) are tunable parameters of the controller, and κ is given by Assumption 7.3. Such type of behavior cannot be obtained using smooth (i.e., Lipschitz continuous) ISCs.

Numerical Examples

In this section, we consider particular realizations of the model introduced in Section 7, as well as numerical examples of the proposed ISCs. Fast Driver Behavior We first consider a scenario where the driver dynamics are qualitatively faster than the average traffic dynamics. Specifically, we borrow the parameters and structure used in [174], based on traffic data of the first dynamic-pricing toll system implemented in the United States: the MnPASS. Thus, the costs of choosing the Express or GP lanes are given by: $c_{\mathrm{EL}}(\rho, u) = a\frac{L}{\overline{v}(\rho)} + bu + \gamma_{\mathrm{EL}}$, and $c_{\mathrm{GP}} = a\frac{L}{\overline{v}(\rho)}\delta + \gamma_{\mathrm{GP}}$, where $\frac{L}{\overline{v}(\rho)}$ represents the estimated travel time on the Express lane, $\gamma_{\mathrm{EL}}, \gamma_{\mathrm{GP}} \in \mathbb{R}$ are offsets used to represent unobservable quantities, a, b are positive weights, and where $\delta \geq 1$ models the fact that the travel time through the GP lane is assumed to be longer or equal than the one of the

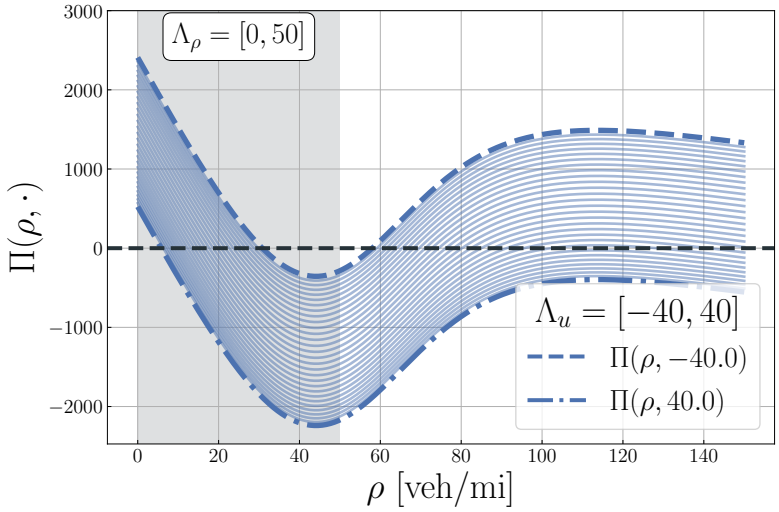


Figure 7.6. Evaluation of viability conditions for highway system with fast driver behavior and static input flow map q_{EL}.

Express lane. On the other hand, the macroscopic driver behavior is assumed to have the form:

$$q_{\text{EL}}(\rho, u) = \frac{Q}{1 + \exp\left(c_{\text{EL}}(\rho, u) - c_{\text{GP}}(\rho)\right)}, \tag{7.14}$$

which is a logistic function of the marginal cost of choosing the Express lane over the GP lane. The choice of static map in (7.14) implies that whenever the perceived cost c_{EL} of choosing the Express lane is lower than the cost c_{GP} of choosing the GP lane, the input flow of vehicles to the Express lane will increase. Moreover, we note that an equal inflow of vehicles to the Express and GP lanes is achieved when the marginal cost is equal to zero. To simulate the ISCs we use the parameters $a = 0.334, b = 0.335, \gamma_{\text{EL}} = 1.71781, \gamma_{\text{GP}} = 0, v_{\text{jam}} = 5[\text{mph}], v_{\text{free}} = 65[\text{mph}], \rho_{\text{jam}} = 80[\text{veh/mi}], \rho_{\text{critical}} = 25[\text{veh/mi}]$ and $L = 0.7[\text{mi}]$. To establish a reference density that guarantees free-flow conditions $\overline{v}(\rho_{\text{ref}}) \approx v_{\text{free}}$ with a moderate occupation of the lane so that the system is not underutilized, we consider the reference seeking performance index φ_{ref} given in (7.8), with $\rho_{\text{ref}} = 0.8\rho_{\text{critical}} = 20$ [veh/mi]. Furthermore, we fix the demand to be $Q = 2170$ vehicles per hour and set $\tilde{\tau} = 3$. Using these parameters, we conduct a numerical study that verifies that Assumptions 7.1-7.3 are satisfied: First, we plot in Figure 7.6 the vector field (7.5) for different values of

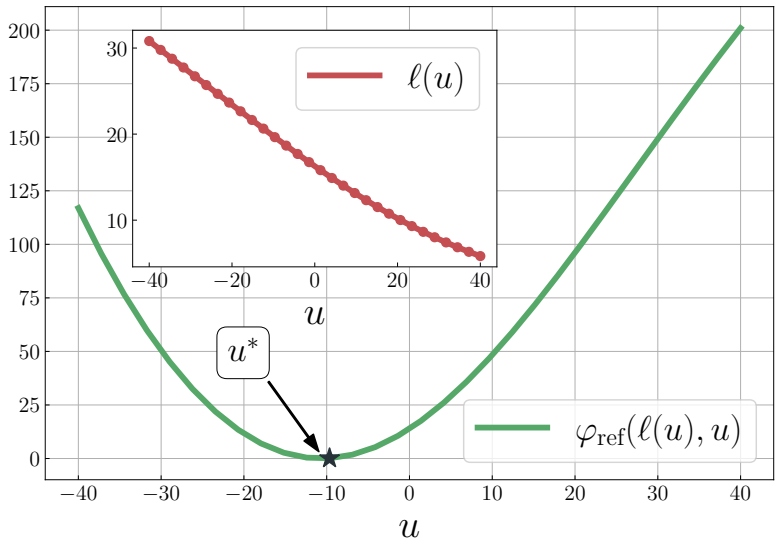


Figure 7.7. Suitability of performance index φ_{ref} and response map $\ell(\cdot)$ for the ISC dynamics presented in this chapter.

$u \in \Lambda_u := [-40, 40]$ and with $\rho = \theta \in [0, 150]$. By arguing graphically, we note that there exists a compact set (interval) $\Lambda_\rho := [0, 50]$, such that restricted to values $(\rho, u) \in \Lambda_\rho \times \Lambda_u$, for each u there exists a unique asymptotically stable equilibrium $\rho^*(u) \in \Lambda$. Therefore, we can define the function $\ell : \Lambda_u \to \mathbb{R}$ as $\ell(u) = \rho^*(u)$, which is shown in Figure 7.7.

Since Π in (7.7) is locally Lipschitz, the previous arguments imply that the socio-technical model of Section 7, with the particular parameters listed above, satisfy the conditions of Assumption 7.1. On the other hand, φ_{ref} satisfies Assumptions 7.2 and 7.3 by construction and the convexity of $\tilde{\varphi}(u) = \varphi(\ell(u), u)$ in Λ_u. Following the closed-loop structure of Figure 7.1, we implement the different ISCs introduced in Section 7 interconnected with the highway dynamics of (7.7). For all the controllers, we set $k = 1$ and use the dithering frequency $\omega = 1$. In the case of the hybrid ISC we choose $\sigma = 0, T_0 = 0.1$ and $T = 20$. For the non-smooth ISC we use $\alpha = 0.5$. We simulate the trajectories of the closed-loop systems using $\varepsilon_f = 1,\ \varepsilon_a = 0.1$, and $\varepsilon_p = 0.01$. These parameters guarantee enough time-scale separation between the different elements of the controller, and also between the controller and the highway-dynamics. We uniformly sample 60 different

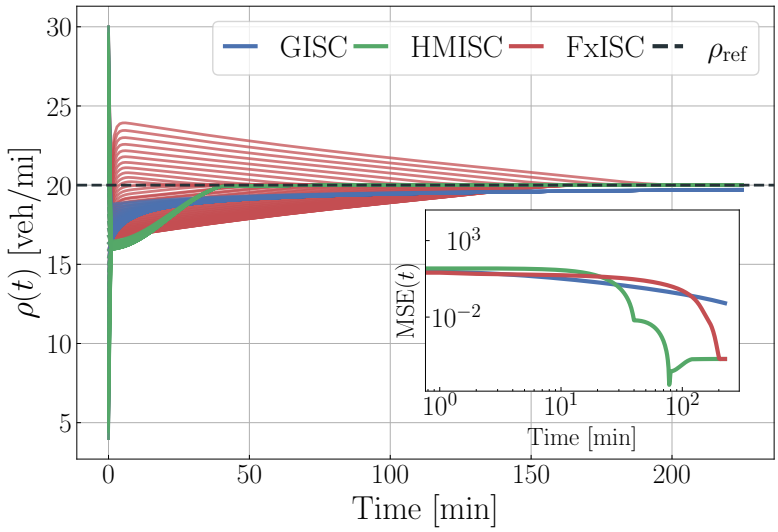


Figure 7.8. Trajectories corresponding to 60 different initial conditions of the average density ρ sampled uniformly between $\rho(0) = 4$ and $\rho(0) = 30$ vehicles per mile. The inset shows the MSE as a function of time of the 60 trajectories.

initial conditions for ρ between 4 and 30. We use the initial incentive $u(0) = 1$, and we plot the resulting trajectories in Figure 7.8. Additionally, we compute and plot the mean squared error (MSE) $\text{MSE}(t) = \frac{1}{60}\sum_{i=1}^{60} |\rho_i(t) - \rho_{\text{ref}}|^2$, where ρ_i corresponds to the trajectory resulting from the i-th initial condition. As shown in the figure, the hybrid and non-smooth ISCs significantly outperform the smooth ISC algorithm (7.11). Note that the hybrid algorithm (7.12) generates the typical oscillatory behavior observed in momentum-based algorithms when the damping is sufficiently small. Note also that the hybrid controller seems to generate better transient performance compared to (7.13), since in certain cases it generates a smaller overshoot. The inset of Figure 7.8 shows that the control signals converge to a small neighborhood of the optimal incentive. Finally, we note that all the ISCs studied in this chapter are well-posed by construction, and therefore they are robust with respect to small bounded additive disturbances acting on the states and dynamics [33, Thm. 7.21]. Moreover, their model-free nature allows them to retain their stability and convergence properties when the parameters of the highway change (slowly) over time. For example, Figure 7.9 shows the impact of variations on the parameter γ_{EL}.

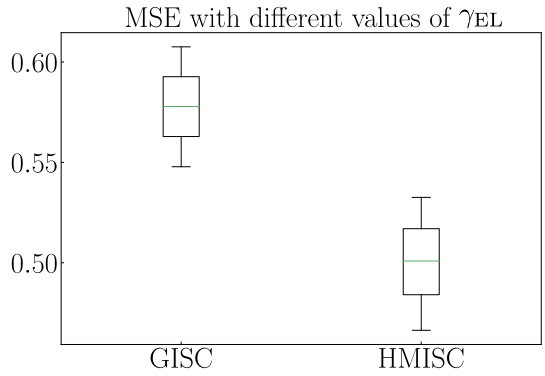


Figure 7.9. Distribution of the mean of the time-averaged MSE for different values of γ_{EL} sampled uniformly within 15% of its nominal value.

Here, we sampled uniformly 20 different values of γ_{EL} with a maximum variation of 15% with respect to the nominal value 1.71781, and we simulated the closed-loop dynamics for each one of the ISCs. For each value of γ_{EL} we computed the mean of the time-average MSE, $\overline{\text{tMSE}} \coloneqq \frac{1}{t_f}\int_0^{t_f} \text{MSE}(\tau)d\tau$, where $t_f = 225[\text{min}]$ is the final time of a simulation run, over 5 trajectories obtained by choosing different initial conditions for ρ on the range $[4, 30]$. As seen in Figure 7.9, the results are consistent with the previous results shown in Figure 7.8. Dynamics for Macroscopic Driver Behavior When the macroscopic driver behavior and the average vehicle density evolve in similar time scales, the dynamics capturing the driver's response due to changes in the incentives need to be considered in the closed-loop system. In this case, we consider the following dynamics to describe the evolution of q_{EL}:

$$\dot{q}_{EL} = \Psi(q_{\text{EL}}, u) \coloneqq -\Big(c_{\text{EL}}(q_{\text{EL}}, u) - c_{\text{GP}}\left(q_{\text{EL}}\right)\Big), \tag{7.15}$$

meaning that the rate of change of the input flow of vehicles to the Express lane is directly determined by the marginal cost of choosing that lane over the GP lane. Consequently, when the perceived cost c_{EL} of choosing the Express lane is lower than the cost c_{GP} of choosing the GP lane, the rate of growth will be instantaneously positive, thus increasing

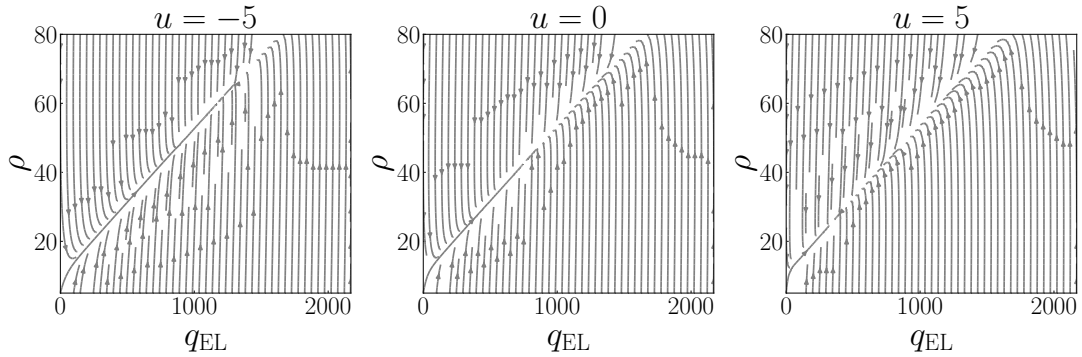


Figure 7.10. Evaluation of the viability conditions for highway system with dynamics describing the macroscopic driver behavior, based on the phase plane of the system.

the input flow of vehicles to the Express lane. We considered the marginal cost

$$c_{\mathrm{EL}}(q_{\mathrm{EL}}, u) - c_{\mathrm{GP}}(q_{\mathrm{EL}}) = \left(q_{\mathrm{EL}} - \frac{Q}{2}\right) + \tilde{a}u,$$

where $\tilde{a} > 0$. For our simulations we use $\tilde{a} = 100$ and the same values of v_{jam}, v_{free}, ρ_{jam}, ρ_{critical}, Q and ρ_{ref} considered in Section 7.3. To study the stability properties of the equilibrium points of the highway dynamics, we analyze the phase planes associated to system (7.5) using $\Phi = \Psi$, $k_m = 1, k_\rho = 1$, and $u \in \Lambda_u = [-40, 40]$, and shown in Figure 7.10 for three particular values of u. In all cases, there exists a compact set $\Lambda_\theta \subset [0, 160] \times [0, Q]$, such that for each u there exists a unique asymptotically stable equilibrium $\theta^*(u) \in \Lambda_\theta$. The same property was numerically confirmed to hold for every $u \in \Lambda_u$ by studying the phase plane plots associated to equally spaced inputs taken from Λ_u, and using continuity of (7.5). Therefore, we can define a response function $\tilde{\ell} : \Lambda_u \to \mathbb{R}^2$ by letting $\tilde{\ell}(u) = \theta^*(u)$. The remaining conditions of Assumption 7.1, and Assumptions 7.2 and 7.3 are verified to hold by following analogous graphical arguments to the ones described in Section 7.3. Indeed, Figure 7.11 shows the corresponding plots describing the response function $\tilde{\ell}$ and the performance index $\varphi_{\mathrm{ref}}\Big(\tilde{\ell}(u), u\Big)$. In this case, we focus our attention on the two ISCs that showed the best performance in Section 7.3, and implement the closed-loop structure of Figure 7.1 using the GISC and HMISC. For both

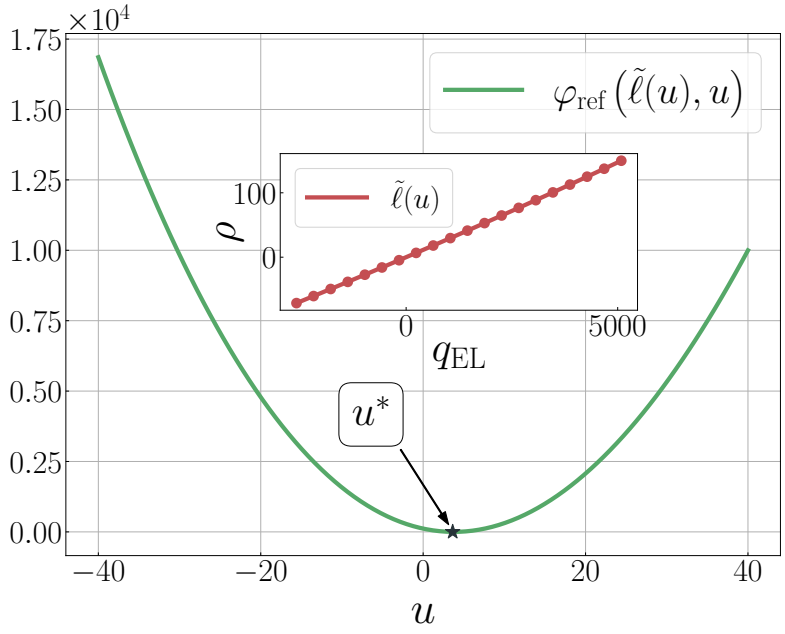


Figure 7.11. Suitability of performance index φ_{ref} and response map $\tilde{\ell}$ for the ISC with highway model including dynamics for the macroscopic driver behavior. The inset shows the response function $\tilde{\ell}$ projected in the phase plane ρ vs. q_{EL} for values of $u \in [-40, 40]$.

controllers we set $\varepsilon_0 = 0.1, \varepsilon_\mu = 0.01, \varepsilon_a = 0.001$, $k = 0.01$, and $\omega = 1$. For the HMISC, we chose $\sigma = 1, T_0 = 0.01$ and $T = 0.5$. The control parameters are selected to guarantee enough time-scale separation between the different elements of the controller and the highway-dynamics. As seen in Figure 7.12, where we plotted the MSE corresponding to 20 different trajectories satisfying $u(0) = 1$ and $q_{\text{EL}}(0) = Q/3$, the HMISC outperforms the GISC, although in a mildly less dominant fashion to what was observed in Section 7.3. This reduction in the performance gap between the GISC and the HMISC, can be mainly attributed to the fact that, in this setup, the strong convexity parameter is relatively high, meaning that the transient performance increase attained via momentum-based dynamics is not as evident as when the costs exhibit shallow convexity properties. Finally, we note that in all our numerical experiments the ISCs were tuned to guarantee that the drivers have enough time to react to changes on the incentives induced by the exploratory signal $\tilde{\mu}$ used by the controllers. This behavior is needed to guarantee real-time learning via

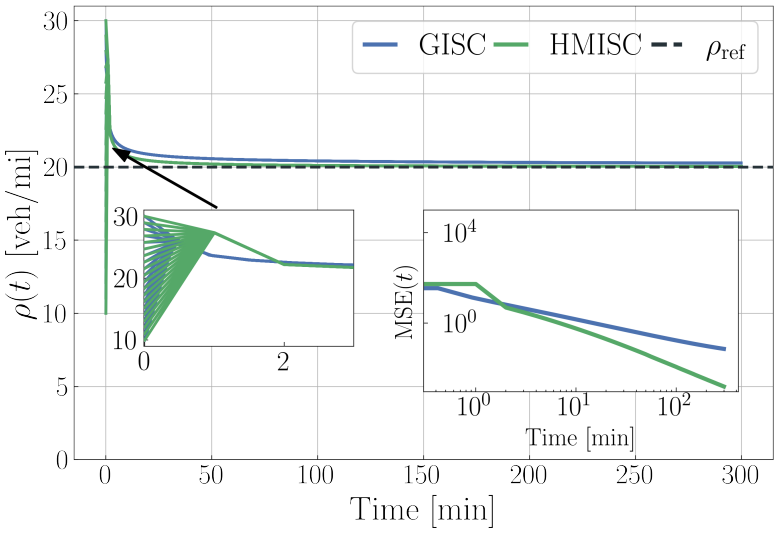


Figure 7.12. Trajectories resulting from the application of the GISC and HMISC to the highway model with dynamics for the macroscopic driver behavior, and corresponding to 20 different initial conditions of the average density ρ sampled uniformly between $\rho(0) = 10$ and $\rho(0) = 30$ vehicles per mile. The inset shows the MSE as a function of time of the 20 trajectories.

feedback measurements of the output of the highway network, and it has also been studied in algorithms based on adaptive control [111] and reinforcement learning [186], to name just a few. Potential extensions that could relax these real-time exploration requirements might be studied in the future by incorporating historical data into the controllers, which can be periodically updated during days or weeks to retain sub-optimality of the incentives. Such controllers will naturally be modeled as hybrid dynamical systems. In this chapter we introduced a new class of incentive-seeking controllers (ISCs) that can learn optimal incentives using only output measurements from traffic in transportation systems, while simultaneously guaranteeing closed-loop stability. We illustrated the benefits of the proposed controllers via numerical experiments in a socio-technical model of a highway system with managed lanes, including the advantages of using nonsmooth and hybrid controllers. The algorithms are agnostic to the exact model of the highway, and robust to small additive disturbances. Future research directions will focus on incorporating past recorded data to minimize real-time exploration in the controllers.

Acknowledgements

Chapter 7, in full, is a reprint of the material as it appears in "High-Performance Optimal Incentive-Seeking in Transactive Control for Traffic Congestion" [7]. The dissertation author was first author of this paper.

Part III

Time-Varying Methods

Chapter 8

Dynamic Gains for Asymptotic-Behavior Shaping in Hybrid Dynamic Inclusions

The ability to control the convergence rate to equilibrium points or optimal solutions is critical across many applications of dynamical systems theory and optimization algorithms. In machine learning, faster convergence can accelerate training of deep neural networks. For embedded control systems, it enables more responsive and robust performance. In computational sciences, it facilitates solving complex simulations and design problems orders of magnitude quicker.

To address the slow asymptotic convergence of traditional methods, significant recent research efforts have explored alternative dynamical systems formulations aimed at achieving faster transient performance. These approaches span momentum-based dynamics [19, 18] that incorporate inertia effects, finite-time stability techniques [187], hyperexponential stability methods [188], and fixed-time stability formulations [189, 190] where the convergence time is uniformly bounded for all initial conditions. Within this line of work, prescribed-time stability has gained increasing attention over the past five years [191, 192, 193, 194, 195, 196]. This form of stability guarantees convergence to the desired target set within a pre-specified time interval, regardless of the initial conditions and parameters of the problem.

The state-of-the-art prescribed-time stability approaches rely on incorporating time-varying or non-Lipschitz vector fields that exhibit finite-time blow-up behavior to

drive the system to equilibrium before the prescribed convergence time. A viewpoint recently explored by the authors for prescribed-time stability of switched systems [9] is to achieve this blow-up effect via a cascade interconnection between an exogenous dynamical system governing the evolution of a dynamic gain and a time-invariant dynamical system with suitable stability properties for the target set. This interconnection technique for inducing pre-established time specifications has been also studied in the context of adaptive control for parameter estimation [197]. However, it has not been comprehensively analyzed for general nonlinear and hybrid dynamical systems combining continuous and discrete behaviors.

This chapter aims to fill this gap by presenting a unifying framework to analytically shape the asymptotic behavior of nonlinear dynamical systems. Our approach hinges on the design of a suitable exosystem that governs the evolution of a dynamic gain. By analyzing the continuous-time deformations that the flow of this system induces on hybrid time domains, we provide sufficient conditions that ensure the original system's stability properties, without the dynamic gain, are transferable under the continuous-time deformation to the full interconnected dynamics. We develop these results by leveraging tools from hybrid dynamical systems theory [33], and formulating an appropriate bijective map that relates the solution sets between the original and cascade systems.

This formulation enables addressing a wide array of time deformations: from simple constant rescaling of time to prescribed-time scalings. Our framework can accommodate systems with hybrid dynamics arising from logic-based switched controllers, state resets, and other discrete-time behaviors coupled with physical dynamics. By providing a unified analysis of the interconnection structures across different classes of dynamic gains and system models this work contributes to the understanding of achievable asymptotic behaviors in dynamical systems.

8.1 Motivational Example

We begin by presenting a motivating example that highlights the key ideas and challenges explored in this chapter. Consider a cost function $f(x) = \frac{x^4}{4}$, and its associated gradient-flow dynamics

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}s} = -\nabla f(\hat{x}) = -\hat{x}^3. \tag{8.1}$$

By the method of separation of variables, the unique solution to (8.1) is given by

$$\hat{x}(s) = \frac{x_0}{\sqrt{2x_0^2 s + 1}} \quad \forall x_0 \in \mathbb{R}.$$

It follows that the value of the cost function along the trajectories of the gradient flow system (8.1) satisfies

$$f(x(s)) = \frac{|x_0|^4}{4\left(2x_0^2 s + 1\right)^2}, \quad \forall\ s \geq 0. \tag{8.2}$$

To improve over the convergence rate of equation (8.2), we consider the case where the gradient flow dynamics in (8.1) are interconnected in a cascade configuration with a dynamic gain μ with dynamics $\frac{\mathrm{d}\mu}{\mathrm{d}t} = 1$. The set of feasible initial conditions is assumed to satisfy $\mu(0) = \mu_0 \in [1, \infty)$, and the interconnected system can be written as the HDS

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\mu \cdot \nabla f(x), \quad \frac{\mathrm{d}\mu}{\mathrm{d}t} = 1, \quad (x, \mu) \in \mathbb{R} \times [1, \infty). \tag{8.3}$$

Since the dynamics of μ are independent of those of the state x, we can first solve for μ to obtain that $\mu(t) = t + \mu_0,\ \mu_0 \in [1, \infty)$. Replacing this result in (8.3) yields

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -(t + \mu_0)x^3.$$

This is a separable ODE, which results in the following solution:

$$x(t) = \frac{x_0}{\sqrt{2x_0^2\left(\frac{t^2}{2} + \mu_0 t\right) + 1}}, \tag{8.4}$$

for all $t \geq 0$, and the corresponding value of the cost function is:

$$f(x(t)) = \frac{|x_0|^4}{4\left(2x_0^2\left(\frac{t^2}{2} + \mu_0 t\right) + 1\right)^2}. \tag{8.5}$$

Note that, the improved convergence bound in (8.5) can be obtained from the bound (8.2) by substituting the time variable s with t and then transforming it under the diffeomorphism $\mathcal{D}_{\mu_0}(t) = \frac{(t+\mu_0)^2}{2} - \frac{\mu_0^2}{2}$. This diffeomorphism satisfies $\frac{d^2}{dt^2}\mathcal{D}_{\mu_0}(t) = 1$, which is the flow map for μ in (8.3).

The above motivating example illustrates that by interconnecting a gradient system with a dynamic gain, the convergence rate can be significantly improved from $\mathcal{O}(1/t^2)$ to $\mathcal{O}(1/t^4)$. However, this improvement was contingent upon obtaining the closed-form solution for the state trajectory under interconnected dynamics. For general nonlinear systems where such closed-form solutions are unavailable, it remains unclear what conditions would permit the application of a similar interconnection procedure to achieve faster convergence rates. Additionally, the example only addressed a continuous-time dynamical system, while many applications involve systems with hybrid dynamics. Consequently, two fundamental questions arise:

1. What are the conditions on the original system's dynamics that allow for the application of the interconnection procedure and enable an improvement in the convergence bounds?

2. Can similar improvements be expected when interconnecting systems with hybrid dynamics?

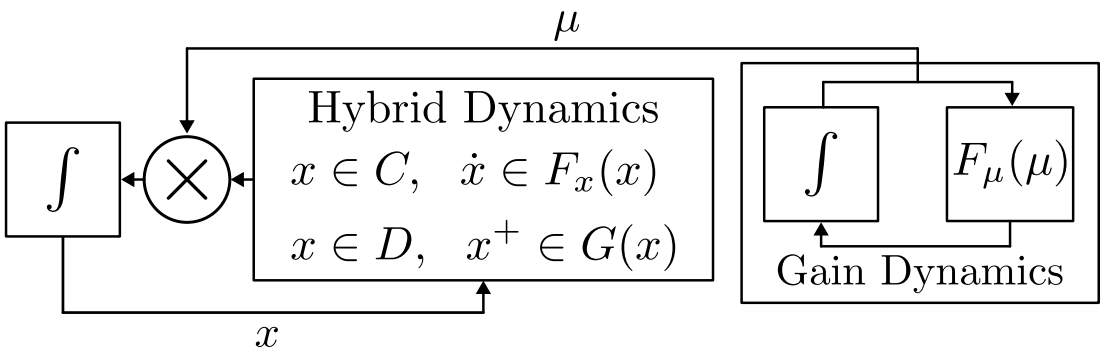


Figure 8.1. Scheme of the cascade interconnection between hybrid dynamics and gain dynamics

To address both questions, we consider a class of dynamical systems arising from the cascade interconnection of a target system exhibiting continuous-time and discrete-time dynamics, and a continuous-time dynamical system that governs the evolution of a dynamic gain. This interconnection idea is shown in Figure 8.1. We model the resulting interconnection as a hybrid dynamical system with state $z \coloneqq (x, \mu) \in \mathbb{R}^n \times \mathcal{X}_\mu$, where $\mathcal{X}_\mu \coloneqq [1, \infty)$. The hybrid dynamics are characterized by the following data:

$$z \in C_x \times \mathcal{X}_\mu, \quad \frac{\mathrm{d}z}{\mathrm{d}t} \in \mu F_x(x) \times \{F_\mu(\mu)\}, \tag{8.6a}$$

$$z \in D_x \times \mathcal{X}_\mu, \quad z^+ \in G(x) \times \{\mu\}, \tag{8.6b}$$

where $C_x, D_x \subseteq \mathbb{R}^n$. We refer to the HDS (8.6) as the *nominal* HDS and denote it with $\mathcal{H}$. The primary objective of this chapter is to certify the stability properties of a suitable closed set $\mathcal{A} \subset \mathbb{R}^n$ under the nominal HDS $\mathcal{H}$, by deriving them from the stability of $\mathcal{A}$ under a *target* HDS $\hat{\mathcal{H}}$. The target system has state $\hat{z} = (\hat{x}, \hat{\mu})$ and is defined by the following data:

$$z \in C_x \times \mathcal{X}_\mu, \quad \frac{\mathrm{d}\hat{z}}{\mathrm{d}s} \in F_x(\hat{x}) \times \left\{ \frac{1}{\hat{\mu}} F_\mu(\hat{\mu}) \right\}, \tag{8.7a}$$

$$z \in D_x \times \mathcal{X}_\mu, \quad \hat{z}^+ \in G(\hat{x}) \times \{\hat{\mu}\}. \tag{8.7b}$$

To obtain our results, the following assumption will play a key role.

Assumption 8.1. a) For every $\hat{\mu}_0 \in \mathcal{X}_\mu$, there exists a unique and complete solution to the *gain ODE* $\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}s} = \frac{1}{\hat{\mu}} F_\mu(\hat{\mu})$.

b) For each $c \in \mathcal{X}_\mu$ there exists an order preserving diffeomorphism $\mathcal{D}_c : \mathcal{T}_c \to \mathbb{R}_{\geq 0}$, where $\mathcal{T}_c \subseteq \mathbb{R}_{\geq 0}$ and $\min \mathcal{T}_c = 0$, satisfying $\mathcal{D}_c(0) = 0$ for all $c \in \mathcal{X}_\mu$ and:

$$\frac{d}{dt}\mathcal{D}_{\mu_0}(t) = \left(\hat{\mu} \circ \mathcal{D}_{\mu_0}\right)(t), \quad \forall t \in \mathcal{T}_{\mu_0},\ \mu_0 \in \mathcal{X}_\mu, \tag{8.8}$$

where $\hat{\mu}$ is the unique solution to the gain ODE.

Assumption 8.1 restricts the type of dynamic gains we consider for our approach to those compatible with suitable deformations of the continuous-time domain, as characterized by the parameterized diffeomorphisms $\mathcal{D}_c$. Alternatively, this constraint allows us to first consider a time-domain specification characterized by a diffeomorphism $\mathcal{D}_c$, such as the accelerated convergence presented in the example of Section 8.1, and then encode it in the dynamical system governing the evolution of the dynamic gain μ. Below, we present several examples illustrating different dynamic gains, their associated flow-maps F_μ, and the respective diffeomorphisms $\mathcal{D}_c$ satisfying Assumption 8.1.

Example 8.1. a) *Linear*: Let, $F_\mu(\mu) = 0$. Then, the diffeomorphism $\mathcal{D}_c(t) \coloneqq ct$, with $c \in \mathcal{X}_\mu$, satisfies the matching equation (8.8) with $\mathcal{T}_c = \mathbb{R}_{\geq 0}$. This type of diffeomorphism is ubiquitous in the singular-perturbation literature where constant parameters are used to induce sufficient time-scale separations in multi-time scale systems, for instance [22, Ch. 11].

b) *Monomial*: Let $F_\mu(\mu) = (p-1)\mu^{\frac{p-2}{p-1}}$, $p > 1$. The unique solution to the gain

ODE $\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}s} = (p-1)\hat{\mu}^{\frac{1}{1-p}},\ \hat{\mu} \in \mathcal{X}_\mu$, is given by

$$\hat{\mu}(s) = \left(\mu_0^{\frac{p}{p-1}} + s \cdot p\right)^{\frac{p-1}{p}},\ \hat{\mu}_0 \in \mathcal{X}_\mu.$$

Then the map $\mathcal{D}_c : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$:

$$\mathcal{D}_c(t) \coloneqq \frac{\left(t + c^{\frac{1}{p-1}}\right)^p - c^{\frac{p}{p-1}}}{p},\ c \in \mathcal{X}_\mu,$$

satisfies the matching equation (8.8) with $\mathcal{T}_c \coloneqq \mathbb{R}_{\geq 0}$.

c) *Exponential*: Let $F_\mu(\mu) = \mu$. The unique solution to the gain ODE $\dot{\hat{u}} = 1,\ \hat{u} \in \mathcal{X}_\mu$ is given by: $\hat{u}(s) = s + u_0$. Then, the map $\mathcal{D}_c : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ defined as

$$\mathcal{D}_c(t) \coloneqq ce^t - c,\ c \in \mathcal{X}_\mu,$$

satisfies the matching equitation (8.8) with $\mathcal{T}_c = \mathbb{R}_{\geq 0}$.

d) *Prescribed-Time*: Let $F_\mu(\mu) = \frac{\mu^2}{\Upsilon},\ \Upsilon > 0$. The unique solution to the gain ODE $\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}s} = \hat{\mu}/\Upsilon$ is given by

$$\hat{\mu}(s) = \hat{\mu}_0 e^{\frac{s}{\Upsilon}}.$$

Then, the map $\mathcal{D}_c : \mathcal{T}_c \coloneqq [0, \Upsilon/c) \to \mathbb{R}_{\geq 0}$ defined as

$$\mathcal{D}_c(t) \coloneqq \Upsilon \ln\left(\frac{\Upsilon}{\Upsilon - ct}\right),\ c \in \mathcal{X}_\mu,$$

satisfies the matching equation (8.8) with $\mathcal{T}_c \coloneqq [0, \Upsilon/c)$. dynamic gains characterized by a finite-escape time $T_{\mu_0} \coloneqq \Upsilon/\mu_0 > 0$ which evolve according to the flow-map $F_\mu(\mu) = \mu^2/\Upsilon$ have recently gained widespread adoption in

the context of prescribed-time regulation of dynamical systems [191]. In this context, the term T_{μ_0} receives the connotation of the *prescribed-time*.

To leverage the deformations of time-domains induced by the dynamic gains with dynamics satisfying Assumption 8.1, we ask for every solution of the target dynamical system to have hybrid time domains with unbounded continuous time.

Assumption 8.2. Every solution $\hat{z} \in \mathcal{S}_{\hat{\mathcal{H}}}$ satisfies $\sup_t \mathrm{dom}(\hat{z}) = \infty$.

Using the above assumptions and exploiting the cascade interconnection structure of the HDS $\mathcal{H}$, we can obtain the following result, whose proof is provided in Appendix F.

Lemma 8.1. Suppose Assumptions 8.1 and 8.2 are satisfied, and let $\mathbb{D}_c \coloneqq \mathcal{D}_c \times \mathrm{id}_{\mathbb{Z}_{\geq 0}}$ for all $c \in \mathcal{X}_\mu$. Then, the map

$$\mathcal{W}: \quad \mathcal{S}_{\hat{\mathcal{H}}} \quad \longrightarrow \quad \mathcal{S}_{\mathcal{H}} \tag{8.9a}$$

$$\hat{z} = (\hat{x}, \hat{\mu}) \longmapsto \hat{z} \circ \mathbb{D}_{\hat{\mu}(0,0)}, \tag{8.9b}$$

is a bijection between maximal solution sets. Moreover, $\mathrm{dom}\left(\mathcal{W}(\hat{z})\right) = \mathbb{D}^{-1}_{\hat{\mu}(0,0)}\left(\mathrm{dom}(\hat{z})\right)$ for every $\hat{z} \in \mathcal{S}_{\hat{\mathcal{H}}}$.

Equipped with the preceding Lemma, we now present our main result, establishing a connection between the stability properties of the target hybrid dynamical system $\hat{\mathcal{H}}$ and the nominal hybrid dynamical system $\mathcal{H}$.

Theorem 8.1. Let $\mathcal{A}_0 \coloneqq \mathcal{A} \times \mathcal{X}_\mu$, with $\mathcal{A} \subset \mathbb{R}^n$ a closed set, and suppose that Assumptions 8.1 and 8.2 are satisfied. Assume that $\mathcal{A}_0$ is UGAS for $\hat{\mathcal{H}}$. Then, for

every $z_0 = (x_0, \mu_0) \in (\overline{C}_x \cup D_x) \times \mathcal{X}_\mu$, and every $z \in \mathcal{S}_{\mathcal{H}}(z_0)$ it follows that

$$|z(t,j)|_{\mathcal{A}_0} \leq \beta(|z_0|_{\mathcal{A}_0}, \mathcal{D}_{\mu_0}(t), j),$$

for all $(t,j) \in \mathrm{dom}(z)$.

Remark 8.1. When the function $\beta \in \mathcal{KL}$ takes an exponential form, and the dynamic gain dynamics follow the structure given in Examples 8.1a)-c), Theorem 8.1 certifies that the nominal system exhibits hyperexponential stability properties in the sense defined by [188]. Conversely, if the gain dynamics are specified as in Example 8.1d), then Theorem 8.1 recovers the *prescribed-time stability results via flows* studied in [9] for hybrid systems, and in [198] for ODEs.

8.2 Applications

This section showcases Theorem 8.1 through two optimization examples. For each example, we establish the stability properties of a compact set under a target hybrid dynamical system. Then we leverage the results of the theorem to certify the stability of the same set for the nominal dynamics.

8.2.1 Acceleration of Gradient Flows

Let $f \in C^1(\mathbb{R}^n)$ be a strictly convex and radially unbounded function. Consider the nominal HDS denoted as $\mathcal{H}_g$, with continuous-time dynamics

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\mu \nabla f(x), \quad \frac{\mathrm{d}\mu}{\mathrm{d}t} = F_\mu(\mu), \ (x,\mu) \in \mathbb{R}^n \times \mathcal{X}_\mu, \tag{8.10}$$

where we suppose that the map F_μ satisfies Assumption 8.1, and with no discrete-time evolution, i.e., with $D = \emptyset$.

We establish the stability of the compact set $\mathcal{A} \coloneqq \{x^\star\}$ for $\mathcal{H}_g$, where $x^\star$ is the unique minimizer of f. We begin by considering the target HDS $\hat{\mathcal{H}}_g$ with continuous-time evolution:

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}s} = -\nabla f(\hat{x}), \quad \frac{\mathrm{d}\hat{\mu}}{\mathrm{d}s} = \frac{1}{\hat{\mu}} F_\mu(\hat{\mu}), \quad \hat{z} = (\hat{x}, \hat{\mu}) \in \mathbb{R}^n \times \mathcal{X}_\mu,$$

and no discrete-time dynamics. We assume that the ODE $\frac{\mathrm{d}\hat{\mu}}{\mathrm{d}s}$ satisfies Assumption 8.1. Next, we augment the system with a timer state $\hat{\tau}$ which has dynamics $\frac{\mathrm{d}\hat{\tau}}{\mathrm{d}s} = 1$, $\hat{\tau} \in \mathbb{R}_{\geq 0}$. We denote the resulting HDS as $\hat{\mathcal{H}}_\tau$, and study the stability properties of the set $\mathcal{A}_0 \times \mathbb{R}_{\geq 0},\ \mathcal{A}_0 \coloneqq \mathcal{A} \times \mathcal{X}_\mu$. To this end, consider the candidate Lyapunov function

$$V(\hat{z}, \hat{\tau}) = \hat{\tau}(f(\hat{x}) - f^\star) + \frac{1}{2}|\hat{x}|^2_{\mathcal{A}},$$

where $f^\star \coloneqq f(x^\star)$. Note that $V(\hat{z}, \hat{\tau}) \geq 0$ for all $(\hat{z}, \hat{\tau}) \in \mathbb{R}^n \times \mathcal{X}_\mu \times \mathbb{R}_{\geq 0}$. Additionally, the continuous-time derivative of V along the trajectories of $\hat{\mathcal{H}}_\tau$ is given by:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s} V(\hat{z}, \hat{\tau}) &= \langle \hat{\tau}\nabla f(\hat{x}) + (\hat{x} - x^\star),\ -\nabla f(\hat{x}) \rangle \\
&\quad + \langle f(\hat{x}) - f^\star,\ 1 \rangle \\
&= -\big(f^\star - [f(\hat{x}) + \langle \nabla f(\hat{x}), x^\star - \hat{x} \rangle]\big) \\
&\quad - \hat{\tau}|\nabla f(\hat{x})|^2. \qquad (8.11)
\end{aligned}$$

Letting $d_f(x^\star, x) \coloneqq f^\star - [f(\hat{x}) + \langle \nabla f(\hat{x}), x^\star - \hat{x} \rangle]$, by the strict convexity of f, it follows that $d_f(x^\star, x) > 0$ for all $\hat{x} \in \mathbb{R}^n \setminus \mathcal{A}$, and that $d_f(x^\star, x) = 0$ and $\nabla f(\hat{x}) = 0$ if and only if $\hat{x} \in \mathcal{A}$. From (8.11), we obtain that there exists $\rho_c \in \mathcal{PD}$ such that:

$$\frac{\mathrm{d}}{\mathrm{d}s} V(\hat{z}, \hat{\tau}) = -\rho_c\left(|\hat{x}|_{\mathcal{A}}\right) = -\rho_c(|(\hat{z}, \hat{\tau})|_{\mathcal{A}_0 \times \mathbb{R}_{\geq 0}}), \qquad (8.12)$$

for all $(\hat{z}, \hat{\tau}) \in \mathbb{R}^n \times \mathcal{X}_\mu \times \mathbb{R}_{\geq 0}$, and where in the last equality we have used the fact that $|(\hat{\mu}, \hat{\tau})|_{\mathcal{X}_\mu \times \mathbb{R}_{\geq 0}} = 0$ for all $(\hat{\mu}, \hat{\tau}) \in \mathcal{X}_\mu \times \mathbb{R}_{\geq 0}$. By [199, Theorem 3.19.3a)], this implies that there exists $\beta_g \in \mathcal{KL}$ such that $\mathcal{A}_0 \times \mathbb{R}_{\geq 0}$ is β_g-UGAS for $\mathcal{H}_{\hat{\tau}}$. Additionally, since for every solution $(\hat{z}, \hat{\tau}) \in \mathcal{S}_{\hat{\mathcal{H}}_\tau}$ it follows that $|\hat{\tau}(s)|_{\mathbb{R}_{\geq 0}} = 0$ for all $s \in \mathrm{dom}(\hat{z}, \hat{\tau}) = \mathbb{R}_{\geq 0}$, we obtain that $\mathcal{A}_0$ is β_g-UGAS for $\hat{\mathcal{H}}_g$. Furthermore, for any such solution $(\hat{z}, \hat{\tau})$, by using (8.12), we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}s} V(\hat{z}(s), \hat{\tau}(s)) = -\rho_c \left(|\hat{x}(s)|_{\mathcal{A}} \right),$$

for every $s \in \mathbb{R}_{\geq 0}$. Integrating both sides and using the definition of V yields

$$\hat{\tau}(s)(f(\hat{x}(s)) - f^\star) + \frac{1}{2}|\hat{x}(s)|^2_{\mathcal{A}} = V_0 - \int_0^s \rho_c(|x(\tilde{s})|_{\mathcal{A}}) d\tilde{s},$$

where $V_0 \coloneqq V(z_0, \tau_0)$ and $(z_0, \tau_0) \coloneqq (\hat{z}(0), \hat{\tau}(0))$. Then, using the positive definiteness of ρ_c together with the fact that $\hat{\tau}(s) = \tau_0 + s$, it follows that the following bound is satisfied:

$$f(\hat{x}(s)) - f^\star \leq \frac{V_0}{s + \hat{\tau}_0} \leq \frac{V_0}{s}, \tag{8.13}$$

for any $s \in \mathbb{R}_{\geq 0}$. Leveraging these results, together with Lemma 8.1 and Theorem 8.1, obtains the following proposition, whose proof is presented in Appendix F.

Proposition 8.1. There exists $\beta_g \in \mathcal{KL}$ such that for any $(x_0, \mu_0) \in \mathbb{R}^n \times \mathcal{X}_\mu$ and any solution (x, μ) to the accelerated gradient flow dynamics (8.10), the following bound holds

$$|x(t) - x^\star| \leq \beta_g(|x_0 - x^\star|, \mathcal{D}_{\mu_0}(t)), \tag{8.14}$$

for all $t \in \mathrm{dom}((x, \mu))$, where $\mathcal{D}_{\mu_0}$ is as defined in Assumption 8.1. Moreover, for any

such solution, the sub-optimality measure $f(\cdot) - f^\star$ satisfies

$$f(x(t)) - f^\star \leq \frac{c_0}{\mathcal{D}_{\mu_0}(t)}, \tag{8.15}$$

where $c_0 \in \mathbb{R}_{\geq 0}$ is a constant that depends on the initial condition x_0.

Remark 8.2. Proposition 8.1 extends the results of the example in Section 8.1 to encompass strictly convex functions, and arbitrary dynamic gains with dynamics satisfying Assumption 8.1. In particular, if F_μ is chosen as in Example 8.1-d), i.e., if we let $F_\mu(\mu) = \frac{\mu^2}{\Upsilon}$, the gradient flow dynamics with dynamic gain described in (8.10) achieve convergence of the function $f(x(t))$ to the optimal value $f^\star$ within prescribed-time $T_{\mu_0} = \mu_0/\Upsilon$, adjustable to the user preference by the choice of $\Upsilon > 0$.

8.2.2 Acceleration of Hybrid Dynamics with Momentum

In cases where the convex function f in the previous section has low curvature, the accelerated gradient flow dynamics in (8.10) might still suffer from poor transient performance. In the case where $\mu \equiv 1$ and $F_\mu(\mu) = 0$, this issue has been alleviated in the existing literature by the use of momentum-based dynamics, see [19],[18],[24]. Inspired by these approaches, in this section we consider a momentum-based nominal HDS, denoted as $\mathcal{H}_m$. This HDS has state $z \coloneqq (x, \mu)$, where $x \coloneqq (x_1, x_2, x_3) \in \mathbb{R}^n \times \mathbb{R}^n \times [\underline{T}, \overline{T}]$, and continuous-time dynamics[1]:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = \mu \cdot F_m(x) \coloneqq \mu \begin{pmatrix} \frac{2}{x_3}(x_2 - x_1) \\ -2x_3 \nabla f(x_1) \\ \frac{1}{2} \end{pmatrix}, \ \dot{\mu} = F_\mu(\mu),$$

[1]Here $\dot{x}_i$ stands for $\frac{\mathrm{d}x_i}{\mathrm{d}t}$.

allowed to evolve whenever $z \in C_m \times \mathcal{X}_\mu$, where the map F_μ is assumed to satisfy Assumption 8.1, $C_m \coloneqq \mathbb{R}^n \times \mathbb{R}^n \times [\underline{T}, \overline{T}]$, and $\overline{T} > \underline{T} > 0$ are tunable parameters. The discrete-time dynamics of the nominal HDS $\mathcal{H}_m$ evolve according to

$$\begin{pmatrix} x_1^+ \\ x_2^+ \\ x_3^+ \end{pmatrix} = G_m(x) \coloneqq \begin{pmatrix} x_1 \\ x_1 \\ 0 \end{pmatrix}, \tag{8.16}$$

whenever $z \in D_m \times \mathcal{X}_\mu$, where $D_m \coloneqq \mathbb{R}^n \times \mathbb{R}^n \times \{1\}$. The momentum-based HDS $\mathcal{H}_m$ with $\mu \equiv 1$ and $F_\mu(\mu)$ was introduced in [24, Eq. (8)] as an alternative able to overcome the lack of uniformity in the convergence properties under Nesterov's ODE to the minimizer of the cost function $x^\star$. As shown in [29, Rmk. 2], such lack of uniformity in the attractivity of $\mathcal{A}$ precludes establishing UGAS bounds which are of critical interest for the robust implementation of the optimization dynamics when only noisy measurements or estimates of the gradient ∇f are available.

Indeed, using the results of [24, Theorem 3.1], we can obtain the UGAS of the set

$$\mathcal{A}_0 \coloneqq \mathcal{A} \times \mathcal{X}_\mu, \text{where } \mathcal{A} \coloneqq \{x^\star\} \times \{x^\star\} \times [\underline{T}, \overline{T}],$$

for the target HDS $\hat{\mathcal{H}}_m$, which is of the from in (8.7) with $F_x = F_m$, $C_x = C_m$, and $D_x = D_m$. This fact, together with Lemma 8.1 and Theorem 8.1, allows us to obtain the following proposition, whose detailed proof is presented in Appendix F.

Proposition 8.2. There exists a function $\beta_m \in \mathcal{KLL}$ such that for any $(x_0, \mu_0) \in \mathbb{R}^n \times \mathcal{X}_\mu$ and any solution (x, μ) to the HDS $\mathcal{H}_m$, the following bound holds

$$|x|_{\mathcal{A}} \leq \beta_m(|x_0|_{\mathcal{A}}, \mathcal{D}_{\mu_0}(t), j), \tag{8.17}$$

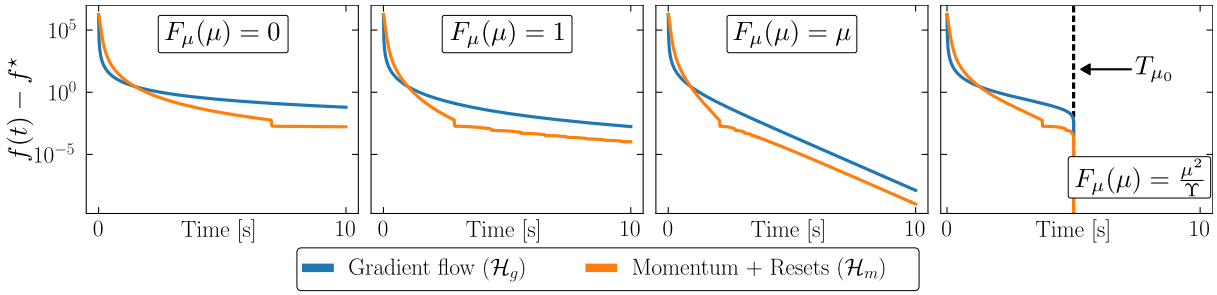


Figure 8.2. Trajectories of solutions to the gradient-flow dynamics with dynamic gain $\mathcal{H}_g$ and the momentum-based dynamics with resets $\mathcal{H}_m$, under different parameterizations of the gain flow maps F_μ.

for all $(t, j) \in \mathrm{dom}((x, \mu))$, where $\mathcal{D}_{\mu_0}$ is as defined in Assumption 8.1. Moreover, any such solution induces the following bound:

$$f(x_1(t, j)) - f^\star \leq \frac{c_j}{(\mathcal{D}_{\mu_0}(t) - \underline{t}_j)^2}, \tag{8.18}$$

where $\underline{t}_j := \min\{t \ : \ (t, j) \in \mathrm{dom}((x, \mu))\}$, and $\{c_j\}_{j=1}^{\infty}$ is a sequence of monotonously decreasing positive constants.

Now, we illustrate the results of Propositions 8.1 and 8.2 by considering the strictly convex cost function $f \in C^1(\mathbb{R}^3)$

$$f(x) = \frac{1}{400}|x - x^\star|^4,$$

where $x^\star = (\pi, 2\pi, 3\pi)$. We simulate the trajectories resulting from the hybrid dynamical systems $\mathcal{H}_g$ (gradient flow with dynamic gain) and $\mathcal{H}_m$ (momentum-based dynamics with resets), using realizations of each gain flow map presented in Example 8.1. For $\mathcal{H}_g$, the gradient flow dynamics are implemented with initial conditions $x_0 = (100, 100, 100)$ and $\mu_0 = 1$. The momentum-based system $\mathcal{H}_m$ is implemented with reset parameters $\underline{T} = 0.01$ and $\overline{T} = 3.5$, and initial conditions $x_1(0, 0) = x_2(0, 0) = x_0$, $x_3(0, 0) = \underline{T}$, and $\mu_0 = 1$.

The resulting trajectories of the suboptimality measure $f(t) - f^*$ are shown in

Figure 8.2, verifying the theoretical bounds derived in Propositions 8.2 and 8.1. As illustrated, incorporating momentum and resets can improve convergence to the cost function minimizer compared to the gradient flow, regardless of the chosen dynamic gain dynamics. However, suitable reset parameters $\underline{T}$ and $\overline{T}$ must be selected in advance.

As the time deformation induced by the gain flow map transitions from constant rescaling (leftmost plot) to prescribed-time scaling (rightmost plot), the frequency of resets naturally increases. This occurs because the state x_3 responsible for triggering resets is scaled by the dynamic gain, while the reset threshold $\overline{T}$ remains constant. This empirical evidence indicates that tuning reset or switching parameters with dynamic gains generally requires accounting for the time-domain deformation induced by the gain flow mappings.

Recent work by the authors in [9] explored this direction by introducing blow-up average-activation time and blow-up average-dwell time conditions for prescribed-time regulation of switched systems. These conditions account for the finite escape times of the dynamic gains employed in prescribed-time stability approaches for switched systems.

This chapter presented a framework to systematically shape the asymptotic convergence behavior of nonlinear dynamical systems, including hybrid systems with combined continuous and discrete dynamics. The key idea was to interconnect the original system with an exogenous dynamic gain designed to induce continuous deformations of the hybrid time domains. By modeling these dynamic gains as dynamical systems and analyzing the diffeomorphisms induced by their flows, we provided sufficient conditions to ensure that a target system's stability properties transfer through the time deformation to the full interconnected dynamics. This formulation enables realizing a diverse range of asymptotic behaviors via time deformations, ranging from constant to prescribed-time scalings. The theory established lays foundations for new techniques to accelerate optimization solvers, enforce real-time control performance, and improve computational efficiency in applications involving nonlinear dynamical systems. Future research directions include extending our approach to the cases where input-to-state, local, semiglobal, and practical stability

certificates are available for the target system.

Acknowledgements

Chapter 8, in its entirety, has been accepted for publication in the proceedings of the 2024 IEEE Conference on Decision and Control. The chapter will appear as a paper titled "Dynamic Gains for Asymptotic-Behavior Shaping in Hybrid Dynamic Inclusions" [8]. The dissertation author is the primary author of this paper.

Chapter 9

Prescribed-Time Stability in Switching Systems with Resets: A Hybrid Dynamical Systems Approach

Recent advances in nonlinear control analysis and design [191, 192, 194, 195] have reinvigorated the concept of Prescribed-Time Stability (PT-S), leading to successful applications across various domains, including nonlinear regulation [191, 192], adaptive control [194], systems with delays [200], partial differential equations [201], and stochastic systems [202]. In contrast to asymptotic or exponential stability, the PT-S property guarantees that the system's trajectories will converge to the desired compact set within a predetermined time, regardless of the initial conditions. As such, achieving this property requires either time-varying or non-Lipschitz vector fields in the dynamics of the system. Non-Lipschitz autonomous systems that achieve convergence to the point (or set) of interest before a fixed time have been studied in [203, 189, 190]. The state of the art of this property, usually called "fixed-time" (FxT) stability, was recently reviewed in [198], with some recent applications in certain classes of hybrid systems under homogeneity conditions [204, 205], continuous-time systems in canonical forms with switching gains [206], and non-switching impulsive systems [207]. In contrast to this line of research, this chapter we study systems that achieve convergence to the target before a prescribed time using the "time-varying gain approach" introduced for ODEs in [191], usually refereed to as "prescribe-time control". This method has a long history in optimal control and tactical missile guidance systems [208], and it has recently gained renewed attention due to breakthroughs

in the design and analysis of nonlinear and adaptive controllers in continuous-time systems with finite-time convergence properties. For a recent survey, see [198] and recent works on adaptive systems [191, 192, 194, 195, 196], PDEs [201, 209, 200, 210], and systems with delays [211, 212]. Since this control approach uses "blow-up" gains over bounded time domains, the solutions of these systems are also defined only over finite-time intervals. For comprehensive discussions on practical applications, strategies to extend the solution domains, and the advantages and limitations of PT control, we refer the reader to recent works [192, 196, 191, 198, 213].

While the study of Prescribed-Time stability properties in continuous-time systems modeled as ordinary differential equations (ODEs) has seen significant progress, PT-S tools for hybrid dynamical systems (HDS) have remained mostly unexplored. For example, switching systems with time-varying gains were studied in [214] using a common Lyapunov function. Similarly, stable controllers that deactivate, or "clip," the high gains before the prescribed time is reached were also discussed in [215]. However, such results consider only one vector field during the convergence phase, and the switching rules can lead to HDS that are not well-posed in the sense of [33]. To the best of our knowledge, general results on PT-S for switching and HDS, similar to those existing for asymptotic or exponential stabilization [216], are still absent in the literature. Since switching and hybrid controllers have been shown to provide powerful solutions to complex control [199, 217], optimization [4, 218], and learning problems [111], there is a clear need for the development of PT-S tools that enable the analysis and design of new algorithms able to simultaneously leverage the advantages of both PT-S and hybrid control.

In this chapter, we address this problem by showing that the PT-S property can be naturally incorporated into a class of HDS that model nonlinear switching systems with resets, allowing the switching signals to incorporate the dynamic effects of time-varying gains, while preserving the structure of the hybrid arcs associated to the solutions of the system. Specifically, the main contributions of this chapter are as follows:

(a) First, we introduce a class of switching signals that preserve the Prescribed-Time Stability (PT-S) property in systems switching between a finite number of PT-S vector fields with exogenous inputs and state resets. To derive these conditions, we reformulate the overall switching system as a hybrid dynamical system (HDS) with dynamic gains that induce appropriate time dilation and contraction in the hybrid time domains of its solutions. By leveraging Lyapunov-based constructions for a suitably normalized HDS evolving on a hybrid dilated time-scale, we show that the original system is PT-Stable, provided the switching signal satisfies a novel "blow-up" average dwell-time (BU-ADT) condition. This condition allows (but does not impose) a non-linear increase in the number of jumps and switches as the total flow time in the system approaches the prescribed convergence time. To study the effect of exogenous inputs and/or disturbances in the system, we establish results via ISS-like bounds "with the convergence property", paralleling those in the literature on PT-S for ODEs [191, Def. 2]. However, unlike the existing results for ODEs, our convergence bounds, presented in Theorem 1, are written in "hybrid time" and highlight the potentially (asymptotically) stabilizing effect of the resets, as well as the order of the dynamics generating the "blow-up" gains. To our knowledge, this is the first result connecting the existing tools on Prescribed-Time Stability for ODEs [191] with the setting of HDS [33].

(b) Next, we incorporate unstable modes into the switching systems, and we characterize a novel "blow-up" average-activation-time (BU-AAT) condition on the amount of time that the system can spend on the unstable modes while preserving the PT-S property. In our model, the unstable modes are also allowed to have "blow-up" time-varying gains with finite-escape times, as well as exogenous inputs and/or disturbances. To study this setting, we construct a HDS with time-ratio monitors, similar in spirit to those considered in [111, 216, 219], but incorporating the blow-up gains into their dynamics, enabling faster switching between the stable and unstable modes as the total amount of flow time in the system approaches the prescribed time. A Lyapunov-based construction

on a dilated-time scale, and a contraction argument on the hybrid time domains, are used to establish in Theorem 9.2 a PT-ISS-like result for switched systems with stable and unstable modes.

(c) To illustrate the applicability of our model and results, we synthesize two different PT-Stable algorithms for the solution of different control and decision-making problems with prescribed-time convergence requirements. First, in Proposition 9.3 we consider the problem of PT regulation of input-affine systems under intermittent feedback, and we propose a new feedback law that extends the results of [191] to plants modeled as switching systems. Finally, we consider the problem of prescribed-time Nash equilibrium seeking in games with switching payoffs via hybrid algorithms with resets. We show in Proposition 9.4 that such algorithms fit into our model and can be studied using the analytical tools presented in the paper.

Switching Systems

In this chapter, we consider switching systems with inputs, with the general form $\dot{x} = \tilde{f}_{\sigma(t)}(x, u, t)$, where $x_0 \in \mathbb{R}^n$ is the initial condition, $x \in \mathbb{R}^n$ is the main state, $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ is an exogenous input assumed to be continuous and bounded, and $\sigma : \mathbb{R}_{\geq 0} \to \mathcal{Q}$ is a right-continuous, piecewise constant, signal that maps the current time t to a finite set of modes $\mathcal{Q} = \{1, 2, \ldots, \overline{q}\}$, where $\overline{q} \in \mathbb{Z}_{\geq 1}$. For each $q \in \mathcal{Q}$, $\tilde{f}_q : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is assumed to be continuous with respect to all arguments. Following the notation of [216], we use $\mathcal{S}$ to denote the set of all right-continuous, piecewise constant, signals from $\mathbb{R}_{\geq 0}$ to $\mathcal{Q}$, with a locally finite number of discontinuities. Such functions are referred to as switching signals. For each signal $\sigma \in \mathcal{S}$, we also define the collection of switching instants $\mathcal{W}(\sigma) \coloneqq \{t \geq 0 : \sigma(t) \neq \sigma(t^-)\}$. In this way, the switching system of interest evolves according to

$$\dot{x}(t) = \tilde{f}_{\sigma(t)}(x, u, t), \quad \forall\, t \notin \mathcal{W}(\sigma), \tag{9.1}$$

where the solutions x to (9.1) are understood in the Caratheodory sense over any interval $[t_a, t_b)$ where σ is constant. During switching times $t \in \mathcal{W}(\sigma)$, we allow "jumps" in the state x via mode-dependent *reset maps* of the form

$$x(t) = R_{\sigma(t^-)}(x(t^-)), \quad \forall\, t \in \mathcal{W}(\sigma), \tag{9.2}$$

where the function $R_q : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be continuous for each $q \in \mathcal{Q}$. Throughout the paper, we will refer to switching systems of the form (9.1)-(9.2) as R-Switching systems.

> **Remark 9.1.** By taking R_q equal to the identity map, system (9.1)-(9.2) recovers a standard switching system [217]. However, other choices of reset maps open the door to study PT-S results in reset control systems [220] (such as impulsive systems by taking $\mathcal{Q} = \{1\}$) as well as more general switched reset controllers (when $|\mathcal{Q}| > 1$), see [216]. It is also possible to consider discontinuous functions $\tilde{f}_q, R_q$ by working with their corresponding Krasovskii regularizations [33, Def. 4.13]. However, for the sake of clarity, we focus on R-switching systems with continuous maps $\tilde{f}_q$ and R_q.

9.1 PT-ISS in Hybrid Dynamical Systems

Motivated by the PT-S property studied for ODEs [191, 192, 194, 195, 200], and before specializing our results to R-switching systems of the form (9.1)-(9.2), in this section we introduce PT-S properties for general HDS with inputs. In particular, we consider systems with state $z = (\psi, \mu_k) \in \mathbb{R}^n \times \mathbb{R}_{\geq 1}$, set C given by:

$$C := \Psi_C \times \mathbb{R}_{\geq 1}, \tag{9.3a}$$

and flow-map defined as:

$$\dot{z} = \begin{pmatrix} \dot{\psi} \\ \dot{\mu}_k \end{pmatrix} \in F(z,u) \coloneqq \begin{pmatrix} \mu_k \cdot F_\Psi(\psi, \mu_k, u), \\ \frac{k}{T}\mu_k^{1+\frac{1}{k}} \end{pmatrix}, \tag{9.3b}$$

where $T > 0$ and $k \geq 1$ are tunable parameters, and $F_\Psi : \mathbb{R}^n \times \mathbb{R}_{\geq 1} \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is a set-valued mapping that we will specify below. The set D is given by

$$D = \Psi_D \times \mathbb{R}_{\geq 1}, \tag{9.3c}$$

and the jump map is given by:

$$z^+ = \begin{pmatrix} \psi^+ \\ \mu_k^+ \end{pmatrix} \in G(z) \coloneqq \begin{pmatrix} G_\Psi(\psi) \\ \mu_k \end{pmatrix}, \tag{9.3d}$$

where $G_\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is also to be specified. We denote the HDS with data given by (9.3) as $\mathcal{H}$. It is assumed that this system satisfies the following standad hybrid basic conditions [33, Assumption 6.5]. These conditions are standard in the hybrid dynamical systems literature [21], and they will be satisfied by construction later when we specialize the results of this section to R-Switching systems with unstable and stable modes.

Assumption 9.1. The sets $\Psi_C, \Psi_D \subset \mathbb{R}^n$ are closed. The set-valued maps F_Ψ and G_Ψ are OSC and LB with respect to Ψ_C, and Ψ_D, respectively; and F_Ψ is convex for all $(\psi, \mu_k, u) \in \Psi_C \times \mathbb{R}_{\geq 1} \times \mathbb{R}^m$.

Remark 9.2. By Definition 1.4, solutions to (1.3) are required to satisfy $\mathrm{dom}(z) = \mathrm{dom}(u)$. To establish this correspondence, we obtain the input u in (1.3) from u in (9.1) using (with some abuse of notation) $u(t,j) = u(t)$ during flows for each fixed j, and by keeping u constant during the jumps (1.3b).

Since in (9.3b) the dynamics of μ_k are independent of ψ, system (9.3) has a cascade structure. However, for system (9.3) the dynamics of ψ will mostly determine the structure of the HTDs of the solutions z, e.g., purely continuous, purely discrete, eventually continuous, etc. To study PT-S properties, in this chapter we consider signals μ_k generated by (9.3b), exhibiting finite escape times that are "controlled" by the parameters (T, k) and by $\mu_k(0)$. This property can be established for the dynamics of μ_k in (9.3b) by direct integration, and it is formalized in Lemma 9.1. The proof is presented in Appendix G.

Lemma 9.1. Let $k \geq 1$, and consider the "*blow-up*" (BU)-ODE $\dot{\mu}_k = \frac{k}{T}\mu_k^{1+\frac{1}{k}}$ with $\mu_k(0) = \mu_0 \in \mathbb{R}_{\geq 1}$. Then, its unique solution satisfies:

$$\mu_k(t) = \frac{T^k}{(\Upsilon_{T,k} - t)^k} \geq 1, \qquad \forall\, t \in [0, \Upsilon_{T,k}), \tag{9.4}$$

where $\Upsilon_{T,k} \coloneqq T\mu_0^{-\frac{1}{k}}$.

For each $k \geq 1$, $\mu_k(\cdot)$ is continuous in its domain, strictly increasing, and satisfies $\lim_{t\to\Upsilon_{T,k}} \mu_k(t) = \infty$. Hence, the next lemma follows directly by the definition of solutions to HDS.

Lemma 9.2. (Bounded Flow-Time) Let z be a maximal solution to $\mathcal{H}$. Then, the HTD of z satisfies $\sup_t(\mathrm{dom}(z)) \leq \Upsilon_{T,k}$.

Lemma 9.2 states that the total amount of flow-time of *every* solution of $\mathcal{H}$ will be upper bounded by $\Upsilon_{T,k}$. We will refer to this quantity as the *prescribed time* (PT), and we emphasize its dependency on the initial value μ_0 and the constants (T, k). In the literature on PT-S in continuous-time, μ_0 is usually equal to one. However, we will consider any $\mu_0 \in \mathbb{R}_{\geq 1}$.

A useful property of the BU-ODE studied in Lemma 9.1, is that, when normalized

by μ_k, the resulting ODE has solutions that are complete and lower bounded by 1. The following Lemma is also proved in Appendix G.

Lemma 9.3. Let $k \geq 1$, and consider the normalized-by-μ_k BU-ODE $\frac{d\hat{\mu}_k}{ds} = \frac{k}{T}\hat{\mu}_k^{\frac{1}{k}}$ with $\hat{\mu}_k(0) = \mu_0 \in \mathbb{R}_{\geq 1}$, evolving in the s-time scale. Then, its unique solution satisfies: (a) For $k = 1$: $\hat{\mu}_k(s) = \mu_0 e^{\frac{s}{T}} \geq 1$ for all $s \geq 0$; (b) For $k > 1$: $\hat{\mu}_k(s) = \left(\frac{(k-1)}{T}s + \mu_0^{\frac{k-1}{k}}\right)^{\frac{k}{k-1}} \geq 1$, for all $s \geq 0$.

9.1.1 Time-Scaling of Hybrid Time Domains

The signals μ_k generated by the dynamics (9.3b) will be used to define a suitable dilation and contraction on the HTD of the solutions to $\mathcal{H}$. To do this, for each $(T, k) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 1}$, and $1 \leq a \leq b$, let the function $\omega_k : \mathbb{R}_{\geq 1} \times \mathbb{R}_{\geq 1} \to \mathbb{R}_{\geq 0}$ be defined as

$$\omega_k(b, a) \coloneqq \frac{T}{k}\left(\frac{b^{\rho(k)} - a^{\rho(k)}}{\rho(k)}\right), \quad \forall\, k > 1, \tag{9.5}$$

and $\omega_1(b, a) \coloneqq \lim_{k \to 1^+} \omega_k(b, a)$, where $\rho(k) \coloneqq \frac{k-1}{k}$. The following proposition states some important properties of $\omega_k(\cdot, \cdot)$ when evaluated along μ_k. All the proofs are presented in Appendix G.

Proposition 9.1. Let $(T, k) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 1}$, μ_k be given by (9.4), and let $\mathcal{T}_k : [0, \Upsilon_{T,k}) \to \mathbb{R}_{\geq 0}$ be the function

$$\mathcal{T}_k(t) \coloneqq \omega_k(\mu_k(t), \mu_k(0)), \quad \forall\, t \in [0, \Upsilon_{T,k}). \tag{9.6}$$

Then, $\mathcal{T}_k(\cdot)$ satisfies the following properties:

(P1) $\lim_{t \to \Upsilon_{T,k}} \mathcal{T}_k(t) = \infty$.

(P2) For any pair $t_2, t_1 \in [0, \Upsilon_{T,k})$ such that $t_2 \geq t_1$:

$$\mathcal{T}_k(t_2) - \mathcal{T}_k(t_1) = \omega_k(\mu_k(t_2), \mu_k(t_1)).$$

(P3) For all $t \in [0, \Upsilon_{T,k})$, we have

$$\frac{d\mathcal{T}_k(t)}{dt} = \mu_k(t), \quad \mathcal{T}_k(0) = 0. \tag{9.7}$$

(P4) For all $t \in [0, \Upsilon_{T,k})$, $\mathcal{T}_k$ has a well-defined inverse $\mathcal{T}_k^{-1} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which is given by

$$\mathcal{T}_k^{-1}(s) = \Upsilon_{T,k} \left(1 - \left(1 + \frac{(k-1)s}{\Upsilon_{T,k}\mu_0} \right)^{\frac{1}{1-k}} \right), \quad k > 1, \tag{9.8}$$

and by $\mathcal{T}_1^{-1}(s) = \lim_{k \to 1^+} \mathcal{T}_k^{-1}(s)$.

(P5) For all $s \in \mathbb{R}_{\geq 0}$, $\mathcal{T}_k^{-1}$ satisfies

$$\frac{d}{ds}\mathcal{T}_k^{-1}(s) = \frac{1}{\mu_k\left(\mathcal{T}_k^{-1}(s)\right)}, \quad \mathcal{T}_k^{-1}(0) = 0. \tag{9.9}$$

(P6) $\lim_{T \to \infty} \mathcal{T}_k(t) = \mu_0^{\frac{k-1}{k^2}} t$ for $k > 1$, and $\lim_{T \to \infty} \mathcal{T}_1(t) = \mu_0 t$ for all $t \geq 0$.

Remark 9.3. To contextualize Proposition 9.1, consider the special case $k = 1$, which is commonly used in the literature on PT-control of ODEs [191, 192]. In this case, Proposition 9.1 yields the following "standard" mappings:

$$\mathcal{T}_1^{-1}(s) = \Upsilon_{T,1} \left(1 - e^{-\frac{1}{T}s} \right), \quad \forall\ s \in \mathbb{R}_{\geq 0}, \tag{9.10a}$$

$$\mathcal{T}_1(t) = T \left(\ln \left(\frac{\mu_1(t)}{\mu_1(0)} \right) \right), \quad \forall\ t \in [0, \Upsilon_{T,1}). \tag{9.10b}$$

Indeed, note that (9.8) can be written as: $\mathcal{T}_k^{-1}(s) = \Upsilon_{T,1}\left(1 - \left(1 + \frac{s}{n(k)T}\right)^{-n(k)}\right)$, with $n(k) = \frac{1}{k-1}$. Using $e^{\frac{s}{T}} = \lim_{n\to\infty}\left(1 + \frac{s}{nT}\right)^n$ and the fact that $n \to \infty$ when $k \to 1^+$, we obtain (9.10a). Similarly, using $\lim_{\rho\to 0}\frac{\mu_1^\rho - 1}{\rho} = \ln(\mu_1)$, and the fact that $\rho(k) \to 0$ if and only if $k \to 1$, (9.10b) follows directly from (9.5) and the definition of ω_1 by applying the product law for limits.

The properties established in Proposition 9.1 are used to derive the following result, which provides a suitable dilation/contraction of the HTDs of $\mathcal{H}$ with data defined by (9.3) when analyzed in a different hybrid time scale (s, j) induced by the transformation $s = \mathcal{T}_k(t)$, see Figure 9.1. Note that, since μ_k does not change during the jumps (9.3d), when evaluating (9.6) along (hybrid) solutions of μ_k generated by (9.3c) we can omit the dependence of $\mathcal{T}_k$ on j.

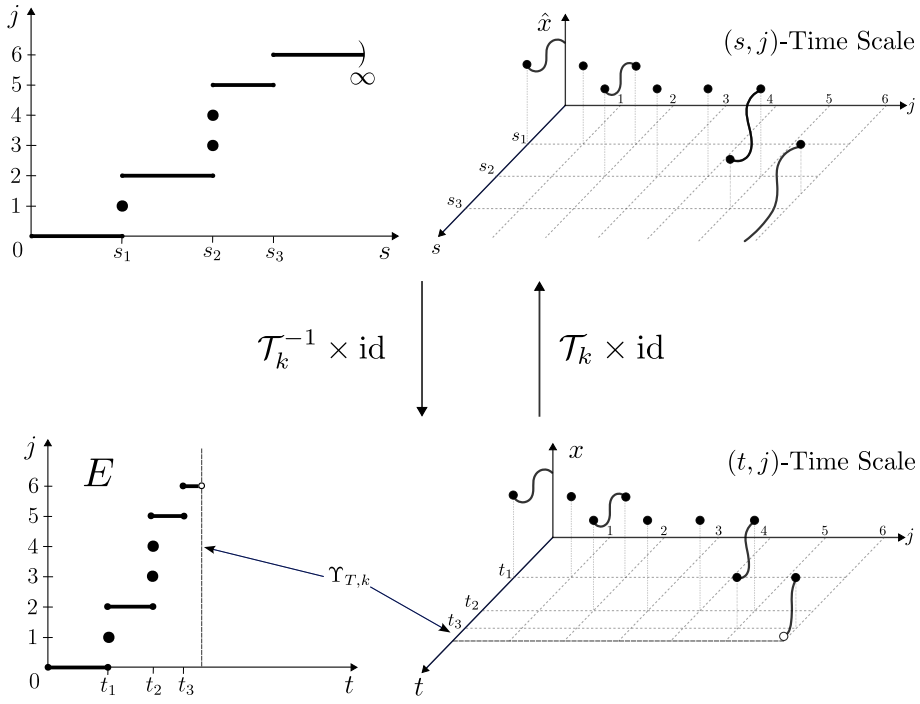


Figure 9.1. Dilation and contraction of hybrid time domains and hybrid arcs. The structure of the hybrid time domain E in the (t, j)-time scale is preserved under the diffeomorphism $\mathcal{T}_k \times \mathrm{id}$ in the (s, j)-time scale.

Proposition 9.2. (Dilation and Contraction of HTD) Let $(T, \mu_0, k) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 1} \times \mathbb{R}_{\geq 1}$, and $\mathcal{T}_k$ be given by (9.6). Consider the following HDS, denoted by $\widehat{\mathcal{H}}$, evolving on the (s, j)-hybrid time scale, with state $\hat{z} = (\hat{\psi}, \hat{\mu}_k)$ and input $\hat{u}$:

$$(\hat{z}, \hat{u}) \in \tilde{C} = C \times \mathbb{R}^m, \quad \dot{\hat{z}}_s \in \frac{1}{\hat{\mu}_k} F(\hat{z}, \hat{u}). \tag{9.11a}$$

$$(\hat{z}, \hat{u}) \in \tilde{D} = D \times \mathbb{R}^m, \quad \hat{z}^+ \in G(\hat{z}). \tag{9.11b}$$

where $(\tilde{C}, F, \tilde{D}, G)$ in (9.11) are the same as in (9.3), and where $\dot{\hat{z}}_s := \frac{\mathrm{d}}{\mathrm{d}s}\hat{z}$.

a) If $(\hat{z}, \hat{u})$ is a maximal solution pair of $\widehat{\mathcal{H}}$ from the initial condition z_0, then the pair of hybrid signals defined as $(z(t, j), u(t, j)) := (\hat{z}(s, j), \hat{u}(s, j))$, for all $(s, j) \in \mathrm{dom}(\hat{z})$, is also a maximal solution pair of $\mathcal{H}$ from the initial condition z_0 via the time dilation $s = \mathcal{T}_k(t)$.

b) If (z, u) is a maximal solution pair of $\mathcal{H}$ from the initial condition z_0, then the pair of hybrid signals defined as $(\hat{z}(s, j), \hat{u}(s, j)) := (z(t, j), u(t, j))$ for all $(t, j) \in \mathrm{dom}(z)$, is also a maximal solution pair of $\widehat{\mathcal{H}}$ from the initial condition z_0 via the time contraction $t = \mathcal{T}_k^{-1}(s)$.

Remark 9.4. Proposition 9.2 establishes a relationship between the solutions of the HDS $\mathcal{H}$ in the (t, j) time scale, and the solutions of $\widehat{\mathcal{H}}$ in the (s, j) time scale via the family of k-parameterized dilations $s = \mathcal{T}_k(t)$ and contractions $\mathcal{T}_k^{-1}(s)$. In particular, the function $\mathcal{T}_k : [0, \Upsilon_{T,k}) \to \mathbb{R}_{\geq 0}$ will define a *diffeomorphism* that preserves the structure of the HTD of the hybrid arcs of $\widehat{\mathcal{H}}$. This observation is central to our analysis, as it enables us to conduct the stability analysis of the original HDS $\mathcal{H}$ by first studying the qualitative behavior of the solutions of system $\widehat{\mathcal{H}}$. In particular, note that $\widehat{\mathcal{H}}$ has a flow map that is normalized by $\hat{\mu}_k$, which removes the finite escape times in $\hat{\mu}_k$ (c.f., Lemma 9.3). This normalized HDS can be viewed as a "target"

system that can be first designed and studied using the extensive set of tools available in the literature on HDS [33, 199].

Remark 9.5. Using (9.4) with $k > 1$, $\mathcal{T}_k$ can be written as

$$\mathcal{T}_k(t) = \frac{T\mu_0^{\frac{k-1}{k}}}{k-1}\left(\frac{T^{k-1}}{\left(T - t\mu_0^{\frac{1}{k}}\right)^{k-1}} - 1\right), \ \forall\ t \in [0, \Upsilon_{T,k}), \tag{9.12}$$

which recovers the common dilation used for ODEs when $\mu_0 = 1$, see [191]. Other types of transformations are presented in [195] for the study of finite-time control of ODEs. Proposition 9.2 provides an extension of these results to hybrid systems.

Remark 9.6. Analyses of HDS based on the time scaling of the flow map are not new, and they have been extensively explored in the context of singular perturbations [221, 30] and averaging theory [222, 111]. However, in contrast to (9.12), the time scaling in those scenarios is usually *linear*.

9.1.2 PT-S via Flows in HDS

Since solutions to system $\mathcal{H}$, whose data is described by (9.3), can only flow for a total amount of time upper bounded by $\Upsilon_{T,k}$, in this chapter we are interested in regulating the state z to a general closed set $\mathcal{A}$, as $t \to \Upsilon_{T,k}$ (or before $\Upsilon_{T,k}$), where

$$\mathcal{A} = \mathcal{A}_\psi \times \mathbb{R}_{\geq 1}, \tag{9.13}$$

and where $\mathcal{A}_\psi$ is an application-dependent compact set. For systems with inputs, the following definition aims to capture this property, which makes use of the transformation $\mathcal{T}_k$ defined in (9.6), and which extends [191, Def. 1] from ODEs to HDS.

Definition 9.1. Let $\mathcal{A}$ be given by (9.13), where $\mathcal{A}_\psi \subset \mathbb{R}^n$ is compact. The set $\mathcal{A}$ is said to be *Prescribed-Time Input-to-State Stable via Flows* (PT-ISS$_\text{F}$) for the HDS $\mathcal{H}$ if there exists $\beta \in \mathcal{KLL}$ and $\gamma \in \mathcal{K}$ such that for every $z(0,0) \in C \cup D$, all solutions z satisfy:

$$|z(t,j)|_\mathcal{A} \leq \beta\big(|z(0,0)|_\mathcal{A}, \mathcal{T}_k(t), j\big) + \gamma\left(|u|_{(t,j)}\right), \tag{9.14}$$

for all $(t,j) \in \text{dom}(z)$. If (9.14) holds with $u \equiv 0$, the set $\mathcal{A}$ is said to be *Prescribed-Time Stable via Flows* (PT-S$_\text{F}$).

In some cases, it might be possible to completely suppress the residual effect of the input u in the bound (9.14) via PT feedback. This property, termed *PT-ISS with Convergence* in [191, Def. 1], can also be obtained in hybrid systems:

Definition 9.2. Let $\mathcal{A}$ be given by (9.13), where $\mathcal{A}_\psi \subset \mathbb{R}^n$ is compact. The set $\mathcal{A}$ is said to be *Prescribed-Time Input-to-State Stable with Convergence via Flows* (PT-ISS-C$_\text{F}$) for the HDS $\mathcal{H}$ if there exists $\beta \in \mathcal{KLL}$, $\gamma \in \mathcal{K}$, and $\beta_c \in \mathcal{KL}$ such that for every $z(0,0) \in C \cup D$, all solutions z satisfy:

$$|z(t,j)|_\mathcal{A} \leq \beta_c\left(\beta\left(|z(0,0)|_\mathcal{A}, \mathcal{T}_k(t), j\right) + \gamma\left(|u|_{(t,j)}\right), \mathcal{T}_k(t)\right), \tag{9.15}$$

for all $(t,j) \in \text{dom}(z)$.

Remark 9.7. (On the use of $\mathcal{KLL}$ functions) The use of $\mathcal{KLL}$ functions in Definitions 9.1 and 9.2 enable us to differentiate convergence behaviors in the continuous-time domain from those in the discrete-time domain. This type of comparison function is common in the analysis of HDS with inputs [21]. Additionally, since by construction $|z(t,j)|_\mathcal{A} = |\psi(t,j)|_{\mathcal{A}_\psi}$ for all $(t,j) \in \text{dom}(z)$ (because $|\mu_k(t,j)|_{\mathbb{R}_{\geq 1}} = 0$), we can

equivalently express the bounds (9.14)-(9.15) with z replaced by ψ, and $\mathcal{A}$ replaced by $\mathcal{A}_\psi$.

Remark 9.8. (On the lack of uniformity with respect to μ_0) Definitions 9.1 and 9.2 extend Prescribed-Time Stability (PT-S) notions, studied in the literature of ODEs, [191, Def. 1] to hybrid systems. The $\mathcal{KLL}$ function β and the $\mathcal{KL}$ function β_c in the bounds (9.14) and (9.15) are independent of the initial conditions on $z = (\psi, \mu)$. However, as defined in (9.6), the diffeomorphism $\mathcal{T}_k$ clearly depends on the initial value of μ_k via (9.6), which parameterizes the prescribed time $\Upsilon_{T,k}$. Yet, the bounds (9.14) and (9.15) are uniform across the initial conditions of ψ, which is the main state of interest in the system.

The following example, which follows as a particular case of the main results in the next section, illustrates the previous discussions:

Example 9.1. Consider the HDS $\mathcal{H}$ with $k = 1$, $T = 1$, $\psi = (x, \tau)$, $F_\Psi = \{-x + u\} \times \{1\}$, $G_\Psi = \{\frac{1}{2}x\} \times \{0\}$, $\Psi_C = \mathbb{R}^n \times [0, 1]$, $\Psi_D = \mathbb{R}^n \times \{1\}$, and u is continuous and bounded. Then, every solution $z = (x, \tau, \mu_1)$ satisfies the following bound (see proof of Theorem 9.1):

$$|\psi(t,j)|_{\mathcal{A}_\psi} \leq k_1 e^{-k_2 \mathcal{T}_1(t)} \left(e^{-k_3(\mathcal{T}_1(t)+j)} |\psi(0,0)|_{\mathcal{A}_\psi} + k_4 |u|_{(t,j)} \right),$$

where $k_i > 0$ and $\mathcal{A}_\psi = \{0\} \times [0, 1]$, for all $(t, j) \in \mathrm{dom}(z)$. Moreover, using (9.10b), the above bound can be written as:

$$|\psi(t,j)|_{\mathcal{A}_\psi} \leq \frac{\mu_1(0,0)^{\alpha_1}}{\mu_1(t,j)^{\alpha_2}} \left(\frac{e^{-qj}}{\mu_1(t,j)^{\alpha_3}} |\psi(0,0)|_{\mathcal{A}_\psi} + \alpha_4 \cdot |u|_{(t,j)} \right),$$

where $\alpha_i > 0$, $\mu_1(0,0) = \mu_0 \geq 1$, and for all $(t, j) \in \mathrm{dom}(z)$. It follows that $\lim_{(t,j)\in\mathrm{dom}(z), t\to\Upsilon_{1,1}} \psi(t, j) = 0$.

It is important to note that, unlike ODEs, for HDS the existence of bounds of the form (9.14)-(9.15) does not necessarily guarantee that the internal state ψ will converge to $\mathcal{A}_\psi$ as $t \to \Upsilon_{T,k}$, for any $\Upsilon_{T,k} > 0$, even if $u \equiv 0$ and z is complete. The following scalar example illustrates this scenario.

Example 9.2. Consider the HDS $\mathcal{H}$ with $k = 1$, main state $\psi \in \mathbb{R}$, functions $F_\Psi = \{-\psi\}$, $G_\Psi = \frac{1}{2}\psi$, and sets $\Psi_C = (-\infty, -1] \cup [1, \infty)$, and $\Psi_D = [-1, 1]$. For this system, we can study stability of ψ with respect to the set $\mathcal{A}_\psi = \{0\}$. For any initial condition to $\mathcal{H}$, $z(0,0) = (\psi_0, \mu_0)$, satisfying $|\psi_0| > 1$ and $\mu_0 = 1$, the unique maximal solution to the HDS satisfies $\psi(t,0) = \psi_0 \left(\frac{T-t}{T}\right)^T$, for all $(t,j) \in [0,t'] \times \{0\}$, where $t' = T(1 - |\psi_0|^{-\frac{1}{T}})$, and $\psi(t,j) = \left(\frac{1}{2}\right)^j \psi(t',0)$, for all $(t,j) \in \bigcup_{j \in \mathbb{Z}_{\geq 1}} \{t'\} \times \{j\}$. It follows that $\psi(t,j) \to \mathcal{A}_\psi$ only as $j \to \infty$. Yet, every solution of the HDS satisfies (9.14) with $u = 0$. This follows by a direct application of item (a) of Proposition 9.2, the result of [223, Thm. 1], and item (b) of Proposition 9.2, in that order.

The previous example shows that bounds of the form (9.14) or (9.15) only guarantee PT-S-like behaviors via the *flows* of the HDS. Therefore, to emulate the existing PT-S bounds obtained for ODEs [191, 192], the "target" HDS $\widehat{\mathcal{H}}$ in (9.11) must generate maximal solutions with hybrid time domains E satisfying $\sup_t E = \infty$, such as those in Example 9.1. In general, this is not possible whenever $C = \emptyset$, or whenever $\widehat{\mathcal{H}}$ has eventually discrete, Zeno, or purely discrete solutions. However, as shown in the next section, for R-Switching systems, discrete solutions can be ruled out by designing appropriate switching signals generated by hybrid automatons that additionally exploit the "blow-up" nature of the functions μ_k.

9.2 PT-ISS in R-Switching Systems

In this section, we apply Proposition 9.2 to study a class of R-switching systems

(9.1)-(9.2) characterized by the following dynamics:

$$\dot{x} = \mu_k(t) \cdot f_{\sigma(t)}(x, \mu_k(t), u, \tau), \qquad t \notin \mathcal{W}(\sigma), \tag{9.16a}$$

$$x(t) = R_{\sigma(t^-)}\left(x(t^-)\right), \qquad t \in \mathcal{W}(\sigma). \tag{9.16b}$$

For generality, in (9.16a) we allow f_σ to depend on μ_k and also on a signal τ that is generated by the following hybrid dynamics

$$\dot{\tau} \in \left[0, \frac{\mu_k(t)}{\tau_d}\right], \qquad t \notin \mathcal{W}(\sigma), \tag{9.17a}$$

$$\tau^+ = \tau - 1, \qquad t \in \mathcal{W}(\sigma), \tag{9.17b}$$

where μ_k is given by (9.4) and $\tau_d > 0$. To contextualize this model, some remarks are in order.

Remark 9.9. When $\mu_k \equiv 1$, $R_\sigma = \mathrm{id}(\cdot)$, and f_σ does not depend on τ and u, equation (9.16) coincides with the conventional nonlinear switching systems examined in [224, 225]. On the other hand, when f_σ depends on u, (9.16a) captures nonlinear switching systems with inputs, similar to those studied [226, 216].

Remark 9.10. When $\mu_k \equiv 1$ and f_σ depends on τ, system (9.16) describes a class of τ-parameterized nonlinear switching systems. In this class, τ is not necessarily constant throughout time, and the function $t \mapsto \tau(t)$ may not be differentiable or even continuous. Such models emerge in, for example, a class of time-triggered reset systems [25, 1] suitable for optimization and learning problems; see also Section 9.3.2 for a specific application.

Remark 9.11. In many applications, the system of interest might not match the exact form of (9.16). This is often the case in PT-regulation and feedback

control of affine dynamical systems with non-zero drift, where multiplying the entire vector fields by the gain μ_k is not feasible. However, as shown later in Section 9.3, appropriate feedback design or variable transformation can reformulate these systems into the form (9.16).

To have a well-posed system, we make the following regularity assumption on system (9.16a):

Assumption 9.2. For each $q \in \mathcal{Q}$, $f_q : \mathbb{R}^n \times \mathbb{R}_{\geq 1} \times \mathbb{R}^m \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is locally Lipschitz, $R_q : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, and $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ is continuous and bounded.

We consider R-switching systems (9.16) with a mix of stable and unstable modes. We denote the set of stable modes as $\mathcal{Q}_s$ and the set of unstable modes as $\mathcal{Q}_u$, such that $\mathcal{Q}_s \cup \mathcal{Q}_u = \mathcal{Q}$ and $\mathcal{Q}_u \cap \mathcal{Q}_s = \emptyset$. To leverage this partition and derive prescribed-time stability results, we proceed to introduce specific stability assumptions for our "target" HDS $\widehat{\mathcal{H}}$ defined in (9.11). Central to these assumptions is the role of a function $\Delta(\hat{\mu}_k)$ that characterizes the effect of the time-varying gain $\hat{\mu}_k$ on the input u in (9.16). In our subsequent analysis, we focus on three specific cases: $\Delta(\hat{\mu}_k) = 0$, $\Delta(\hat{\mu}_k) = 1$, and $\Delta(\hat{\mu}_k) = \hat{\mu}_k^{-\ell}$ with $\ell > 0$.

Assumption 9.3. There exist $\tau_d \in \mathbb{R}_{>0}$, $N_0 \in \mathbb{R}_{\geq 1}$, smooth functions $V_{\hat{q}} : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, where $\hat{q} \in \mathcal{Q}$, and constants $c_{\hat{q},i} > 0$, $i \in \{1, 2, 3, 4, 5\}$, $p > 0$, such that:

a) For all $(\hat{x}, \hat{\tau}, \hat{q}) \in \mathbb{R}^n \times [0, N_0] \times \mathcal{Q}$:

$$c_{q,1}|\hat{x}|^p \leq V_{\hat{q}}(\hat{x}, \hat{\tau}) \leq c_{q,2}|\hat{x}|^p. \tag{9.18a}$$

b) For all $(\hat{x}, \hat{\tau}, \hat{q}, \hat{\mu}_k, \eta) \in \mathbb{R}^n \times [0, N_0] \times \mathcal{Q}_s \times \mathbb{R}_{\geq 1} \times [0, \tau_d^{-1}]$ and for all $u \in \mathbb{R}^m$,

we have:

$$\left\langle \nabla V_{\hat{q}}(\hat{x},\hat{\tau}), \begin{pmatrix} f_{\hat{q}}(\hat{x},\hat{\mu}_k,u,\hat{\tau}) \\ \eta \end{pmatrix} \right\rangle \leq -c_{\hat{q},3} V_{\hat{q}}(\hat{x},\hat{\tau}) + c_{\hat{q},4}\Delta(\hat{\mu}_k)|u|^p. \tag{9.18b}$$

c) For all $(\hat{x},\hat{\tau},\hat{q},\hat{\mu}_k,\eta) \in \mathbb{R}^n \times [0,N_0] \times \mathcal{Q}_u \times \mathbb{R}_{\geq 1} \times [0,\tau_d^{-1}]$ and for all $u \in \mathbb{R}^m$, we have:

$$\left\langle \nabla V_{\hat{q}}(\hat{x},\hat{\tau}), \begin{pmatrix} f_{\hat{q}}(\hat{x},\hat{\mu}_k,u,\hat{\tau}) \\ \eta \end{pmatrix} \right\rangle \leq c_{q,5} V_{\hat{q}}(\hat{x},\hat{\tau}) + c_{\hat{q},4}\Delta(\hat{\mu}_k)|u|^p. \tag{9.18c}$$

d) For all $(\hat{x},\hat{\tau}) \in \mathbb{R}^n \times [1,N_0]$ and $\hat{o},\hat{q} \in \mathcal{Q}$ such that $\hat{q} \neq \hat{o}$:

$$V_{\hat{q}}(R_{\hat{o}}(\hat{x}),\hat{\tau}-1) \leq \chi V_{\hat{o}}(\hat{x},\hat{\tau}), \tag{9.18d}$$

where $\chi > 0$.

Remark 9.12. Inequalities (9.18a)-(9.18b) are common in the context of exponential stability in continuous-time and hybrid systems. For the case when the vector field f_q in (9.16a) does not depend on τ, the function $V_{\hat{q}}$ can also be taken to be independent of $\hat{\tau}$. This is the most common situation in switching systems and systems with resets. An example where f_q does depend on τ will be studied in Section 9.3.2.

Remark 9.13. Inequality (9.18b) in item (b) gives a standard decrease condition on the Lyapunov functions $V_{\hat{q}}$, for each stable mode $\hat{q} \in \mathcal{Q}_s$, and up to a neighborhood of the origin, whose size is parameterized by $\Delta(\hat{\mu}_k)|u|^p$. When $\Delta(\hat{\mu}_k) = 0$, and by [223, Thm. 1], conditions (9.18a)-(9.18b) imply that each mode $\hat{q} \in \mathcal{Q}_s$ renders the origin exponentially stable in the dilated time scale $s = \mathcal{T}_k(t)$ (see Proposition 9.2).

When $\Delta(\hat{\mu}_k) = 1$, and by [21, Prop. 1], conditions (9.18a)-(9.18b) imply that each mode $\hat{q} \in \mathcal{Q}_s$ renders the origin ISS with exponential decay in the dilated time scale. The case $\Delta(\hat{\mu}_k) = \hat{\mu}_k^{-\ell}$, with $\ell > 0$, will emerge in the context of PT-regulation where convergence bounds of the form (9.15) are sought-after. An example in this direction is presented in Section 9.3.

Remark 9.14. Inequality (9.18c) in item (c) rules out finite escape times for the unstable modes $\hat{q} \in \mathcal{Q}_u$. Similar assumptions are considered in the context of asymptotic/exponential stability in switching systems [226, 111]. When $\mathcal{Q}_u = \emptyset$ (i.e., there are no unstable modes), item (c) holds vacuously.

Remark 9.15. Inequality (9.18d) in item (d) considers the effect of the resets on the Lyapunov functions related to each of the modes. Usually (e.g., in standard switching systems) $R_{\hat{q}} = \mathrm{id}(\cdot)$ and $V_{\hat{q}}$ is independent of $\hat{\tau}$, and in this case, inequality (9.18d) holds trivially with $\chi = 1$. When $V_{\hat{q}}$ is independent of $\hat{\tau}$ but $R_{\hat{q}} \neq \mathrm{id}(\cdot)$, item (d) recovers the main assumptions of [216].

9.2.1 Blow-Up Average Dwell-Time Conditions

To achieve asymptotic stability in systems switching between a finite number of stable modes, it is common to assume that for all times $t_2 \geq t_1 \geq 0$, the switching signal σ satisfies an average dwell-time (ADT) condition of the form:

$$N(t_2, t_1) \leq \frac{1}{\tau_d}(t_2 - t_1) + N_0, \tag{9.19}$$

where $N(t_2, t_1)$ is the number of switches of σ in the interval $(t_1, t_2]$, $\tau_d > 0$ is called the dwell-time, and $N_0 \geq 1$ is the chatter bound, see [224, 225], [33, Ch. 2.4]. However, unlike asymptotic convergence results, PT-S properties are defined only over the finite interval $[0, \Upsilon_{T,k})$. Therefore, we consider switching signals defined on similar intervals, which are

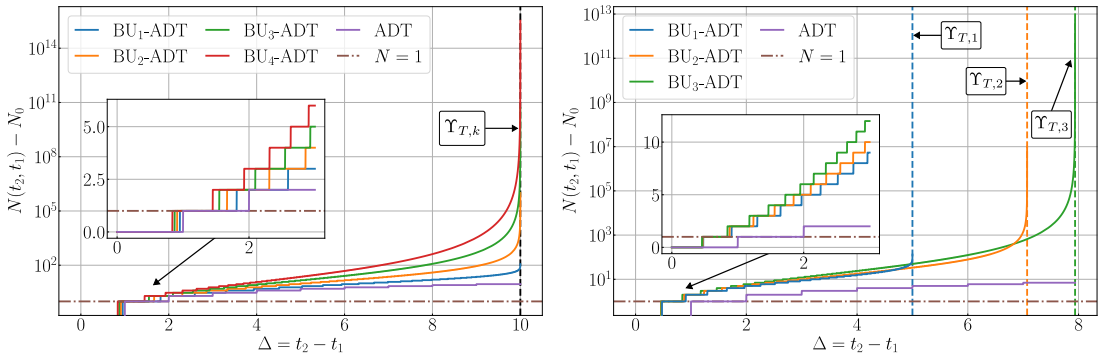


Figure 9.2. BU$_k$-ADT condition (9.20) for $k \in \{1, 2, 3, 4\}$. Left: When $\mu_0 = 1$, $T = 10$, and $t_1 = 0$, there exists a single common terminal time $T = \Upsilon_{T,k}$ for all k. Right: When $\mu_0 = 2$, $T = 10$, and $t_1 = 0$, the dependence of $\Upsilon_{T,k}$ on μ_0 (see Proposition 9.1) leads to the emergence of three distinct terminal times.

additionally allowed to have a switching frequency that becomes unbounded as $t \to \Upsilon_{T,k}$.

Definition 9.3. Let μ_k be given by (9.4). A switching signal $\sigma : [0, \Upsilon_{T,k}) \to \mathcal{Q}$ is said to satisfy the *blow-up average dwell-time condition of order k* (BU$_k$-ADT) if there exist $N_0 \geq 1$ and $\tau_d > 0$ such that for all $t_2, t_1 \in \text{dom}(\sigma)$:

$$N(t_2, t_1) \leq \frac{1}{\tau_d} \omega_k \left(\mu_k(t_2), \mu_k(t_1) \right) + N_0, \tag{9.20}$$

where $\omega_k(\cdot, \cdot)$ is given by (9.5). We use $\Sigma_{\text{BU}_k\text{-ADT}}(\tau_d, N_0, T, \mu_0)$ to denote the family of such signals.

Figure 9.2 illustrates the BU$_k$-ADT condition by comparing various bounds derived from (9.20) (plotted on a logarithmic scale) as functions of $\Delta = t_2 - t_1$, with $t_1 = 0$, and for different values of $k \in \mathbb{Z}_{\geq 1}$, with $\mu_0 = 1$ (left plot) and $\mu_0 = 2$ (right plot). The standard ADT bound (9.19) is also shown in color purple. Unlike the ADT bound, the BU$_k$-ADT bound grows to infinity as $\Delta \to \Upsilon_{T,k}$, allowing an increasing number of switches as $t \to \Upsilon_{T,k}$. However, in any compact sub-interval of $[0, \Upsilon_{T,k})$ the allowable number of switches is bounded. The following lemma shows that switching signals satisfying the ADT condition (9.19) also satisfy the BU$_k$-ADT condition (9.20) when their domain is

appropriately restricted. The implication follows directly because the right-hand side of (9.19) can be upper-bounded by the right-hand side of (9.20). The proof is presented in Appendix G.

Lemma 9.4. Let $T > 0$, $\mu_0 \geq 1$, and σ be a switching signal satisfying the ADT condition (9.19) with $\tau_d > 0$ and $N_0 \geq 1$. Then, $\sigma(t)$ satisfies the BU_k-ADT condition (9.20) for all $k \in \mathbb{Z}_{\geq 1}$ and all $0 \leq t_1 \leq t_2 < \Upsilon_{T,k}$, with the same τ_d, N_0.

Next, we present a lemma that provides an equivalent formulation of the BU_k-ADT condition, as well as its limiting behavior when the prescribed-time $\Upsilon_{T,k}$ goes to infinity. The proof is presented in Appendix G.

Lemma 9.5. The following holds:

a) If $k = 1$, then (9.20) is equivalent to

$$N(t_2, t_1) \leq \frac{T}{\tau_d} \ln\left(\frac{\Upsilon_{T,1} - t_1}{\Upsilon_{T,1} - t_2}\right) + N_0. \tag{9.21}$$

b) If $k \in \mathbb{Z}_{>1}$, then (9.20) is equivalent to

$$N(t_2, t_1) \leq \frac{\gamma_k(t_1, t_2)}{\tau_d} \left((t_2 - t_1) + \sum_{\ell=2}^{k-1} \tilde{c}_{\ell,k} \left(t_2^\ell - t_1^\ell\right)\right) + N_0,$$

where $\tilde{c}_{\ell,k} \coloneqq (-1)^{\ell+1} \frac{b_{k,l}}{k-1} \Upsilon_{T,k}^{1-\ell}$, $b_{k,l} = \frac{(k-1)!}{\ell!(k-\ell-1)!}$ and

$$\gamma_k(t_1, t_2) \coloneqq \mu_0 \left(\frac{\Upsilon_{T,k}^2}{(\Upsilon_{T,k} - t_2)(\Upsilon_{T,k} - t_1)}\right)^{k-1}.$$

c) For all $k \in \mathbb{Z}_{\geq 1}$ and all $t_2 \geq t_1 \geq 0$ the bound (9.20) satisfies

$$\lim_{T \to \infty} \frac{1}{\tau_d} \omega_k\left(\mu_k(t_2), \mu_k(t_1)\right) + N_0 = \frac{\mu_0}{\tau_d}(t_2 - t_1) + N_0,$$

thus recovering the ADT condition (9.19) when $\mu_0 = 1$.

9.2.2 PT-ISS in R-Switching Systems with Stable Modes

When all the modes f_q are stable, i.e., $\mathcal{Q}_u = \emptyset$ and $\mathcal{Q} = \mathcal{Q}_s$, we can study PT-S properties of (9.16) by considering switching signals that satisfy the BU$_k$-ADT bound. In this case, the R-Switching system (9.16) can be analyzed by considering the HDS $\mathcal{H}$ with data (9.3), state $\psi = (x, \tau, q) \in \mathbb{R}^{n+2}$, and

$$F_\Psi(\psi, \mu_k, u) \coloneqq \{f_q(x, \mu_k, u, \tau)\} \times \left[0, \frac{1}{\tau_d}\right] \times \{0\}, \tag{9.22a}$$

$$G_\Psi(\psi, u) \coloneqq \{R_q(x)\} \times \{\tau - 1\} \times \mathcal{Q}_s \backslash \{q\}, \tag{9.22b}$$

$$\Psi_C = \mathbb{R}^n \times [0, N_0] \times \mathcal{Q}_s, \quad \Psi_D = \mathbb{R}^n \times [1, N_0] \times \mathcal{Q}_s. \tag{9.22c}$$

As established in the next lemma, there is a close connection between the HTDs of the solutions of system $\mathcal{H}$ with data (9.22), and the signals σ that satisfy the BU$_k$-ADT condition.

Lemma 9.6. Let $(F_\Psi, G_\Psi, \Psi_C, \Psi_D)$ be given by (9.22a)-(9.22c), and consider the HDS $\mathcal{H}$ under Assumptions 9.2 and 9.3. Then, Assumption 9.1 holds, and:

a) For every maximal solution z and for any pair $(t_1, j_1), (t_2, j_2) \in \mathrm{dom}(z)$, with $t_2 > t_1$, inequality (9.20) holds with $N(t_2, t_1) = j_2 - j_1$.

b) For every HTD satisfying property (a), there exists a solution z of the HDS $\mathcal{H}$ having the said HTD.

One of the main consequences of the equivalence established in Lemma 9.6 is that analyzing the stability properties of the R-switching system (9.16) under the family of switching signals $\Sigma_{\text{BU-ADT}}(\tau_d, N_0, T, \mu_0)$ is equivalent to examining the stability properties

of the HDS $\mathcal{H}$ with $(F_\Psi, G_\Psi, \Psi_C, \Psi_D)$ defined by (9.22a)-(9.22c). In this case, we can study the stability properties of this HDS with respect to the set $\mathcal{A}$ given by (9.13), where $\mathcal{A}_\psi$ is the following compact set

$$\mathcal{A}_\psi = \{0\} \times [0, N_0] \times \mathcal{Q}_s. \tag{9.23}$$

The following Theorem is the first main result of this chapter.

Theorem 9.1. Let $N_0 \geq 1$, $\mathcal{Q}_s \neq \emptyset$, $\mathcal{Q}_u = \emptyset$, and consider the HDS $\mathcal{H}$ with $(F_\Psi, G_\Psi, \Psi_C, \Psi_D)$ given by (9.22a)-(9.22c). Suppose that Assumptions 9.2-9.3 hold, and

$$\tau_d > \frac{\ln(r)}{\min_{q\in\mathcal{Q}} c_{q,3}}, \tag{9.24}$$

where $r := \max\{1, \chi\}$, and $\chi > 0$ is given in Assumption 9.3. For each $(T, k) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 1}$, the following holds:

a) If $\Delta(\mu_k) = 0$, then the set $\mathcal{A}$ is PT-$\mathrm{S_F}$ for $\mathcal{H}$.

b) If $\Delta(\mu_k) = 1$, then the set $\mathcal{A}$ is PT-$\mathrm{ISS_F}$ for $\mathcal{H}$.

c) If $\Delta(\mu_k) = \mu_k^{-\ell}$, then for any $\ell > 0$ the set $\mathcal{A}$ is PT-ISS-$\mathrm{C_F}$ for $\mathcal{H}$.

The following Corollary covers the case $k = 1$, which is the most common in the literature of PT-S [191, 215].

Corollary 9.1. Suppose that all the assumptions of Theorem 9.1 hold, and that $k = 1$. Then, for every solution $z = (x, \tau, q, \mu_k)$ to $\mathcal{H}$, and all $(t, j) \in \mathrm{dom}(z)$, the state x satisfies the following properties:

1. If (9.18b) holds with $\Delta(\mu_1) = 0$ or $\Delta(\mu_1) = 1$, then

$$|x(t,j)| \leq \kappa_1 \left(\frac{\mu_0}{\mu_1(t,j)}\right)^{\kappa_2 T} e^{-\kappa_2 j}|x(0,0)| + \kappa_3 \Delta |u|_{(t,j)}, \tag{9.25}$$

   where $\kappa_i > 0$ for $i \in \{1,2,3\}$.

2. If (9.18b) holds with $\Delta(\mu_1) = \mu_1^{-\ell}$, then:

$$|x(t,j)| \leq \frac{\alpha_1 \mu_0^{\alpha_2}}{\mu_1(t,j)^{\alpha_3}} \left(\frac{e^{-\alpha_4 j}}{\mu_1(t,j)^{\alpha_5}}|x(0,0)| + \alpha_6 |u|_{(t,j)}\right), \tag{9.26}$$

   where $\alpha_i > 0$ for $i \in \{1,2,\ldots,6\}$.

9.2.3 PT-ISS in R-Switching Systems with Unstable Modes

We now consider the scenario where some of the modes f_q in (9.16) are unstable, i.e., $\mathcal{Q}_u \neq \emptyset$ and $\mathcal{Q} = \mathcal{Q}_s \cup \mathcal{Q}_u$. To study this case, we introduce a *blow-up average activation-time* (BU$_k$-AAT) condition on the amount of time that the unstable modes can remain active in any sub-interval of $[0, \Upsilon_{T,k})$.

Definition 9.4. A switching signal $\sigma : [0, \Upsilon_{T,k}) \to \mathcal{Q}$ is said to satisfy the *blow-up average activation-time condition of order k* (BU$_k$-AAT) if there exist $T_0 > 0$ and $\tau_a > 1$ such that for each pair of times $t_2, t_1 \in \text{dom}(\sigma)$:

$$\int_{t_1}^{t_2} \mu_k(t) \cdot \mathbb{I}_{\mathcal{Q}_u}(\sigma(t))dt \leq \frac{1}{\tau_a}\omega_k\left(\mu_k(t_2), \mu_k(t_1)\right) + T_0, \tag{9.27}$$

where μ_k is given by (9.4). We denote the family of such signals as $\Sigma_{\text{BU}_k\text{-AAT}}(\mathcal{Q}_u, \tau_a, T_0, T, \mu_0)$.

Remark 9.16. For asymptotic and exponential stability results in switching systems with both stable and unstable modes [226, 111, 216], it is common to restrict the family of admissible switching signals to those that satisfy the ADT condition (G.6) *and* the following average activation-time (AAT) condition:

$$\int_{t_1}^{t_2} \mathbb{I}_{\mathcal{Q}_u}(\sigma(t))dt \leq \frac{1}{\tau_a}(t_2 - t_1) + T_0, \tag{9.28}$$

where $\tau_a > 1$, and $T_0 > 0$. This bound can be recovered from (9.27) by taking the limit as $T \to \infty$ in both sides of (9.27) and using $\mu_0 = 1$. Also, note that for $k = 1$, the BU_1-AAT condition reduces to:

$$\int_{t_1}^{t_2} \frac{\mathbb{I}_{\mathcal{Q}_u}(\sigma(t))}{\Upsilon_{T,1} - t}dt \leq \frac{1}{\tau_a} \ln\left(\frac{T - t_1\mu_0}{T - t_2\mu_0}\right) + T_0.$$

Similar bounds can be obtained for $k \in \mathbb{Z}_{\geq 2}$ using (9.4).

Figure 9.3 compares the BU_k-AAT bounds and the traditional AAT bound (9.28). The left plot shows the left-hand side of (9.27) for different values of k, under a particular switching signal σ that switches between one stable mode and one unstable mode. The classic AAT bound is shown in purple color. The right plot shows (9.27) for $k = 1$ and different values of τ_a.

To study the PT-S properties of the R-Switching system (9.16) when $\mathcal{Q}$ contains unstable modes, we now consider the HDS $\mathcal{H}$ with state $\psi = (x, \tau, \rho, q) \in \mathbb{R}^{n+3}$, set-valued mappings:

$$F_\Psi \coloneqq \{f_q(x, \mu_k, u, \tau)\} \times \left[0, \frac{1}{\tau_d}\right] \times \left(\left[0, \frac{1}{\tau_a}\right] - \mathbb{I}_{\mathcal{Q}_u}(q)\right) \times \{0\}, \tag{9.29a}$$

$$G_\Psi \coloneqq \{R_q(x)\} \times \{\tau - 1\} \times \{\rho\} \times \mathcal{Q}\backslash\{q\}, \tag{9.29b}$$

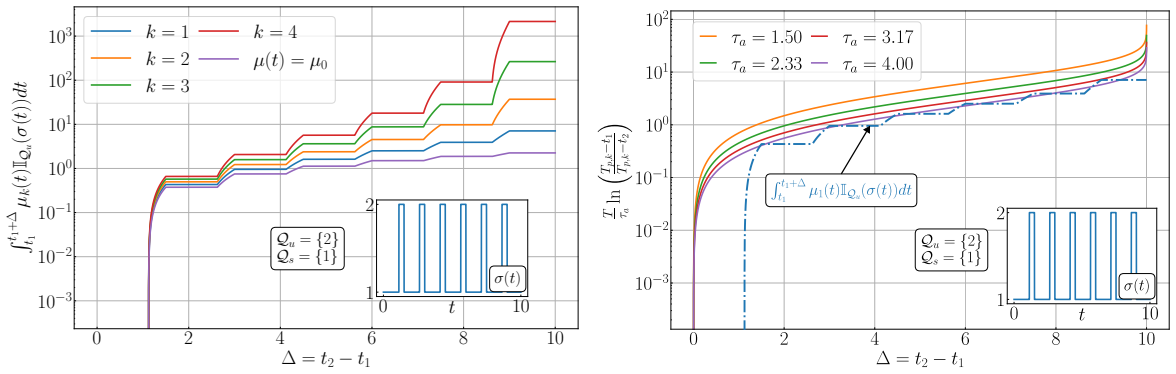


Figure 9.3. Functions appearing in the BU_k-AAT condition (9.27) using the switching signal $\sigma(\cdot)$ (see inset), $T = 10$, and $\mu_0 = 1$.

and sets:

$$\Psi_C = \mathbb{R}^n \times [0, N_0] \times [0, T_0] \times \mathcal{Q}, \tag{9.29c}$$

$$\Psi_D = \mathbb{R}^n \times [1, N_0] \times [0, T_0] \times \mathcal{Q}. \tag{9.29d}$$

There is a close connection between the hybrid time domains of the solutions generated by the HDS $\mathcal{H}$ with data (9.29), and the switching signals that simultaneously satisfy (9.20) and (9.27).

Lemma 9.7. Let $(F_\Psi, G_\Psi, \Psi_C, \Psi_D)$ be given by (9.29a)-(9.29c), and consider the HDS $\mathcal{H}$ given by (9.3), under Assumption 9.2-9.3. Then, Assumption 9.1 holds, and:

a) For every maximal solution z to $\mathcal{H}$ and for any pair $(t_1, j_1), (t_2, j_2) \in \text{dom}(z)$, with $t_2 > t_1$, inequality (9.20) holds with $N(t_2, t_1) = j_2 - j_1$, and inequality (9.27) holds with $\sigma(t) = q(t, \underline{j}(t))$, where $\underline{j}(t) := \min\{j \in \mathbb{Z}_{\geq 0} : (t, j) \in \text{dom}(z)\}$.

b) For every HTD satisfying property (a), there exists a solution z of $\mathcal{H}$ having the said HTD.

Similar to Lemma 9.6, the result of Lemma 9.7 enables the study of the stability

properties of the R-Switching system (9.16), under switching signals σ satisfying (9.20) *and* (9.27), by studying the stability properties of the HDS (G.11). In this case, we consider the set $\mathcal{A}$ given by (9.13), where $\mathcal{A}_\psi$ is now given by

$$\mathcal{A}_\psi = \{0\} \times [0, N_0] \times [0, T_0] \times \mathcal{Q}. \tag{9.30}$$

The next theorem is the second main result of this chapter.

Theorem 9.2. Let $N_0 \geq 1$, $T_0 > 0$, $\mathcal{Q}_u \neq \emptyset$, $\mathcal{Q}_s \neq \emptyset$, and consider the HDS $\mathcal{H}$ given by (9.3) with $(F_\Psi, G_\Psi, \Psi_C, \Psi_D)$ given by (9.29a)-(9.29c). Suppose that Assumptions 9.2-9.3 hold, and that

$$1 > \frac{1}{\underline{c}_3 \tau_d} \ln(r) + \frac{1}{\tau_a} \left(1 + \frac{\overline{c}_5}{\underline{c}_3} \right), \tag{9.31}$$

where $r = \max\{1, \chi\}$, $\chi > 0$ is given in Assumption 9.3, $\underline{c}_3 = \min_{p \in \mathcal{Q}} c_{q,3}$, and $\overline{c}_5 = \max_{p \in \mathcal{Q}} c_{q,5}$. For each $(T, k) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geq 1}$ the following holds:

a) If $\Delta(\mu_k) \triangleq 0$, then the set $\mathcal{A}$ is PT-$\mathrm{S_F}$.

b) If $\Delta(\mu_k) \triangleq 1$, then the set $\mathcal{A}$ is PT-$\mathrm{ISS_F}$.

c) If $\Delta(\mu_k) \triangleq \mu_k^{-\ell}$, $\ell > 0$, then the set $\mathcal{A}$ is PT-ISS-$\mathrm{C_F}$.

Remark 9.17. (Switching with Non-PT Unstable Modes) It is reasonable to consider a situation where the unstable modes in (9.16a) do not have time-varying gains, i.e., $\mu_k \equiv 1$ when $q \in \mathcal{Q}_u$. In particular, consider a system switching between the following two families of systems:

$$\dot{x} = \mu_k f_q(x), \quad q \in \mathcal{Q}_s, \quad \text{and} \quad \dot{x} = f_p(x), \quad p \in \mathcal{Q}_u,$$

where the modes in $\mathcal{Q}_s$ satisfy (9.18b), and the modes in $\mathcal{Q}_u$ satisfy (9.18c) with

$u \equiv 0$. Following the same approach of Theorem 9.2, and operating in the s-time scale for the flows, we now obtain the following two type of modes:

$$\dot{\hat{x}}_s = f_q(\hat{x}), \quad q \in \mathcal{Q}_s, \quad \text{and} \quad \dot{\hat{x}}_s = \frac{1}{\hat{\mu}_k} f_p(x), \quad p \in \mathcal{Q}_u.$$

For this system, the same Lyapunov-based analysis can be applied as in the proof of Theorem 9.2 to obtain the bound (G.15) for all $q \in \mathcal{Q}_s$. On the other hand, for $q \in \mathcal{Q}_u$, we now obtain $\langle \nabla W_2(\hat{z}), \dot{\hat{z}}_s \rangle \leq -(\underline{c}_3 - \delta) W_2(\hat{z}) - \overline{c}_5 \left(1 - \frac{1}{\hat{\mu}_k}\right) W_2(\hat{z})$. Note that $1 - \frac{1}{\hat{\mu}_k} \geq 0$ since $\hat{\mu}_k \geq 1$ by Lemma 9.3. This implies that $\langle \nabla W_2(\hat{z}), \dot{\hat{z}}_s \rangle \leq -(\underline{c}_3 - \delta) W_2(\hat{z})$. From here, the proofs follow the same steps as in the proof of Theorem 9.2.

Remark 9.18. While all our results assumed that the resets (9.16b) were stabilizing, or at least, not destabilizing, it is possible to extend Theorems 9.1-9.2 to cases where the resets are destabilizing, provided the flows of the HDS are "sufficiently" frequent compared to the jumps. In this case, stability can be established by a simple modification of the Lyapunov functions used to study the target systems $\widehat{\mathcal{H}}$ as in [33, Prop. 3.29].

We conclude this section by noting that, with some additional effort, the stability results of Theorems 9.1-9.2 could be extended to systems for which Lyapunov functions with monomial bounds do not exist. While this represents an interesting research direction, such characterizations are beyond the scope of this chapter and could be more appropriately studied in the future within the context of integral-ISS, as described in [216]. For our applications of interest, discussed in the next section, as well as others not detailed here due to space constraints (e.g., concurrent learning [161], extremum seeking [196], feedback-optimization), Assumption 9.3 is typically satisfied.

9.3 Applications to PT-Control and PT-Decision Making

This section presents two applications that illustrate our main results. Throughout this section, the state q and the blow-up gain μ_k are assumed to follow the hybrid dynamics $\mathcal{H}$ defined in (9.3), with data given by (9.22) or (9.29). Since practical implementations of PT-Stable algorithms typically involve early terminations to avoid numerical instabilities, as well as techniques such as clipping and saturation [192, 196, 198], for all our numerical simulations we employ a fourth-order Runge-Kutta method with fixed time step $\delta t = 10^{-6}$ and we saturate the blow-up gain μ_k at 1×10^3.

9.3.1 PT-Regulation with Intermittent Feedback

Consider a switched input-affine system with intermittent feedback, of the form:

$$\dot{x} = d_q(x) + \mathbb{I}_{\mathcal{Q}_s}(q)\, b_q(x) u_q(x, \mu_k), \tag{9.32}$$

where $x \in \mathbb{R}^n$, $q \in \mathcal{Q} = \mathcal{Q}_s \cup \mathcal{Q}_u$ is a logic state and $\mathcal{Q}_u \neq \emptyset$. The blow-up gain μ_k is as defined in (9.4), $d_q(x) \in \mathbb{R}^n$ and $b_q(x) \in \mathbb{R}^{n\times n}$ denote mode-dependent drift and input vector fields, respectively, $u_q : \mathbb{R}^n \times \mathbb{R}_{\geq 1} \to \mathbb{R}^n$ is the control input, and $\mathbb{I}_{\mathcal{Q}_s}(q)$, is an indicator function representing the intermittent nature of the feedback. Such input-affine switching systems model diverse phenomena, ranging from gene regulatory networks in biology [227] to hybrid locomotion in robotics [228]. Incorporating intermittent feedback enhances the practical relevance of these models by addressing challenges such as limited sensor availability, and adversarial operating environments. The implementation of prescribed time controllers proves crucial in scenarios demanding strict time constraints thereby extending the applicability of these models to time-sensitive applications.

We assume that $b_q(\cdot)$ and $d_q(\cdot)$ are unknown locally Lipschitz functions, which

satisfy the following properties:

$$|d_q(x)| \leq \overline{d}_q(x), \qquad \forall\ q \in \mathcal{Q},\ x \in \mathbb{R}^n,$$
$$b_q(x) + b_q(x)^\top \succeq \epsilon I_n, \quad \forall\ q \in \mathcal{Q}_s,\ x \in \mathbb{R}^n,$$

where $\epsilon > 0$, and $\overline{d}_q(x) > 0$ is a known scalar-valued function assumed to be continuous for all $x \in \mathbb{R}^n$ and all $q \in \mathcal{Q}$. We also assume that $\overline{d}_q(x)$ is ℓ_q-globally Lipschitz for all $q \in \mathcal{Q}_u$. To regulate the state x to the origin in a prescribed time, we consider the following switching feedback-law:

$$u_q(x, \mu_k) = -\mu_k \left(\eta_q + \delta_q \overline{d}_q(x)^2\right) x, \tag{9.33}$$

with $\delta_q > 0$ and $\eta_q > 0$ and $k \geq 2$. The closed-loop system has the form of the HDS $\mathcal{H}$ with data (9.29) and continuous-time dynamics of x given by:

$$\dot{x} = \mu_k(t) f_{\sigma(t)}(x, \mu_k), \tag{9.34}$$

where, for every $q \in \mathcal{Q}$, $f_q : \mathbb{R}^n \times \mathbb{R}_{\geq 1} \to \mathbb{R}^n_{\geq 0}$ is given by

$$f_q(x, \mu_k) \coloneqq -\mathbb{I}_{\mathcal{Q}_s}(q) \left(\eta_q + \delta_q \psi_q(x)^2\right) b_q(x) x + \frac{1}{\mu_k} d_q(x).$$

The following proposition extends the results of [191, Sec. 3] to the scenario where the system switches between multiple stable and unstable modes:

Proposition 9.3. There exists $\tau_d > 0$ and $\tau_a > 0$ such that the set $\mathcal{A}_\psi \times \mathbb{R}_{\geq 1}$ is PT-ISS-C_F for the closed-loop system, where $\mathcal{A}_\psi$ is as given in (9.30). Additionally, the switching feedback-law u_q is bounded over the continuous-time interval $[0, \Upsilon_{T,k})$ and converges to 0 as $t \to \Upsilon_{T,k}$.

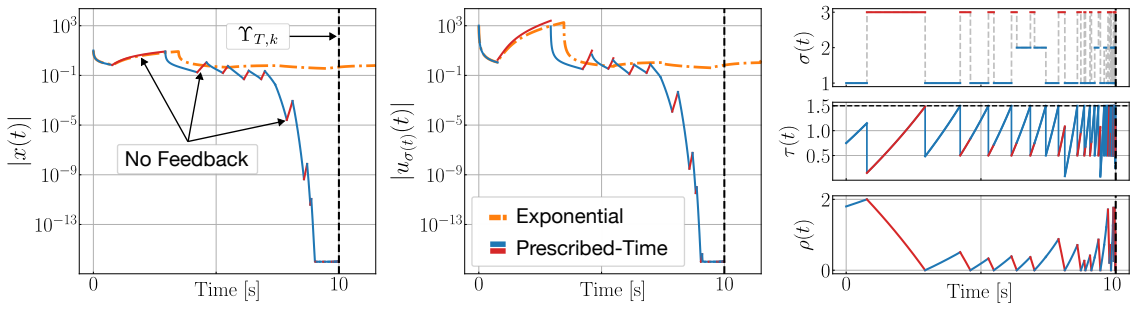


Figure 9.4. Comparison between controller with Exponential convergences and PT-Regulation with intermittent feedback. Left: Trajectory of system's state norm plotted in logarithmic scale. Center: Trajectories of the switching feedback law u_q. Right: Trajectories of the switching signal σ (top), the dwell-time state τ (middle), and the monitor state ρ (bottom) for the PT-Regulation mechanism with intermittent feedback.

To illustrate Proposition 9.3 with a numerical example, consider $\mathcal{Q}_s = \{1, 2\}$, $\mathcal{Q}_u = \{3\}$, and $x \in \mathbb{R}$. Let $d_q(x) = q\tanh(x)$, $b_q(x) = 1$, $\forall q \in \mathcal{Q}$, and consider the control-law $u_q(x,t) = -\mu_2(t)(1+q|x|^2)x$. Then, all the conditions to apply Proposition 9.3 are satisfied. We numerically verify the PT-ISS-C_F property by using a switching signal $\sigma \in \Sigma_{\text{BU-ADT}}(\tau_d, N_0, T, \mu_0) \cap \Sigma_{\text{BU-AAT}}(\mathcal{Q}_u, \tau_a, T_0, T, \mu_0)$ with $\tau_a = 2$, $\tau_d = 1$, $T = 10$, $T_0 = 2$, and $N_0 = 1.5$. Figure 9.4 displays the trajectories of the norm of the state x plotted in logarithmic scale, the switching feedback-law u_q, the switching signal σ, and the associated average dwell-time and average activation time states τ and ρ. As shown in the figure, the state x and the switching feedback-law u_q rapidly approach zero as $t \to \Upsilon_{T,1}$ and converge faster than using a switching feedback with static gains (for exponential convergence). The overshoot occur when the system is in one of the modes without feedback.

9.3.2 PT-Decision-Making in Switching Games

Consider a non-cooperative game with $n \in \mathbb{Z}_{\geq 2}$ players [190], where the cost functions defining the game are allowed to switch in time. Specifically, for each $i \in \mathcal{V} = \{1, 2, \cdots, n\}$, the i^{th} player has an associated mode-dependent and continuously differentiable cost function $\phi_q^i : \mathbb{R}^n \to \mathbb{R}$, where $q \in \mathcal{Q}$. We refer to the q^{th} game as the game with the set of cost functions $\left\{\phi_q^i\right\}_{i\in\mathcal{V}}$. The action of the i^{th} player is denoted by

$x_1^i \in \mathbb{R}$, and the action profile of the game is given by the vector $x_1 \coloneqq (x_1^1, x_1^2, \ldots, x_1^n) \in \mathbb{R}^n$. The goal of the players is to converge to the unique common Nash equilibrium (NE) of the games [35, 190], defined as the vector $\tilde{x} \in \mathbb{R}^n$ that satisfies:

$$\phi_q^i\left(\tilde{x}^i,\ \tilde{x}^{-i}\right) = \inf_{x_1^i \in \mathbb{R}} \phi_q^i\left(x_1^i,\ \tilde{x}^{-i}\right), \quad \forall i \in \mathcal{V},$$

for all $q \in \mathcal{Q}$, where $x_1^{-i} \in \mathbb{R}^{n-1}$ denotes the vector that contains all actions except those of player i. To study this problem, let $\mathcal{G}_q : \mathbb{R}^n \to \mathbb{R}^n$ denote the pseudo-gradient of the q^{th} game, which is given by:

$$\mathcal{G}_q(x_1) \coloneqq \left(\frac{\partial \phi_q^1}{\partial x_1^1},\ \frac{\partial \phi_q^2}{\partial x_1^2}, \ldots, \frac{\partial \phi_q^n}{\partial x_1^n},\right).$$

For all $q \in \mathcal{Q}$, we assume that there exists $\kappa_q > 0$ and $\ell_q > 0$ such that $\mathcal{G}_q$ is a κ_q-strongly monotone and ℓ_q-globally Lipschitz mapping. These properties are common in NE seeking problems and they guarantee the existence and uniqueness of the NE $\tilde{x}$ [190]. To efficiently achieve convergence to the NE in a prescribed time, we introduce *PT high-order NE-seeking dynamics with momentum and resets* (PT-NESmr). The proposed algorithm is modeled as a HDS $\mathcal{H}$ with data (9.22) and maps f_q and R_q defined as follows:

$$f_q(x, \tau) = \begin{pmatrix} \dfrac{2}{\eta(\tau)}\left(x_2 - x_1\right) \\ -2\eta(\tau)\mathcal{G}_q(x_1) \end{pmatrix}, \quad R_q(x) = \begin{pmatrix} x_1 \\ \\ x_1 \end{pmatrix}, \tag{9.35}$$

where $x \coloneqq (x_1, x_2) \in \mathbb{R}^{2n}$, and $x_2 \coloneqq (x_2^1, x_2^2, \ldots, x_2^n) \in \mathbb{R}^n$, and where $\eta : [0, N_0] \to [\underline{\eta}, \overline{\eta}]$ is an affine bounded mapping defined as:

$$\eta(\tau) \coloneqq \tau \frac{\left(\overline{\eta} - \underline{\eta}\right)}{N_0} + \underline{\eta} \tag{9.36}$$

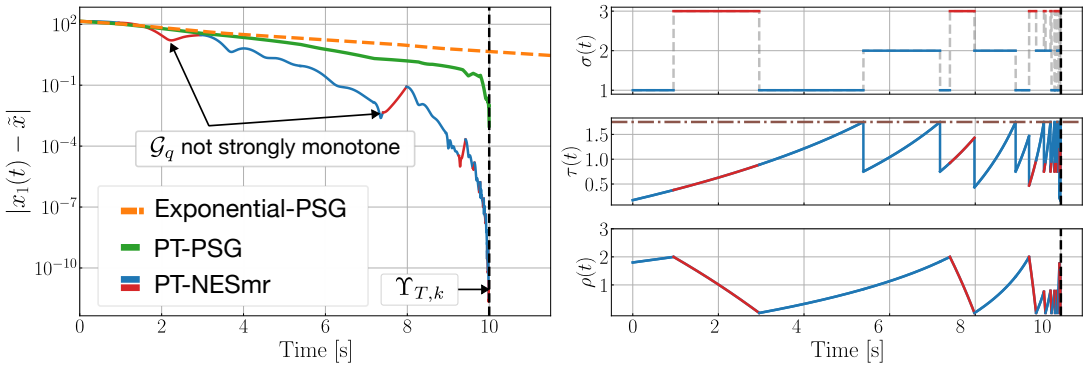


Figure 9.5. Comparison between Pseudo-Gradient Flow (PSG) with exponential convergence and PT Nash-Equilibrium Seeking in a Switching Game. Left: Trajectory of the errors to the NE generated by the PT-NESmr, the PT-PSG, and the Exponential PSG dynamics. Right: Trajectories of the switching signal $\sigma(t)$ (top), the dwell-time state τ (middle), and the monitor state $\rho(t)$ (bottom) for the PT-NESmr dynamics.

with $\overline{\eta} > \underline{\eta} > 0$ being tunable parameters. In the context of asymptotic convergence, mappings of the form (9.35), which incorporate momentum (via the state x_2) and resets (via the update $x_2^+ = x_1$), have been recently shown to improve the transient performance of NE-seeking dynamics in (stable) strongly monotone games [1]. To further make the convergence time independent of both the initial conditions and of the monotonicity properties of the game, we study convergence to the NE in prescribed-time.

For every $q \in \mathcal{Q}$, let $\sigma_q > 0$ be such that $\sigma_{\max}\left(I - \partial \mathcal{G}_q(x)\right) \leq \sigma_q$ for all $x \in \mathbb{R}$, with $\sigma_{\max}(\cdot)$ denoting the maximum singular value of its argument. Such σ_q always exists since the pseudo-gradient $\mathcal{G}_q$ is assumed to be globally Lipschitz for all $q \in \mathcal{Q}$. We make the following assumption on the parameters of the game and the selection of the tunable parameters in (9.35)-(9.36).

Assumption 9.4. (Tuning Guidelines) There exist $0 \leq \underline{\eta} \leq \overline{\eta}$, $\delta_\eta > 0$, and $\delta_d > 0$ satisfying $\delta_\eta + \delta_d := \delta \in (0, 1)$ and:

$$\overline{\eta}^2 \leq \delta_\eta \frac{\min_{q \in \mathcal{Q}} \zeta_q}{\left(\max_{q \in \mathcal{Q}} \sigma_q\right)^2}, \qquad \frac{1}{\tau_d} \leq \delta_d \frac{N_0}{\overline{\eta} - \underline{\eta}} \min_{q \in \mathcal{Q}} \zeta_q, \tag{9.37}$$

for some $\tau_d > 0$ and $N_0 \geq 1$, where $\zeta_q \coloneqq \kappa_q/\ell_q^2$.

The stability properties of the states x_1, x_2 are studied with respect to the following set

$$\mathcal{A}_x \coloneqq \{\tilde{x}\} \times \{\tilde{x}\} \subset \mathbb{R}^n \times \mathbb{R}^n. \tag{9.38}$$

The following proposition establishes PT-S_F of the set $\mathcal{A}_x$ under the PT-NESmr dynamics.

Proposition 9.4. Suppose that Assumption 9.4 is satisfied. Then, the PT-NESmr dynamics render the set $\mathcal{A}_x \times [0, N_0] \times \mathcal{Q} \times \mathbb{R}_{\geq 1}$ PT-S_F, provided

$$\tau_d > \frac{\max\left\{3, 2\left(\frac{1}{\underline{\kappa}^2} + \overline{\eta}^2\right)\right\} \ln(r)}{4\underline{\eta}\nu}, \tag{9.39}$$

where $\nu = \frac{(1-\delta_d-\delta_\eta)\overline{\sigma}^2}{\delta_\eta(1-\delta_d)\underline{\zeta}+\overline{\sigma}^2}$, $\overline{\sigma} \coloneqq \max_{q\in\mathcal{Q}} \sigma_q$, $\underline{\zeta} \coloneqq \min_{q\in\mathcal{Q}} \zeta_q$, and $r = \max\left\{1, \frac{\overline{\ell}^2}{\underline{\kappa}^2}\frac{\eta(N_0-1)^2}{\eta(1)^2} + \frac{1}{2\underline{\kappa}^2\eta(1)^2}\right\}$.

Remark 9.19. (PT-NESmr with non-monotone $\mathcal{G}_q$) Unlike [1], the results of Proposition 9.4 can be directly extended to switching games where some modes lack strong monotonicity in their pseudo-gradients. In this case, we can use the HDS $\mathcal{H}$ with data (9.29) and leverage Theorem 9.2, paralleling the approach followed in Section V.A to study unstable plants. In this case, we obtain conditions on τ_d and τ_a in $\mathcal{H}$, characterizing admissible switching signals under which PT-NESmr dynamics attain prescribed-time stability. This broadens PT-NESmr's applicability to switching games with temporary loss of strong monotonicity.

To illustrate the previous discussion, let $\mathcal{Q} = \{1, 2, 3\}$ and $\mathcal{G}_q(x_1) = \vartheta A_q(x_1 - \tilde{x})$, with $\tilde{x} = (1, 1)$, $A_1 = [6, -1.5; -1.5, 6]$, $A_2 = [8, -2; 2, 8]$, $A_3 = [4, 6; 5; 2]$, and $\vartheta = 5 \times 10^{-2}$. The pseudo-gradient $\mathcal{G}_q(\cdot)$ is κ_q-strongly monotone only for $q \in \{1, 2\} =: \mathcal{Q}_s$ and ℓ_q-globally

Lipschitz for all $q \in \mathcal{Q}$. Using $k = 1$, $\tau_d = \tau_a = 2.5$, $N_0 = 1.75$, $T_0 = 2$ we simulate the system using a switching signal $\sigma \in \Sigma_{\text{BU-ADT}}(\tau_d, N_0, T, \mu_0)$ with $T = 10$. We compare our results with the continuous-time prescribed-time pseudo-gradient-flows (PT-PSG), recently introduced in [229], and given by $\dot{x}_1 = \mu_1(t)\mathcal{G}_{\sigma(t)}(x_1)$. The resulting trajectories are shown in Figure 9.5. As shown in the figure, under the PT-NESmr and the PT-PSG dynamics, the state x_1 rapidly approaches zero as $t \to \Upsilon_{T,1}$ and converges faster than using the standard pseudo-gradient flows with exponential convergence guarantees (Exponential-PSG). Also, note that the synergistic incorporation of momentum, resets, and PT techniques leads to an improvement compared to the continuous-time PT-PSG algorithm under the same switching signal. The overshoots occur when the Nash-equilibrium seeking algorithms operate with a pseudo-gradient that is not κ-strongly monotone, or equivalently when $q \in \mathcal{Q}_u = \mathcal{Q} \setminus \mathcal{Q}_s$.

In this chapter, the property of prescribed-time stability was studied and extended for a class of hybrid dynamical systems incorporating switching nonlinear vector fields with time-varying increasing gains, exogenous inputs, and resets. Novel switching conditions that preserve the prescribed-time stability properties of the system were derived using tools from hybrid dynamical systems theory and under a suitable contraction/dilation of the hybrid time domains. The switching conditions allow the incorporation of unstable modes. The results were illustrated in two applications in the context of control and decision-making. Future applications will include prescribed-time concurrent learning and prescribed-time switching extremum seeking. Future work will also include studying the synergies between non-smooth and prescribed-time tools, as well as consistent discretization mechanisms for HDS, similar to [230].

Acknowledgements

Chapter 9, in full, is a reprint of the material as it appears in "Prescribed-Time Stability in Switching Systems with Resets: A Hybrid Dynamical Systems Approach" [9]. The dissertation author is the first author of this paper.

Part IV

Global Asymptotic Guarantees via Hybrid Dynamical Systems

Chapter 10

Robust Global Optimization on Smooth Compact Manifolds via Hybrid Gradient-Free Dynamics

This chapter studies algorithms for the *global* solution of optimization problems of the form

$$\min \ \phi(z) \quad \text{subject to} \quad z \in M, \tag{10.1}$$

where ϕ is a smooth cost function and (M, g) is an n-dimensional Riemannian manifold to be formally defined in Section 10.1. The mathematical form of ϕ and its derivatives is assumed to be *unknown*. It is only assumed that ϕ is available through *measurements* or *evaluations* on M. This class of problems arises in various practical applications, spanning from aerospace engineering [231] to power systems [232] and quantum control [233]. One of the simplest and most successful algorithms for optimization is the gradient-descent method, which has been studied in the context of manifolds since at least the end of the last century [234]. Recently, these methods have gained considerable interest due to their potential applications in estimation, machine learning, and data science pipelines [235]. In the context of dynamical systems described by ordinary differential equations (ODEs), real-time optimization problems defined on a manifold M are common in robotics, mechanical systems, and aerospace control problems evolving under kinematic constraints. For example, controlling unicycles [236, Sec. 2.2] or navigating in obstacle-occluded spaces

[25]. In such problems, the restriction to evolve on M limits the feasible directions that any onboard algorithm can exploit in real-time. For comprehensive introductions to ODE-based optimization algorithms on manifolds, we refer the reader to [232, 237]. One of the primary challenges in solving optimization problems on smooth (boundaryless) compact manifolds stems from the fact that in such spaces, a point cannot be *robustly globally* asymptotically stabilized using continuous feedback in ordinary differential equations (ODEs) [238, Thm. 1]. This result extends to compact Lie groups and non-contractible spaces in general, as shown in [239, Thm. 21]. This well-known property implies that standard gradient flows or Newton-like flows cannot achieve robust *global* convergence to the minimizer of a continuously differentiable cost function for every type of smooth compact manifold. The reason behind this incompatibility lies in the fundamental mismatch between the topological nature of the basin of attraction of a point under continuous dynamics, and the topological properties of a compact boundaryless manifold [239, Thm. 21]. Specifically, the basin of attraction of a point under continuous feedback is contractible, while a compact manifold is not. Many results in the literature overcome this issue by focusing on asymptotic stability properties that overlook measure-zero sets containing the critical points of the cost function that are not solutions to the optimization problem under study [240, 241], such as local maximizers and saddle points. However, algorithms with *almost* global convergence certificates have been shown to be susceptible to arbitrarily small (adversarial) disturbances. Under such disturbances, the set of problematic initial conditions from which convergence is not achieved is not of measure zero anymore, but rather an open set. Examples illustrating this susceptibility can be found in [236], [25, Ex. 1], and [242].

Alternatively, other works have circumvented the obstruction via time-varying [243], or discontinuous feedback [244], finding success in achieving global convergence in certain applications. However, as shown in [245, Cor. 21], time-varying approaches can only circumvent the issue when the optimization dynamics operate in nominal conditions.

In particular, when the system is subject to (arbitrarily) small disturbances, *robust* and global stabilization of a point in compact boundaryless manifolds cannot be achieved by merely using discontinuous or time-varying feedback strategies. To address this issue, in [246] the authors introduced a *hybrid* controller that synergistically switches between different continuous vector fields, generated from a family of potential functions, to globally stabilize a point. Recent works have employed the synergistic framework to solve attitude stabilization problems in SO(3) [247], stabilization by hybrid backstepping [248], and for the robust stabilization of trajectories in multi-rotor aerial vehicles [249]. However, since these works address stabilization problems, where the point to be stabilized is known *a priori*, in general, they cannot be directly used for the solution of optimization problems where the set of optimizers is unknown, or in cases where the potential functions are only accessible via measurements or evaluations.

Optimization problems where the cost function is unknown and only accessible through measurements or evaluations are common across many applications. These problems have traditionally been studied using gradient-free methods, such as *zeroth-order* optimization algorithms. While the literature of continuous-time zeroth-order optimization dynamics, also known as extremum seeking, is quite rich [250, 251, 111], most of the algorithms applicable to smooth compact manifolds are characterized by smooth gradient-free dynamical systems that aim to emulate, via averaging or other "approximation" technique, the behavior of a target gradient-flow on the manifold [251, 252]. In these settings, the stability properties of gradient-free dynamics are usually inherited from the stability properties of the target system being approximated. Therefore, the challenges of robust global optimization extend to the gradient-free counterparts whenever the target system is characterized by a smooth ODE. Moreover, existing results that achieve global optimization via switching control [253] do not preserve the forward invariance of the manifold due to the use of dither signals that do not evolve in the manifold's tangent space, a requirement that is relevant for practical applications where the evolution on

manifolds is enforced by physical constraints, or in problems where the cost function is defined only on the manifold.

To address the above challenges and limitations of existing approaches, the main contribution of this chapter is the introduction of a novel class of gradient-free algorithms for the *global* solution of optimization problems defined on compact boundaryless connected Riemannian manifolds. The algorithms are characterized by a family of hybrid gradient-free controllers that switch between different zeroth-order feedback laws that implement exploratory geodesic dithers to extract suitable "descent directions" from the cost function ϕ. The switches in the algorithms are implemented in both the exploration and the exploitation components of the dynamics. In particular, to globally navigate and explore manifolds that are not parallelizable (e.g., $\mathbb{S}^2$), the exploratory geodesic dithers switch between different local frames using a hystheresis-based mechanism. To achieve global convergence to (a neighborhood of) the set of minimizers, the algorithms implement a class of switching diffeomorphisms adapted to the cost function of interest. Such diffeomorphisms can be constructed under mild qualitative assumptions on the cost functions and for different types of manifolds. Our main result establishes *robust global* practical asymptotic stability of the set of minimizers of the cost function for the proposed hybrid gradient-free dynamics. Compared to previous approaches for gradient-free optimization on manifolds, e.g. [251], [252] and [254], our convergence results are global rather than local or almost global. Compared to existing switching algorithms [253], our gradient-free dynamics are designed to evolve on the manifold and preserve its invariance via geodesic dithering. The results presented in this chapter are also applicable to a larger class of manifolds and optimization problems. Our results also provide an alternative approach to the solution of gradient-free optimization and extremum seeking problems with multiple critical points, typical in non-convex settings, a problem that has also been recently studied in [255] using other techniques.

10.1 Preliminaries on Differential Geometry

We introduce the main differential geometric concepts used in the chapter. For more details, we refer the reader to [256] and [257]. The concept of smooth manifold will play an important role in this chapter:

Smooth manifolds: An n-dimensional manifold is a second-countable Hausdorff topological space that is locally Euclidean of dimension n. A coordinate chart for M is a pair (U, φ) where $U \subset M$ is an open set and $\varphi : U \to \hat{U} \subset \mathbb{R}^n$ is a homeomorphism. Two coordinate charts (U, φ) and (V, ψ) are said to be smoothly compatible if the transitions maps $\psi \circ \varphi^{-1}$ and $\varphi \circ \psi^{-1}$ are diffeomorphisms. A smooth structure on M is a maximal collection of coordinate charts for which any two charts are smoothly compatible; a smooth coordinate chart is any chart that belongs to a smooth structure. Then, a *smooth manifold* is a manifold endowed with a particular smooth structure. Given a smooth manifold M, the set of all smooth real-valued functions $f : M \to \mathbb{R}$ is denoted by $C^\infty(M)$.

Tangent space and Vector Fields: Dynamical systems evolving on smooth manifolds are defined by vector fields that lie within their tangent spaces. For each $z \in M$, a tangent vector at z is a linear map $v : C^\infty(M) \to \mathbb{R}$ that satisfies $v(fh) = f(z) \cdot v(h) + h(z) \cdot v(f)$, for $f, h \in C^\infty(M)$. The set of all tangent vectors at z is denoted by T_zM and is called the *tangent space* of M at z. The tangent bundle TM is defined to be the disjoint union of the tangent spaces at all points in the manifold, i.e., $TM \coloneqq \bigsqcup_{z \in M} T_zM$. A *smooth vector field* is a smooth map $X : M \to TM$ satisfying $X(z) \in T_zM$ for all $z \in M$. We use $\mathfrak{X}(M)$ to denote the set of all smooth vector fields on M.

The differential of a function $f \in C^\infty(M)$, denoted by $df : TM \to \mathbb{R}$, is a map defined pointwise by:

$$df_z(v) = v(f), \qquad \forall\ v \in T_zM. \tag{10.2}$$

Using the differential, we define the sets of *critical points and critical values* of $f \in C^\infty(M)$

as follows:

$$\mathrm{Crit}f \coloneqq \{z \in M \ : \ df_z = 0\}\,, \tag{10.3}$$

$$\mathrm{Val}f \coloneqq \{a \in \mathbb{R} \ : \ a = f(z),\ z \in \mathrm{Crit}f\}\,. \tag{10.4}$$

A local frame for M is defined as a tuple of vector fields $(X_1, \ldots, X_n)$ defined on an open set $U \subset M$, that is linearly independent and spans T_zM at each $z \in M$. If this frame is defined in the entire manifold ($U = M$), it is called a global frame. When M admits a global frame, the manifold is said to be *parallelizable*. Parallelizability will play an important role in our algorithms.

Riemannian Manifolds: in this chapter, we will focus on Riemannian manifolds. An n-dimensional *Riemannian manifold* is a pair (M, g), where M is an n-dimensional smooth manifold, and g is a Riemannian metric whose value at each point $z \in M$, denoted as g_z, is an inner product defined on T_zM. The Riemannian metric g enables the definition of the gradient of f, $\boldsymbol{\nabla} f : M \to TM$, as the continuous map satisfying:

$$df_z(v) = g\left(\boldsymbol{\nabla} f|_z, v\right), \quad \text{for all } z \in M, v \in T_zM, \tag{10.5}$$

where $\boldsymbol{\nabla} f|_z \in T_zM$ represents the value of the gradient of f at z.

To guarantee a suitable exploration of M, while preserving its invariance, we will work with algorithms that implement geodesic dithers:

Geodesics: Geodesics are defined as curves $\gamma : [a, b] \to M$ on a Riemannian manifold, satisfying

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = 0, \tag{10.6}$$

where $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is the Levi-Civita connection [257, Ch. 5]. To generate the dither signals used by the gradient-free optimization algorithms considered

in this chapter, we use the *restricted exponential map* $\exp_z : T_zM \to M$, defined by $\exp_z(v) = \gamma_v(1)$, where γ_v is the unique maximal geodesic satisfying $\gamma_v(0) = z$ and $\dot{\gamma}_v(0) = v$.

Throughout the chapter, we make use of the following standing assumption.

Assumption 10.1. The Riemannian manifold (M, g) is compact, boundaryless, and connected.

In particular, Assumption 10.1 guarantees the existence of a path between any two points in M [257, Prop 2.50], which facilitates the definition of a notion of distance.

Riemannian Distance: The Riemannian distance, denoted by $d_g(z_1, z_2)$ is defined to be the infimum of the lengths of all admissible curves between a pair of points in the manifold [256, Ch 2.]. Formally, the Riemannian distance $d_g : M \times M \to \mathbb{R}_{\geq 0}$ is defined by $d_g(z_1, z_2) \coloneqq \inf_{\gamma \in \mathbb{A}(z_1,z_2)} \int_{t_1}^{t_2} \sqrt{g\left(\dot{\gamma}(t), \dot{\gamma}(t)\right)} dt$, where $\mathbb{A}(z_1, z_2)$ represents the set of all admissible curves connecting z_1 and z_2, and $t_1, t_2 \in \mathbb{R}$ are such that $\gamma(t_1) = z_1$ and $\gamma(t_2) = z_2$ for $\gamma \in \mathbb{A}(z_1, z_2)$.

Stability Notions: In this chapter we consider algorithms modeled as hybrid dynamical systems (HDS) of the form:

$$x \in C, \qquad \dot{x} = F(x) \tag{10.7a}$$

$$x \in D, \qquad x^+ \in G(x), \tag{10.7b}$$

where some components of the state x will be constrained to evolve in the manifold M. By endowing the manifold M with the distance function d_g, M constitutes a metric space [257, Thm 2.55]. Accordingly, we can use stability notions analogous to those studied in the Euclidean space.

Definition 10.1. The compact set $\mathcal{A} \subset C \cup D \subset M$ is said to be *uniformly globally asymptotically stable* (UGAS) for (10.7) if $\exists\ \beta \in \mathcal{KL}$ such that for all solutions x:

$$|x(t,j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t+j), \tag{10.8}$$

$\forall\ (t,j) \in \mathrm{dom}(x)$, where $|z|_{\mathcal{A}} = \min_{s\in\mathcal{A}} d_g(z,s)$.

We also consider ε-parameterized HDS $\mathcal{H}_\varepsilon$ of the form

$$x \in C_\varepsilon,\ \ \dot{x} = F_\varepsilon(x), \quad \text{and} \quad x \in D_\varepsilon,\ x^+ \in G_\varepsilon(x),$$

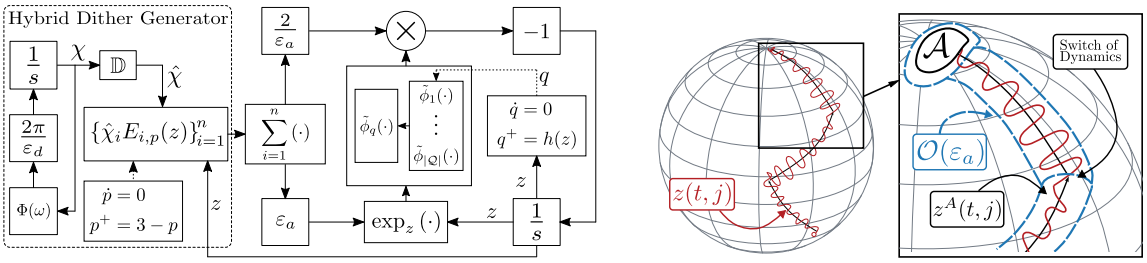
where $\varepsilon > 0$. For these systems, we will study *global practical stability* properties as $\varepsilon \to 0^+$.

Definition 10.2. The compact set $\mathcal{A} \subset C \cup D$ is said to be *Globally Practically Asymptotically Stable* (GP-AS) as $\varepsilon \to 0^+$ for system (10.7) if $\exists\ \beta \in \mathcal{KL}$ such that for each $\nu > 0$ there exists $\varepsilon^* > 0$ such that for all $\varepsilon \in (0, \varepsilon^*)$ and $x(0,0) \in M$, every solution of $\mathcal{H}_\varepsilon$ satisfies

$$|x(t,j)|_{\mathcal{A}} \leq \beta(|x(0,0)|_{\mathcal{A}}, t+j) + \nu, \tag{10.9}$$

$\forall\ (t,j) \in \mathrm{dom}(x)$.

The notion of GP-AS can be extended to systems that depend on two parameters $\varepsilon = (\varepsilon_1, \varepsilon_2)$. In this case, we say that $\mathcal{A}$ is GP-AS as $(\varepsilon_2, \varepsilon_1) \to 0^+$ where the parameters are tuned in order starting from ε_1.



(a) Block diagram of the Zeroth-Order Hybrid Dynamics $\mathcal{H}_0$.

(b) Example of a trajectory z and its average z^A.

Figure 10.1. Left: Block diagram of the proposed hybrid zeroth-order dynamics with geodesic dithering. Right: Cartoon of the trajectories of the system evolving on a manifold M.

Hybrid Gradient-Free Synergistic Optimization

Approaches for optimization in Euclidean spaces with global convergence certificates usually rely on convexity properties of ϕ. For Riemannian manifolds, convexity is characterized along geodesics. However, under Assumption 10.1 geodesic convexity has little utility since, in compact Riemannian manifolds, geodesically convex functions are necessarily constant [258, Cor. 2.5]. Given the limitations imposed by convexity in compact Riemannian manifolds, this chapter opts for an alternative by relying on the following regularity assumption on ϕ, which is closely related to the decomposability of invariant sets introduced in [259, Assumption 1].

Assumption 10.2. The cost function ϕ has a finite amount of critical values, i.e., there exists $l \in \mathbb{N}$ such that $\operatorname{Val} \phi = \{\phi_1, \phi_2, \ldots, \phi_l\}$, where

$$\underline{\phi} := \phi_1 < \phi_2 \leq \phi_i \leq \phi_l =: \overline{\phi},$$

and $\underline{\phi} \leq \phi(z) \leq \overline{\phi}$ for all $z \in M$. Moreover, the critical points of ϕ are isolated, and ϕ has a unique minimizer.

Let $\mathcal{A} := \left\{ z \in \operatorname{Crit} \phi \ : \ \phi(z) = \underline{\phi} \right\}$ represent the minimizer of ϕ and define $\mathcal{B} :=$

Crit $\phi \setminus \mathcal{A}$. Since M is compact, the set $\mathcal{A}$ is also compact. Note that Assumption 10.2 does not rule out functions ϕ with multiple critical points. Indeed, in our problem setup, $\mathcal{B}$ is not empty since, by Morse theory, there exists at least two critical points for scalar-valued functions on compact boundaryless manifolds. Such critical points correspond to equilibria in traditional gradient flows, rendering them highly susceptible to small (adversarial) disturbances. This robustness issue, thoroughly discussed in [236], [25, Ex. 1], and [242], and illustrated later in Section 10.2.2 via numerical examples, is one of the main motivations for the development of robust hybrid algorithms. In our case, we design the hybrid algorithms to be gradient-free by leveraging tools from averaging theory for hybrid dynamical systems.

> **Remark 10.1.** For the case when ϕ is a Morse function [260, Definition 2.3], Assumption 10.2 is automatically satisfied. Moreover, since the set of Morse functions is an open dense set in the space of differentiable functions [260, Theorem 2.7], we can dispense with Standing Assumption 10.2 by considering a surrogate approximate optimization problem to (10.1), whose solution is the minimizer of a Morse function sufficiently close to ϕ.

> **Remark 10.2.** When the set of minimizers $\mathcal{A}$ forms a submanifold rather than a singleton in M, the basin of attraction is diffeomorphic to a tubular neighborhood of $\mathcal{A}$ in M [245, Cor. 21]. This neighborhood may or may not be contractible. In this case, to assess the applicability of our approach, further assumptions regarding the topological characteristics of $\mathcal{A}$ and its tubular neighborhood are required. To simplify our presentation, we defer this problem to future research.

To solve problem (10.1), the left plot of Figure 10.1 shows a block diagram of the proposed controllers. Before analyzing the mathematical properties of this system, we first briefly describe the main ideas behind the algorithms:

(a) A set of dynamic oscillators, with state χ and frequency proportional to $1/\varepsilon_d$, where $\varepsilon_d > 0$ is a small tunable parameter, is employed to generate exploratory signals defined in $\mathbb{T}^n$. The signals are then suitably combined with a local orthonormal frame $\{E_{i,p}\}_{i=1}^n$, $p \in \mathcal{P} \subset \mathbb{Z}_{\geq 1}$, to obtain a dithering vector field $\mathcal{D}_p$ that drives dithering geodesics along the manifold M. These geodesic dithers will be used for the purpose of *local (real-time) exploration*.

(b) To ensure that a local exploration around every point $z \in M$ is well-defined for all time, we let p be a logic state that selects an orthonormal frame $\{E_{i,p}\}_{i=1}^n$ that locally spans the tangent space at a given point z. This logic state is updated using a *hybrid exploration supervisor* that hysteretically switches between local frames. When the manifold is *parallelizable*, we can dispense with this logic state and its associated hybrid dynamics.

(c) The geodesic dithers, together with measurements or evaluations of the cost ϕ, are used to generate families of vector fields $\{\hat{f}_{q,p}(\cdot,\chi)\}_{q\in\mathcal{Q}}$, $p \in \mathcal{P}$, given by

$$\hat{f}_{q,p}(z,\chi) \coloneqq \frac{2}{\varepsilon_a}\tilde{\phi}_q\left(\exp_z\left(\varepsilon_a \mathcal{D}_p(z)\right)\right)\mathcal{D}_p(z), \tag{10.10}$$

where $\varepsilon_a > 0$ is a tunable gain and $\mathcal{Q} \subset \mathbb{Z}_{\geq 1}$. These vector fields, explained below, are used for the purpose of *exploitation* in the optimization dynamics.

(d) To define the vector fields $\{\hat{f}_{q,p}(\cdot,\chi)\}_{q\in\mathcal{Q}}$, we use a set of diffeomorphisms and generate a family of surrogate warped cost functions $\{\tilde{\phi}_q\}_{q\in\mathcal{Q}}$. The chosen diffeomorphisms shift the points that are not in a neighborhood of the minimizers of ϕ. In this manner, by appropriately *partitioning* the manifold M, for each $q \in \mathcal{Q}$ we can implement the vector field $\hat{f}_{q,p}(\cdot,\chi)$ in a "safe zone" where its average dynamics have no critical points other than $\mathcal{A}$. A *hybrid exploitation supervisor* is then used to switch the logic state q to globally steer the state z to $\mathcal{A}$. These partitions can be constructed under mild qualitative assumptions on the cost function.

(f) As we increase the frequency of the dithers (i.e., $\varepsilon_d \to 0^+$), the trajectories induced by the switching vector fields (10.10) will approximate the trajectories of a class of hybrid gradient flows that will be shown to achieve *robust global* asymptotic stability of $\mathcal{A}$ on M.

The above ideas suggest that the proposed algorithms are similar in spirit to other hybrid controllers studied in the context of robust global *stabilization* problems [246], [199, Ch. 7]. However, the algorithms studied in this chapter do not exactly fit the setting of synergistic hybrid control, since the family $\{\hat{f}_{q,p}(\cdot,\chi)\}_{q\in\mathcal{Q}}$ does not describe gradients of synergistic Lyapunov functions. In fact, unlike standard stabilization problems tackled via hybrid control, the main challenges in problem (10.1) are that the set $\mathcal{A}$ and the function ϕ are unknown. Therefore, to implement the gradient-free hybrid dynamics we need to characterize the family of cost functions ϕ and smooth manifolds (M, g) that admit suitable partitions and deformations to generate feasible adaptive switching rules that induce global stability of $\mathcal{A}$, in a gradient-free way.

10.2 Stability, Convergence, and Robustness Results for Parallelizable Manifolds

To solve problem (10.1), we first focus on manifolds M that are parallelizable, which enables the use of a global orthonormal frame $\{E_i\}_{i=1}^n$. This facilitates the definition of a single *dithering vector field* $\mathcal{D} : M \to TM$ as $\mathcal{D}(z) \coloneqq \sum_{i=1}^n \hat{\chi}_i E_i(z)$, where $\hat{\chi}$ corresponds to the vector that stacks the odd components of χ. This single vector field will drive the dithering geodesics, ensuring global exploration of M without the need of using additional logic states (i.e., with $p \equiv 1$). The study of the non-parallelizable scenario is postponed to Section 10.3.

The closed-loop system describing the gradient-free hybrid dynamics, shown in Figure 10.1a, has three main states: $(z,\ q,\ \chi) \in M \times \mathcal{Q} \times \mathbb{T}^n$, where z is an internal auxiliary state, $q \in \mathcal{Q} := \{1, 2, \ldots, N\}$, $N \in \mathbb{Z}_{\geq 2}$, is a logic decision variable, and χ is the

state of the oscillator. The data of this hybrid system is denoted as:

$$\mathcal{H}_0 = \{C_0,\ F_0,\ D_0,\ G_0\}. \tag{10.11}$$

In this way, the continuous-time dynamics of $\mathcal{H}_0$, with state $y \coloneqq (z,\ q,\ \chi)$ are given by

$$y \in C_0, \quad \dot{y} = F_0(y) \coloneqq \begin{pmatrix} -\hat{f}_q\left(z, \chi\right) \\ 0 \\ \frac{2\pi}{\varepsilon_d}\Psi(\omega)\chi \end{pmatrix}, \tag{10.12}$$

where $\hat{f}_q : M \times \mathbb{T}^n \to TM$ is defined via (10.10) by omitting the state p, and $\Psi : \mathbb{R}^n \to \mathbb{R}^{2n\times 2n}$ is given by

$$\Psi(\omega) \coloneqq \begin{pmatrix} \Omega(\omega_1) & 0 & \dots & 0 \\ 0 & \Omega(\omega_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Omega(\omega_n) \end{pmatrix},\ \Omega(\alpha) \coloneqq \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix},$$

where $\alpha > 0$. Here, ω_i is a positive rational number, and $\varepsilon_d \in \mathbb{R}_{>0}$ and $\varepsilon_a \in \mathbb{R}_{>0}$ are tunable gains. For every $q \in \mathcal{Q}$, the vector field $\hat{f}_q(z, \chi)$ is obtained by geodesically dithering the corresponding warped cost function $\tilde{\phi}_q$ (defined below in Definition 10.3) around the current point z. In particular, the dither is obtained along a geodesic γ, originating from z with an initial velocity parameterized by the dithering amplitudes, denoted by χ.

To model the switches between different vector fields, the discrete-time dynamics G_0 of $\mathcal{H}_0$ are given by the following constrained difference inclusion

$$y \in D_0, \quad y^+ \in G_0(y) \coloneqq \{z\} \times h(z) \times \{\chi\}, \tag{10.13}$$

where the set-valued map $h : M \rightrightarrows \mathcal{Q}$, is defined as

$$h(z) \coloneqq \left\{q \in \mathcal{Q}\ :\ \tilde{\phi}_q(z) = m(z)\right\}, \tag{10.14}$$

and $m : M \to \mathbb{R}$ is defined as:

$$m(z) \coloneqq \min_{q \in \mathcal{Q}} \tilde{\phi}_q(z). \tag{10.15}$$

Namely, $m(z)$ is the minimum value among all the warped cost functions $\tilde{\phi}_q$ at a given point z. To compute $m(z)$, the algorithm only needs measurements or evaluations of $\tilde{\phi}_q(z)$, which preserves the gradient-free nature of the hybrid dynamics. Moreover, the minimum in (10.15) is well-defined since $\mathcal{Q}$ is finite, and obtaining the value of m is not computationally expensive, since the complexity scales linearly with the cardinality of $\mathcal{Q}$.

The final elements needed for the characterization of the hybrid system $\mathcal{H}_0$ are the flow and jump sets C_0 and D_0, respectively. To define these sets, and since the warping induced by the diffeomorphisms is only useful if it modifies the points that are not in a neighborhood of the minimizers, we will use a threshold parameter $\gamma \in \mathbb{R}$ characterized by the following assumption:

Assumption 10.3. There exists a known threshold number $\gamma \in (\underline{\phi}, \phi_2)$.

Remark 10.3. Knowledge of γ does not necessarily imply a precise understanding of the minimizer or the exact mathematical form of ϕ. Instead, Assumption 10.3 requires only a mild qualitative understanding of the values of ϕ near its minimum. Such a qualitative characterization is often available in practical scenarios where the range of ϕ is known to lie within certain broad bounds. An example of this can be found in [261, pp. 131], where a known lower bound on the cost function is employed to design the gain of an exploratory signal for extremum seeking control. In the particular case when $\phi = 0$, the assumption holds for any sufficiently small $\gamma > 0$.

Using γ, we can characterize a *synergistic family of diffeomorphisms* for the solution of problem (10.1).

Definition 10.3. Let M be a smooth manifold, and suppose $\phi \in C^\infty(M)$ satisfies Assumption 10.2. A family of functions $\mathcal{S} = \{S_q\}_{q\in\mathcal{Q}}$ is said to be a *δ-gap synergistic family of diffemorphisms adapted to ϕ* if it satisfies:

(A_1) For every $q \in \mathcal{Q}$, $S_q : M \to M$ is a diffeomorphism.

(A_2) For every $q \in \mathcal{Q}$, $\phi(z) < \gamma \implies S_q(z)=z$.

(A_3) There exists $\delta \in (0, \mu(\mathcal{S}))$, where

$$\mu(\mathcal{S}) \coloneqq \min_{\substack{q\in\mathcal{Q} \\ z\in \mathrm{Crit}\ \tilde{\phi}_q \setminus \mathcal{A}}} \left(\tilde{\phi}_q(z) - \min_{p\in\mathcal{Q}} \tilde{\phi}_p(z) \right),$$

and the *warped cost* $\tilde{\phi}_q : M \to \mathbb{R}$ is given by $\tilde{\phi}_q \coloneqq \phi \circ S_q, \quad \forall\, q \in \mathcal{Q}$.

The family of functions $\mathcal{S}$ satisfying the above properties ensures there are enough ways to distort the manifold (M, g), allowing for the distinction of critical points other than the minimizers of ϕ using only cost *measurements or evaluations*. For each distortion of (M, g), a warped cost $\tilde{\phi}_q$ can be defined, leading to a family of N different vector fields in (10.12). Using Definition 10.3, we state our last main standing assumption

Assumption 10.4. There exists a δ-gap synergistic family of diffeomorphisms adapted to ϕ with finite index set $\mathcal{Q}$.

Remark 10.4. Verifying conditions (A1)-(A3) is clearly application-dependent, and different manifolds typically result in different warped costs. However, we stress that the constructions needed to implement the hybrid dynamics do not require explicit mathematical knowledge of the cost function ϕ, but only knowledge of qualitative properties that could be verified a priori via simple tests or experiments. Particular examples of pairs $(\phi, (M, g))$ that satisfy Standing Assumptions 10.2-10.4 will be presented in Section 10.2.3.

The flow sets and jump sets of the zeroth-order hybrid dynamics $\mathcal{H}_0$, given by (10.11), are given by

$$C_0 \coloneqq \left\{(z, q, \chi) \in M \times \mathcal{Q} \times \mathbb{T}^n : \left(\tilde{\phi}_q - m\right)(z) \leq \delta\right\}$$
$$D_0 \coloneqq \{(z, q, \chi) \in M \times \mathcal{Q} \times \mathbb{T}^n : \left(\tilde{\phi}_q - m\right)(z) \geq \delta\}.$$

Based on the structure of the sets (C_0, D_0), switches of q (i.e., jumps) are allowed whenever the difference $\tilde{\phi}_q(z) - m(z)$ exceeds a δ-threshold. Flows following the vector field (10.12) are allowed when this difference is less or equal to δ. When the difference is exactly equal to δ, flows and jumps are both allowed. This immediately indicates that solutions of $\mathcal{H}_0$ are not unique. However, the structure of the warped cost functions $\tilde{\phi}_q$ and the jump map will prevent the occurrence of infinite consecutive jumps by inducing hysteresis-like behavior. In this manner, whenever a solution approaches a critical point of $\tilde{\phi}_q$ outside the set of minimizers $\mathcal{A}$, the dynamics will transition to a different vector field generated from a warped cost function $\tilde{\phi}_p$ with a lower value. The existence of such a warped cost function is guaranteed by the following technical Lemma. All proofs are presented in Section H.

Lemma 10.1. Suppose that ϕ satisfies Assumption 10.2, and let $\mathcal{S} = \{\mathcal{S}_q\}_{q \in \mathcal{Q}}$ be a family of functions satisfying (A_1) and (A_2) in Definition 10.3. If $\mathcal{S}$ satisfies (A_3), then, for all $q \in \mathcal{Q}$ and every $z \in \text{Crit } \tilde{\phi}_q \setminus \mathcal{A}$, there exists $p \in \mathcal{Q}$ such that:

$$\tilde{\phi}_p(z) + \delta < \tilde{\phi}_q(z). \tag{10.16}$$

Conversely, if for all $q \in \mathcal{Q}$ and every $z \in \text{Crit } \tilde{\phi}_q \setminus \mathcal{A}$, there exists $p \in \mathcal{Q}$ such that (10.16) holds, then $\mathcal{S}$ satisfies (A_3), making it a δ-gap synergistic family of diffeomorphisms adapted to ϕ.

We can now state the first main result of the chapter.

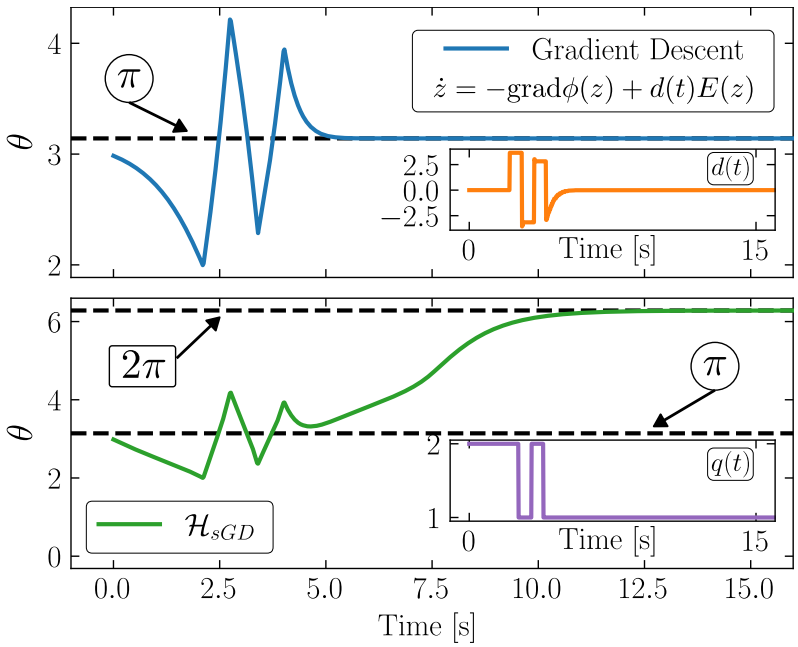


Figure 10.2. Top: Trajectory of a gradient flow under a disturbance $d(t)E(z)$. Bottom: Evolution in time of the main state of $\mathcal{H}_1$ under the same perturbation applied to the z-component of the dynamics. See Example 3.6.

Theorem 10.1. Assume that the manifold M is parallelizable, and consider the hybrid zeroth-order dynamics $\mathcal{H}_0$ with data (10.11). Let the frequencies ω_i in (10.12) satisfy:

$$\omega_i \neq \omega_j, \quad \omega_i \neq 2\omega_j, \ \omega_i \neq 3\omega_j, \quad \text{for all } i \neq j. \tag{10.17}$$

Then, the set $\mathcal{A} \times \mathcal{Q} \times \mathbb{T}^n$ is GP-AS as $(\varepsilon_d, \varepsilon_a) \to 0^+$, and $M \times \mathcal{Q} \times \mathbb{T}^n$ is strongly forward invariant.

The result of Theorem 10.1 establishes global convergence of the trajectories z of $\mathcal{H}_0$ to an arbitrarily small neighborhood of the set of minimizers $\mathcal{A}$, while simultaneously evolving on (and exploring) the manifold M. This behavior is illustrated in Figure 10.1b. To our best knowledge, Theorem 10.1 is the first result in the literature that achieves *global* bounds of the form (10.9) in smooth boundaryless compact Riemannian manifolds via deterministic continuous-time zeroth-order optimization algorithms.

10.2.1 Approximation via 1st-Order Hybrid Dynamics

The result of Theorem 10.1 relies on using averaging theory and perturbation theory (for hybrid systems) to show that, as $(\varepsilon_d, \varepsilon_a) \to 0^+$, the trajectories of $\mathcal{H}_0$ will approximate (on compact time domains) a solution of a first-order hybrid algorithm $\mathcal{H}_1$, with state $x = (z, q)$, continuous-time dynamics given by

$$x \in C_1, \ \dot{x} = F_1(x) \coloneqq \begin{pmatrix} -\sum_{i=1}^{n} \nabla_{E_i} \tilde{\phi}_q(z) E_i(z) \\ 0 \end{pmatrix} \tag{10.18}$$

discrete-time dynamics given by

$$x \in D_1, \quad x^+ \in G_1(x) = \{z\} \times h(z), \tag{10.19}$$

and flow set and jump set given by

$$C_1 \coloneqq \{(z, q) \in M \times \mathcal{Q} \ : \ (\tilde{\phi}_q - m)(z) \leq \delta\} \tag{10.20a}$$

$$D_1 \coloneqq \{(z, q) \in M \times \mathcal{Q} \ : \ (\tilde{\phi}_q - m)(z) \geq \delta\}. \tag{10.20b}$$

Since system (10.18)-(10.20) makes use of first-order information of the warped costs $\tilde{\phi}_q$ via $\nabla_{E_i} \tilde{\phi}_q E_i(z)$, we will refer to this system as the *first-order hybrid dynamics* $\mathcal{H}_1 := \{C_1, F_1, D_1, G_1\}$. In this system, for every $q \in \mathcal{Q}$, the dynamics $\dot{z}$ in (10.18) represents a scaled version of grad $\tilde{\phi}_q$. Similar dynamics have been studied in the literature [252]. They differ from the coordinate representation of

$$\text{grad } \tilde{\phi}_q(z) = \sum_{i,j=1}^{n} \zeta^{ij}(z) \nabla_{E_i} \tilde{\phi}_q(z) E_j(z), \tag{10.21}$$

by excluding the values $\zeta^{ij}(z) \in \mathbb{R}$ that represent the Riemannian metric g at a point $z \in M$, in terms of the basis $\{E_i(z)\}_{i=1}^n$. However, as shown in the following Lemma 10.2, such dynamics do not modify the set of critical points of the warped cost functions.

Lemma 10.2. For all $q \in \mathcal{Q}$ we have that grad $\tilde{\phi}_q|_z = 0$ if and only if $\sum_{i=1}^{n} \nabla_{E_i} \tilde{\phi}_q(z) E_i(z) = 0$.

The following theorem provides a first-order version of Theorem 10.1 for the case when the vector field (10.18) can be explicitly computed or measured in real time, and all the standing assumptions hold.

Theorem 10.2. The first-order hybrid dynamics $\mathcal{H}_1$ render the set $\mathcal{A} \times \mathcal{Q}$ UGAS, and the set $M \times \mathcal{Q}$ is strongly forward invariant.

Similar to Theorem 10.1, the main novelty of Theorem 10.2 is the ability to overcome topological obstructions to global optimization on smooth compact manifolds that emerge in ODEs. In particular, the asymptotic stability result is *global* rather than *almost* global, semi-global, or local. This result, combined with the well-posedness of the dynamics, will allow us to establish important robustness properties with respect to arbitrarily small (potentially adversarial) disturbances, which could also act on the hybrid zeroth-order dynamics $\mathcal{H}_0$.

10.2.2 Robustness Corollaries: Stability Under Adversarial Disturbances

Crucially, the hybrid dynamics $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy the Basic Assumptions of [33, Ch. 6]. Consequently, their stability properties are not drastically affected by small (potentially adversarial) additive disturbances acting on the states and data of the hybrid systems [33, Thm. 7.20]. This is formalized in the following corollary:

Corollary 10.1. Consider the perturbed first-order hybrid dynamics

$$x + d_1 \in C_1, \quad \dot{x} = F_1(x + d_2) + d_3 \tag{10.22a}$$

$$x + d_4 \in D_1, \quad x^+ \in G_1(x + d_5) + d_6 \tag{10.22b}$$

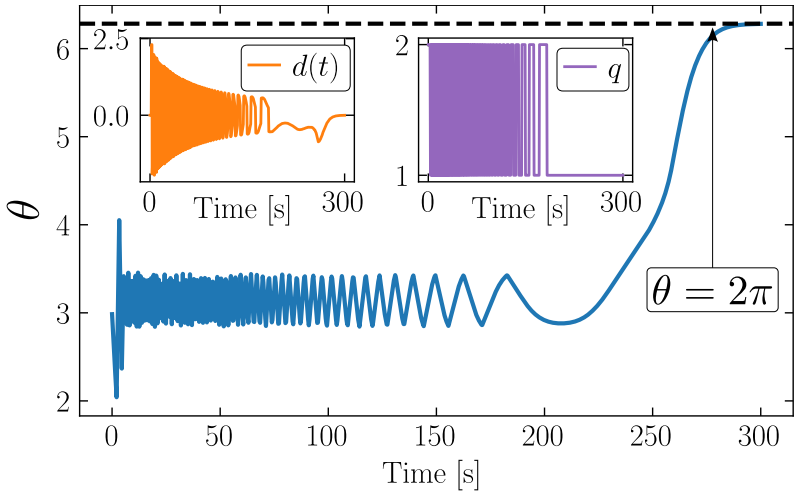


Figure 10.3. Trajectories of $\mathcal{H}_1$, under a small adversarial disturbance generated by a dynamical system. The insets show the amplitude of the injected disturbance, as well as the evolution of the index state q in time. See Example 3.6.

where $\{C_1, F_1, D_1, G_1\}$ is the data of $\mathcal{H}_1$, and the signals $d_j : \text{dom}(x) \to C_1 \cup D_1$, for all $j \in \{1, 2, 4, 5, 6\}$, and $d_3 : \text{dom}(x) \to TC_1$, are measurable functions satisfying $\sup_{(t,j)\in\text{dom}(x)} |d_k(t, j)| \leq d^*$, where $d^* > 0$, for all $k \in \{1, 2, \ldots, 6\}$. Then, system (10.22) renders the set $\mathcal{A} \times \mathcal{Q}$ GP-AS as $d^* \to 0^+$.

Robustness results, such as Corollary 10.1, are relevant for practical applications where measurement noise or numerical approximations induce unavoidable disturbances during implementations. They also hold with respect to adversarial perturbations designed to destabilize the set $\mathcal{A}$, or to stabilize spurious equilibria.

Example 10.1. Let $M = \mathbb{S}^1 \subset \mathbb{R}^2$ be the unit circle, which is a smooth, boundaryless compact parallelizable manifold. We consider the cost function $\phi : \mathbb{S}^1 \to \mathbb{R}$, $z \mapsto 1 - z_1$, where $z_i \in [-1, 1]$ represents the i-th coordinate of $z \in \mathbb{S}^1$ expressed in regular Cartesian coordinates. The cost function ϕ has two critical points in $\mathbb{S}^1$ corresponding to the global minimizer given by (in polar coordinates) $\theta^* = 2\pi$, and a global maximizer, given by $\theta' = \pi$. To find the unknown minimizer of ϕ, we first

implement the first-order dynamics

$$z \in M, \quad \dot{z} = -\nabla_{E(z)} \tilde{\phi}_q(z) E(z) + d(t) E(z), \tag{10.23}$$

where $E : \mathbb{S}^1 \subset \mathbb{R}^2 \to T\mathbb{S}^1$ is the vector field defined by $E(\cos(\theta), \sin(\theta)) = (-\sin(\theta), \cos(\theta))$ and θ denotes the polar coordinate on the circle. By [256, Example 8.10.d)], E constitutes a smooth global frame for $\mathbb{S}^1$. In (10.23), $d(t)E(z)$ is a time-varying perturbation that preserves the invariance of M. The amplitude of this perturbation $d(t)$ was generated by interconnecting (10.23) with an adversarial hybrid system to stabilize the *maximizer* θ'. As shown in Figure 10.2, the adversarial perturbation is always bounded and it succeeds in stabilizing θ'. On the other hand, when this same adversarial signal $d(t)E(z)$ is added in open loop to $\mathcal{H}_1$, as in (10.22), the hybrid dynamics achieve global convergence to the minimizer θ^*, as shown in the bottom plot of Figure 10.2. Finally, we show in Figure 10.3 the performance of the hybrid system $\mathcal{H}_1$ when interconnected to the same adversarial dynamical system used to destabilize θ^* in (10.23). As observed, the hybrid controller still achieves convergence to θ^*.

It is important to note that smooth gradient-free versions of (10.23), obtained via averaging theory, might encounter similar issues as those illustrated in Example 10.1. Specifically, if a small adversarial disturbance can locally stabilize the average dynamics of the system to a point outside $\mathcal{A}$, and if this stabilizing effect of the disturbance is preserved after averaging, then applying the same disturbance to the original dynamics may cause the system to locally converge to a neighborhood of that point, as predicted by standard averaging results for ODEs (e.g., [22, Ch. 10]). An example of this behavior in obstacle avoidance problems was presented in [25, Ex. 1]. The question of systematically constructing such adversarial signals in other manifolds remains application-dependent

and is not further explored in this chapter.

The following corollary parallels the results of Corollary 10.1 for the zeroth-order dynamics $\mathcal{H}_0$.

Corollary 10.2. Consider the perturbed zeroth-order dynamics, given by

$$y + d_1 \in C_0, \quad \dot{y}_0 = F_0(y + d_2) + d_3 \tag{10.24a}$$

$$y + d_4 \in D_0, \quad y_0^+ \in G_0(y + d_5) + d_6 \tag{10.24b}$$

where $\{C_0, F_0, D_0, G_0\}$ is the data of $\mathcal{H}_0$ in (10.11), and the signals $d_j : \mathrm{dom}(y) \to C_0 \cup D_0$, for all $j \in \{1, 2, 4, 5, 6\}$, and $d_3 : \mathrm{dom}(y) \to TC_0$, are measurable functions satisfying $\sup_{(t,j)\in\mathrm{dom}(y)} |d_k(t, j)| \leq d^*$, where $d^* > 0$, for all $k \in \{1, 2, \ldots, 6\}$. Then, system (10.24) renders the set $\mathcal{A} \times \mathcal{Q} \times \mathbb{T}^n$ GP-AS as $(d^*, \varepsilon_2, \varepsilon_1) \to 0^+$.

Remark 10.5. The class of problems for which smooth optimization dynamics cannot achieve robust global certificates on a compact boundaryless manifold M extends beyond the case where the cost has a unique minimizer. Indeed, as briefly stated in Remark 10.2, the basin of attraction of the set of minimizers $\mathcal{A}$ of a continuous cost ϕ under any outer-semicontinuous and locally bounded optimization dynamics F, is diffeomorphic to an open tubular neighborhood of $\mathcal{A}$. In general, this neighborhood is not topologically compatible with M. For instance, when the cost has a finite set of global isolated minimizers $\mathcal{A} = \bigcup_{i\in I} \{\underline{\mathrm{x}}_i\}$, the basin of attraction $\mathcal{B}_F(\mathcal{A}) = \left\{x \in M \ : \ d_g(x, \underline{\mathrm{x}}_i) < \frac{1}{2} \min_{i\neq j} d_g\left(\underline{\mathrm{x}}_j, \underline{\mathrm{x}}_i\right)\right\}$ is not contractible. However, the results of Theorems 10.1 and 10.2 can be directly extended to overcome this type of topological obstruction. We omit this extension due to space limitations.

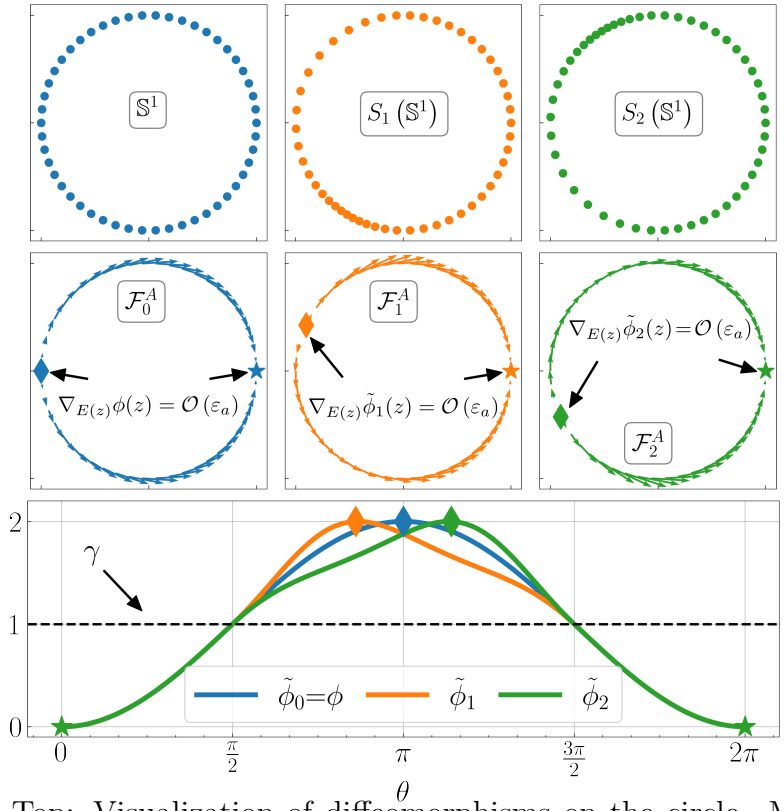


Figure 10.4. Top: Visualization of diffeomorphisms on the circle. Middle: Average gradient-based vector fields derived from warped costs. Bottom: Original and warped costs obtained by precomposing with diffeomorphisms.

10.2.3 Applications: Synthesis of Algorithms

In this section, we showcase the effectiveness of the proposed zeroth-order hybrid dynamics $\mathcal{H}_0$ for solving problems of the form (10.1) on two distinct compact parallelizable Riemannian manifolds. In particular, we show how to synthesize specific algorithms by generating a δ-gap family of diffeomorphisms adapted to smooth cost functions defined in the unitary circle $\mathbb{S}^1$, and in the special orthogonal group SO(3), and we use the hybrid algorithms to achieve *global* gradient-free (practical) optimization while preserving the forward invariance of the manifolds during the real-time exploration.

Gradient-free Feedback Optimization on $\mathbb{S}^1$

Consider the unitary circle $\mathbb{S}^1 = \{z \in \mathbb{R}^2 \ : \ |z|^2 = 1\}$. Given $k_q \in \mathbb{R}$, with q belonging to some index set $\mathcal{Q}$, we define the map $S_q^{(1)} : \mathbb{S}^1 \to \mathbb{S}^1$ as follows:

$$S_q^{(1)}(z) \coloneqq \mathbf{1}_{\{\phi(z) \leq \gamma\}} z + \mathbf{1}_{\{\phi(z) > \gamma\}} e^{k_q \alpha(\phi(z)-\gamma)\Psi} z, \tag{10.25a}$$

where $\Psi \coloneqq e_2 e_1^\top - e_1 e_2^\top \in \mathbb{R}^{2\times 2}$, and $\alpha : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function satisfying: (B_1) $\alpha(0) = 0$; (B_2) $\alpha'(0) = 0$; (B_3) $\alpha'(r) > -1$, $\forall r \geq 0$. The conditions (B_1)-(B_3) ensure that $S_q^{(1)}$ is a continuously differentiable function that constitutes a suitable candidate for a diffeomorphism. In particular, by leveraging [253, Thm 4.1], we have that if

$$|k_q| < \frac{1}{\max\left\{\left|\alpha'\left(\phi(z)-\gamma\right) d\phi_z(\Psi z)\right| \ : \ z \in \mathbb{S}^1, \ \phi(z) \geq \gamma\right\}},$$

then $S_q^{(1)}$ is a diffeomorphism. Although the value of the bound on k_q might not be known (since we do not know the cost function nor its differential) its existence is guaranteed by the continuity of α', ϕ, and $d\phi$, and the compactness of $\{z \in \mathbb{S}^1, \phi(z) \geq \gamma\}$. Estimates of the bound could be obtained by, e.g., a Monte Carlo method that uses measurements of ϕ at different $z \in \mathbb{S}^1$.

Given a cost $\phi^{(1)} : \mathbb{S}^1 \to \mathbb{R}$, and using gains $\{k_q\}_{q\in\mathcal{Q}}$ with corresponding diffeomorphisms defined by (10.25), it is possible to build a suitable δ-gap synergistic family of diffeomorphisms subordinate to $\phi^{(1)}$. To illustrate this process, similarly to Example 10.1, consider the cost function $\phi^{(1)}(z) \coloneqq 1 - z_1$. Assume that only measurements or evaluations of $\phi^{(1)}$ are available for feedback design, but that the intermediate value $\gamma = 1 \in (0, 2) = \left(\underline{\phi}^{(1)}, \phi_2^{(1)}\right)$ and the number of critical points of $\phi^{(1)}$ are known in advance. Let $\alpha(r) = r^2$, and note that it satisfies conditions (B_1)-(B_3). Then, by choosing any two gains satisfying the bound on $|k_q|$, we can obtain a synergistic family of diffeomorphisms subordinate to $\phi^{(1)}$. Indeed, with $\mathcal{Q} = \{1, 2\}, |k_q| < 1, \ q \in \mathcal{Q}, k_1 \neq k_2$

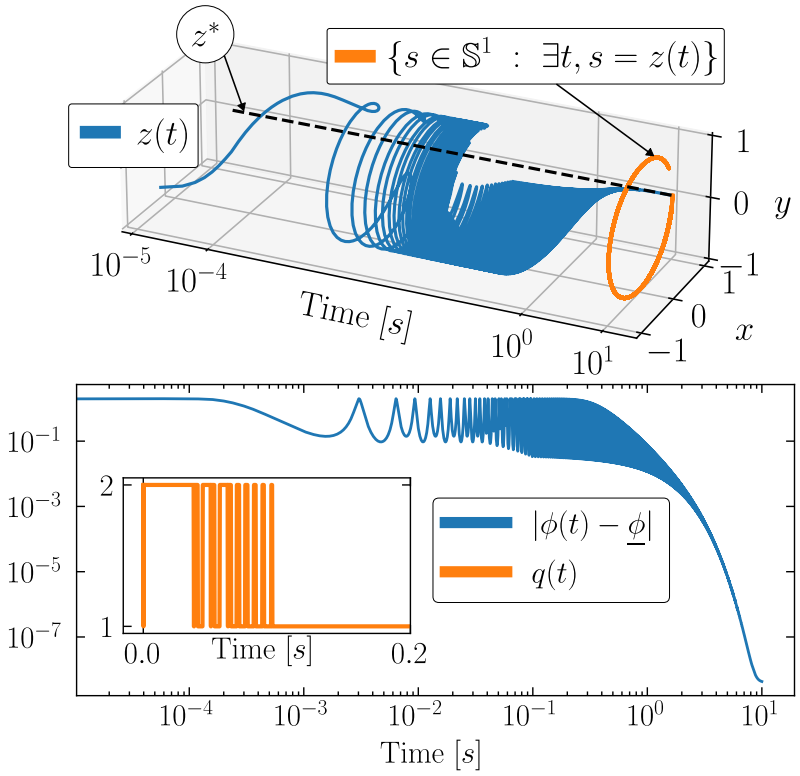


Figure 10.5. Gradient-free global optimization via $\mathcal{H}_0$ on $\mathbb{S}^1$ using Geodesic Dithering.

the set $\mathcal{S}^{(1)} = \left\{S_q^{(1)}\right\}_{q\in\mathcal{Q}}$ is a δ-gap family of diffeomorphisms adapted to $\phi^{(1)}$ with gap $\delta < \mu\left(\mathcal{S}^{(1)}\right)$. In Figure 10.4 we present a visualization of the diffeomorphisms in this family using the choice $k_1 = \frac{1}{2}, k_2 = -\frac{1}{2}$, and we show how these maps warp the original cost function. We also plot the gradient-based vector fields obtained from the warped cost functions which, as shown in Section H, correspond to $\mathcal{O}(\varepsilon_a)$-perturbations of the flows of $\mathcal{H}_1$ in (10.18). In turn, the trajectories of $\mathcal{H}_0$ are shown in Figure 10.5. As observed, the zeroth-order hybrid dynamics with geodesic dithering successfully converge (globally) to the minimizer of $\phi^{(1)}$, $z^* = (1, 0)$, while escaping the other critical point $z' = (-1, 0)$.

Gradient-free Feedback Optimization on SO(3)

As an additional application, we consider the special orthogonal group SO(3), i.e., the group of 3×3 orthogonal matrices with determinant equal to 1 and matrix multiplication as the group operation. By [262, Cor. 3.45], SO(3) forms a 3-dimensional compact Lie

group. The tangent space at z is given by $T_z\mathrm{SO}(3) = \{zX \ : \ X \in \mathbb{R}^{3\times 3}, \ X^\top = -X\}$, see [262, Def. 3.18].

To equip SO(3) with a Riemannian structure, we consider the metric $\langle X, Y\rangle_z = \mathrm{tr}(X^\top Y)$ for all $z \in \mathrm{SO}(3)$, and all $X, Y \in T_z\mathrm{SO}(3)$. Using this choice, the Riemannian exponential map can be written in terms of the matrix exponential $e^{(\cdot)}$ [263, Prop. 21.20] as $\exp_z(X) = ze^{z^{-1}X}$ for $X \in T_z\mathrm{SO}(3)$. Moreover, since SO(3) is a Lie group, it is parallelizable [256, Cor. 8.39]. Indeed, for $i \in \{1, 2, 3\}$ let $E_i : \mathrm{SO}(3) \to T\mathrm{SO}(3)$ be the vector field defined as $E_i(z) = zb_i$, where

$$b_1 \coloneqq \frac{1}{\sqrt{2}}(e_3e_2^\top - e_2e_3^\top), \quad b_2 \coloneqq \frac{1}{\sqrt{2}}(e_1e_3^\top - e_3e_1^\top),$$
$$b_3 \coloneqq \frac{1}{\sqrt{2}}(e_2e_1^\top - e_1e_2^\top).$$

It follows that for every $z \in \mathrm{SO}(3)$, $T_z\mathrm{SO}(3) = \mathrm{span}\,\{E_i(z)\}_{i=1}^3$ and $\langle E_i(z), E_j(z)\rangle_z = \delta_{ij}$, which implies that $\{E_i\}_{i=1}^\infty$ constitutes an orthonormal global frame for SO(3). Using this global frame, we can implement the dithering vector field $\mathcal{D}(z) = \sum_{i=1}^n \hat{\chi}_i E_i(z)$ *everywhere* to extract suitable information from a cost function ϕ at every point in SO(3).

Given a cost $\phi \in C^\infty(\mathrm{SO}(3))$, to establish a suitable family of diffeomorphisms consider the map $S_q^{(2)} : \mathrm{SO}(3) \to \mathrm{SO}(3)$, defined as

$$\mathcal{S}_q^{(2)}(z) = \mathbf{1}_{\{\phi(z)\leq\gamma\}}z + \mathbf{1}_{\{\phi(z)>\gamma\}}e^{k_q\alpha(\phi(z)-\gamma)X}z, \tag{10.26}$$

where $k_q \in \mathbb{R}^n$ and $X \in T_I\mathrm{SO}(3), \ X \neq 0$ are tunable parameters, and $\alpha : \mathbb{R} \to \mathbb{R}$ satisfies the conditions (B_1)-(B_3) defined in Section 10.2.3 to ensure continuous differentiability of the map. The definition of the map $\mathcal{S}_q^{(2)}$, results from modifying the function introduced in [242, Sec 3.4.3] for the angular warping of the two-dimensional sphere by using the function α, and letting the warping act only when ϕ exceeds the threshold γ. For this map we establish the following technical lemma:

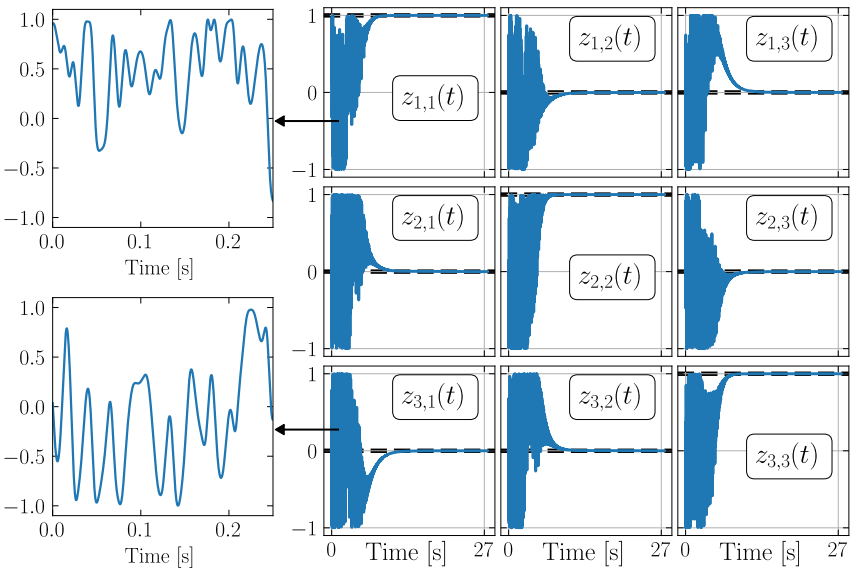


Figure 10.6. Synergistic Gradient-free Optimization Seeking on SO(3) via Geodesic Dithering.

Lemma 10.3. Let k_q satisfy the bound $|k_q| < \overline{k}^{(2)}$, with:

$$\overline{k}^{(2)} := \frac{\|X\|_F^{-1}}{\max\limits_{z \in \mathrm{SO}, \phi(z) \geq \gamma} |\alpha'(\phi(z) - \gamma)| \, \|\mathrm{grad}\ \phi|_z\|_F}, \tag{10.27}$$

and $\|X\|_F = \sqrt{\mathrm{tr}(X^\top X)}$. Then, $\mathcal{S}_q^{(2)}$ is a global diffeomorphism.

To illustrate the application of the zeroth-order hybrid dynamics $\mathcal{H}_0$ in SO(3), we consider the cost function $\phi^{(2)}(z) = \mathrm{tr}\left((I - z)A\right)$, where $A = \frac{3}{\sum_{i=1}^3 a_i} \mathrm{diag}\,(a)$, and $a = (11, 12, 13)$. It follows that $\mathrm{Crit}\phi^{(2)} = \{I\} \cup \bigcup_{i=1}^3 \{I + 2[e_i]_\times^2\}$, where $e_i \in \mathbb{R}^3$ denotes the standard basis vector with a 1 in the i-th position and zeros in the other entries, and where $[u]_\times : \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ is defined as

$$[u]_\times := \sqrt{2} u_1 b_1 + \sqrt{2} u_2 b_2 + \sqrt{2} u_3 b_3. \tag{10.28}$$

For this problem, we consider the threshold value $\gamma = 2 \in \left(\underline{\phi}^{(2)}, \phi_2^{(2)}\right)$, and we select the gains $k_1 = 0.15$ and $k_2 = -0.15$, let $X = [a/|a|]_\times$, $\alpha(r) = \frac{r^2}{2}$, and consider the family of

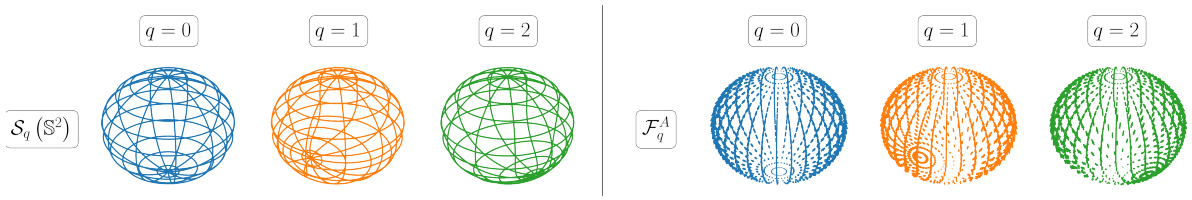


Figure 10.7. Left: Visualization of diffeomorphisms on $\mathbb{S}^2$. Right: Average vector fields derived from warped cost functions

functions $\mathcal{S}^{(2)} \coloneqq \{\mathcal{S}_q^{(2)}\}_{q\in\{1,2\}}$. With these choices, the value of the upper bound $\overline{k}^{(2)}$ in (10.27) is approximately 0.188, which means that $|k_q| < \overline{k}^{(2)}$, and hence, via Lemma 10.3, that the set $\mathcal{S}^{(2)}$ is a family of diffeomorphisms adapted to $\phi^{(2)}$. It can be computed that the family is δ-synergistic with gap $\delta = 0.0796$.

Using $\mathcal{S}^{(2)}$ and the global orthonormal frame $\{E_i\}_{i=1}^3$, we implement the HDS $\mathcal{H}_0$ and obtain the results shown in Figure 10.6. The figure shows the trajectories of the entries of the state z converging (globally) to the optimal values z_i^*, where $z^* = I$.

10.3 Extensions to Non-parallelizable Manifolds: Gradient-free Feedback Optimization on $\mathbb{S}^2$

In this section, we extend our results to manifolds M that are not parallelizable. In such cases, a unique global orthonormal frame is unavailable to define dithering vectors that are valid at every point on M. To address this issue, we employ local orthonormal frames and we introduce a suitable switching mechanism between them to cover M. This mechanism ensures that dithering vectors are always available for real-time exploration. Since the constructions of the controllers in non-parallelizable manifolds are highly dependent on the manifold, we focus our attention on the 2-dimensional sphere $\mathbb{S}^2 \coloneqq \{z \in \mathbb{R}^3 \ : \ z^\top z = 1\}$. However, we stress that the proposed methodology can be extended to other compact non-parallelizable manifolds.

First, we introduce two local orthonormal frames $\{E_{i,p}\}_{i=1}^2$, $p \in \mathcal{P} \coloneqq \{1, 2\}$, which will later be used to generate suitable dithering vector fields. Specifically, inspired by [264],

we use local coordinate frames established through the stereographic projection maps:

$$\varphi_1 : U_1 \coloneqq \mathbb{S}^2 \setminus \{N\} \to \mathbb{R}^2,\ z \mapsto \frac{1}{1-z_3}\,(z_1, z_2)\,, \tag{10.29a}$$

$$\varphi_2 : U_2 \coloneqq \mathbb{S}^2 \setminus \{S\} \to \mathbb{R}^2,\ z \mapsto \frac{1}{1+z_3}(z_1, z_2), \tag{10.29b}$$

where $N \coloneqq (0,0,1)$ and $S \coloneqq (0,0,-1)$, denote the north and south pole of $\mathbb{S}^2$, respectively. The stereographic projections constitute homeomorphisms onto their images, and their inverse functions are given by [263, Ex. 4.1]:

$$\varphi_1^{-1}(u_1, u_2) = \frac{1}{1+|u|^2}(2u_1, 2u_2, |u|^2 - 1) \tag{10.30a}$$

$$\varphi_2^{-1}(u_1, u_2) = \frac{1}{1+|u|^2}(2u_1, 2u_2, 1 - |u|^2). \tag{10.30b}$$

Using (10.29)-(10.30), we define the vector fields $E_{i,p}(z) \coloneqq d(\varphi_j^{-1})_{\varphi_j(z)}(e_i)$ for all $z \in U_j$, $i \in \{1,2\}$, and $p \in \mathcal{P} \coloneqq \{1,2\}$, where e_i denotes the i-th canonical basis vector in $\mathbb{R}^2$. Unwrapping definitions, we obtain:

$$E_{1,1}(z) = \begin{pmatrix} 1 - z_3 - z_1^2 \\ z_1 z_2 \\ (1-z_3)z_1 \end{pmatrix},\ E_{1,2}(z) = \begin{pmatrix} z_1 z_2 \\ 1 - z_3 - z_2^2 \\ (1-z_3)z_2 \end{pmatrix},$$

$$E_{2,1}(z) = \begin{pmatrix} 1 + z_3 - z_1^2 \\ -z_1 z_2 \\ -z_1(1+z_3) \end{pmatrix},\ E_{2,2}(z) = \begin{pmatrix} -z_1 z_2 \\ 1 + z_3 - z_2^2 \\ -z_2(1+z_3) \end{pmatrix}.$$

Next, for each local orthonormal frame $\{E_{i,p}\}_{i=1}^2$, $p \in \mathcal{P}$, we define a corresponding dithering vector field used for the purpose of real-time exploration of M:

$$\mathcal{D}_p = \sum_{i=1}^{2} \hat{\chi}_i E_{i,p}. \tag{10.31}$$

Additionally, for each dithering vector field $\mathcal{D}_p$, and given a δ-gap synergistic family of diffeomorphisms $\mathcal{S} = \{S_q\}_{q\in\mathcal{Q}}$ adapted to a cost function ϕ, we define a family of vector fields $\{\hat{f}_{q,p}(\cdot,\chi)\}_{q\in\mathcal{Q}}$ suitable for exploitation of the information of ϕ learned during the exploration. Specifically, given $q \in \mathcal{Q}$ and $p \in \mathcal{P}$, we let

$$\hat{f}_{q,p}(z,\chi) \coloneqq \frac{2}{\varepsilon_a}\tilde{\phi}_q\left(\exp_z\left(\varepsilon_a \mathcal{D}_p(z)\right)\right)\mathcal{D}_p(z), \tag{10.32}$$

where the vector of oscillating amplitudes $\chi \in \mathbb{T}^2$, and the warped cost function $\tilde{\phi}_q = \phi \circ S_q$ are as defined in Section 10.2. Finally, we modify the zeroth-order hybrid dynamics $\mathcal{H}_0$ to incorporate the switching between frames. The new zeroth-order hybrid system, termed $\tilde{\mathcal{H}}_0$, incorporates an additional logic state $p \in \mathcal{P}$ and implements a hysteresis-based switching mechanism dependent on z. The mechanism enables the robust transition between the families of vector fields $\{\hat{f}_{q,p}(z,\chi)\}_{q\in\mathcal{Q}}$, and ensures that the orthonormal frame associated with the selected family satisfies the condition $\text{span}\left(\{E_{i,p}\}_{i=1}^2\right) = T_z\mathbb{S}^2$ for the current value of z. To the best of our knowledge, this approach has not been studied before in the context of zeroth-order optimization and extremum-seeking.

To define the hysteresis-based switching, we first let $r > 1$, and define the open sets $C_p \coloneqq \varphi_p^{-1}\left(r\mathbb{B}^\circ\right)$. By using the definitions of φ_i and φ_i^{-1}, it follows that $C_1 \cup C_2 = \mathbb{S}^2$, and that $\text{span}(\{E_{i,p}\}_{i=1}^2) = T_z\mathbb{S}^2$ for all $z \in \overline{C}_p$ and all $p \in \mathcal{P}$. Using these sets, we characterize the new dynamics $\tilde{\mathcal{H}}_0$, which describe the evolution of the state $\tilde{y} \coloneqq (z, q, \chi, p) \in \mathbb{S}^2 \times \mathcal{Q} \times \mathbb{T}^2 \times \mathcal{P}$, and have data $\tilde{\mathcal{H}}_0 = \{\tilde{C}_0,\ \tilde{F}_0,\ \tilde{D}_0,\ \tilde{G}_0\}$, with continuous-time dynamics:

$$\tilde{y} \in \tilde{C}_0, \quad \dot{\tilde{y}} = \tilde{F}_0(\tilde{y}) \coloneqq \begin{pmatrix} -\hat{f}_{q,p}(z,\chi) \\ 0 \\ \frac{2\pi}{\varepsilon_d}\Psi(\omega)\chi \\ 0 \end{pmatrix}, \tag{10.33}$$

where $\Psi(\omega) \in \mathbb{R}^{4\times 4}$ and $\omega \in \mathbb{R}^2$ are as defined in Section 10.2. The flow set is defined by

$\tilde{C}_0 \coloneqq \tilde{C}_{0,1} \cup \tilde{C}_{0,2}$, where, for all $p \in \mathcal{P}$, we let

$$\tilde{C}_{0,p} \coloneqq \left\{(z, q, \chi) \in \overline{C}_p \times \mathcal{Q} \times \mathbb{T}^2 : \left(\tilde{\phi}_q - m\right)(z) \leq \delta\right\} \times \{p\}.$$

The jump set $\tilde{\mathcal{D}}_0$ is constructed as the union of two sets: 1) $\tilde{D}_{0,d}$, which enables switching between the families of vector fields $\{\hat{f}_{q,p}(\cdot, \chi)\}_{q \in \mathcal{Q}}$, and 2) $\tilde{D}_{0,s}$, which enables the synergistic switching between vector fields within the selected family, akin to the methodology outlined in Section 10.2. Specifically, we let $\tilde{D}_0 \coloneqq \tilde{D}_{0,d} \cup \tilde{D}_{0,s}$, where $\tilde{D}_{0,d} \coloneqq \bigcup_{p \in \{1,2\}} \tilde{D}_{p,d}$, and

$$\begin{aligned}
\tilde{D}_{p,d} &\coloneqq (\mathbb{S}^2 \setminus C_p) \times \mathcal{Q} \times \mathbb{T}^2 \times \{p\}, \quad \forall p \in \mathcal{P} \\
\tilde{D}_{0,s} &\coloneqq \{(z, q, \chi) \in \mathbb{S}^2 \times \mathcal{Q} \times \mathbb{T}^2 : \left(\tilde{\phi}_q - m\right)(z) \geq \delta\} \times \mathcal{P}.
\end{aligned}$$

The jump map describing the switches of p is given by $\tilde{G}_{0,d}(\tilde{y}) \coloneqq (z, q, \chi, 3 - p),\ \forall \tilde{y} \in \tilde{D}_{0,d}$, which updates the current frame used for the purpose of dithering. On the other hand, the jump map describing the switches of q is given by $\tilde{G}_{0,s}(\tilde{y}) \coloneqq \{z\} \times h(z) \times \{(\chi, p)\},\ \forall \tilde{y} \in \tilde{D}_{0,s}$, where h is the set-valued map defined in (10.14). Using these maps, the overall jump map of the HDS $\tilde{\mathcal{H}}_0$ is given by:

$$\tilde{G}_0(\tilde{y}) \coloneqq \begin{cases} \tilde{G}_{0,s}(\tilde{y}) & \forall \tilde{y} \in \tilde{D}_{0,s} \setminus \tilde{D}_{0,d} \\ \tilde{G}_{0,d}(\tilde{y}) & \forall \tilde{y} \in \tilde{D}_{0,d} \setminus \tilde{D}_{0,s} \\ \tilde{G}_{0,s}(\tilde{y}) \cup \tilde{G}_d(y) & \forall \tilde{y} \in \tilde{D}_{0,d} \cap \tilde{D}_{0,s} \end{cases}.$$

By leveraging our standing assumptions, the following theorem extends the global results of Theorem 10.1 to the non-parallelizable manifold $\mathbb{S}^2$.

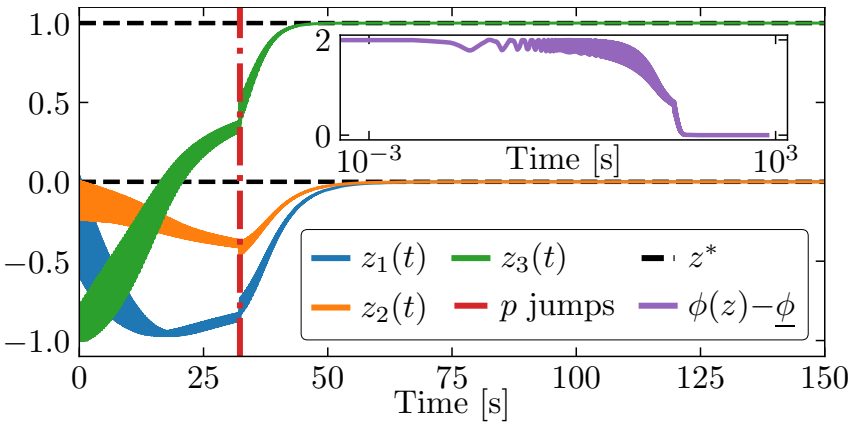


Figure 10.8. Evolution of the coordinates of z under the Synergistic Gradient-Free Optimization Seeking dynamics on $\mathbb{S}^2$.

Theorem 10.3. Consider the zeroth-order hybrid dynamics $\tilde{\mathcal{H}}_0$ and let the vector of frequencies ω in (10.33) satisfy condition (10.17). Then, the set $\mathcal{A} \times \mathcal{Q} \times \mathbb{T}^2 \times \mathcal{P}$ is GP-AS as $(\varepsilon_d, \varepsilon_a) \to 0^+$.

To illustrate the performance of $\tilde{\mathcal{H}}_0$ in $\mathbb{S}^2$, we synthesize the algorithms by using the parameterized transformation $\mathcal{S}_q^{(3)} : \mathbb{S}^2 \to \mathbb{S}^2$ defined as:

$$\mathcal{S}_q^{(3)}(z) = \mathbf{1}_{\{\phi(z) \leq \gamma\}} z + \mathbf{1}_{\{\phi(z) > \gamma\}} e^{k_q \alpha(\phi(z)-\gamma)[u]_\times} z, \tag{10.34}$$

with $k_q \in \mathbb{R}$, $X \in T_I \mathrm{SO}(3)$ and α as defined in Section 10.2.3. Note that $\mathcal{S}_q^{(3)}$ is identical to $\mathcal{S}_q^{(2)}$, except for the fact that its domain and codomain are now $\mathbb{S}^2$ instead of $\mathrm{SO}(3)$. The following Lemma extends the result of Lemma 10.3 to $\mathbb{S}^2$.

Lemma 10.4. Let k_q satisfy $|k_q| < \overline{k}^{(3)}$, where:

$$\overline{k}^{(3)} := \frac{1}{\max\left\{\left|\alpha'\left(\phi(z)-\gamma\right) d\phi_z\left(Xz\right)\right| : z \in \mathbb{S}^2, \phi(z) \geq \gamma\right\}}.$$

Then, $S_q^{(3)}$ is a diffeomorphism.

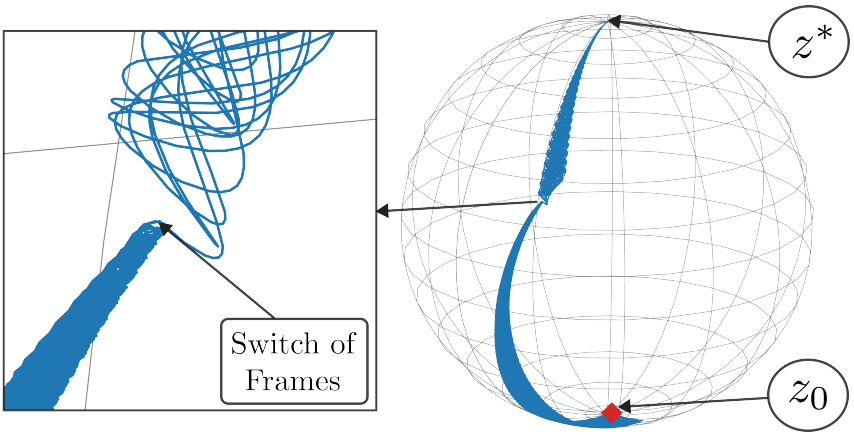


Figure 10.9. Synergistic Gradient-Free Optimization Seeking on $\mathbb{S}^2$ via Geodesic Dithering. The inset shows the moment when the system switches from one local frame to another.

For numerical verification, we consider the cost function $\phi^{(3)} : \mathbb{S}^2 \to \mathbb{S}^2$ defined by $\phi^{(3)}(z) = 1 - z_3$. We choose the threshold value $\gamma = 1$, the gains $k_1 = \frac{1}{2}, k_2 = -\frac{1}{2}$, the matrix $X = [u]_\times \in \mathrm{SO}(3)$, where $u = (0, 1, 0) \in \mathbb{R}^3$ and $[u]_\times$ is as defined in Section 10.2.3, and let $\alpha(r) = r^2$. With this data, we define the family of tranformations $\mathcal{S}^{(2)} \coloneqq \left\{S^{(2)}_{q,u}\right\}_{q\in\{1,2\}}$. Since $|k_q| < \overline{k}^{(3)} = 1$ for all $q \in \mathcal{Q} \coloneqq \{1, 2\}$, via Lemma 10.4, $\mathcal{S}^{(2)}$ is a family of diffeomorphisms. In fact, by Lemma 10.1, $\mathcal{S}^{(2)}$ constitutes a δ-synergistic family of diffeomorphisms adapted to ϕ with gap $\delta < \frac{1}{4}$. Figure 10.7 shows a visualization of the diffeomorphisms in this family with the choice $k_1 = \frac{1}{2}, k_2 = -\frac{1}{2}$. The figure also shows the vector fields obtained from the warped cost functions. We stress that such diffeomorphisms can be constructed using only mild qualitative knowledge of ϕ, namely, under a suitable choice of γ, which can be seen as a tunable parameter of the algorithm. In Figure 10.8, we show the trajectory of the coordinates of the state z and indicate when the local frame used for the dithering switches by showing the moments when the state p jumps. In Figure 10.9 we show the trajectory evolving on the sphere. As observed, the state z converges to the global minimizer $z^* = (0, 0, 1)$, while escaping the critical point $z' = (0, 0, -1)$.

In this chapter, we introduced a novel class of zeroth-order hybrid algorithms for

the global solution of gradient-free optimization problems on smooth, compact, and boundaryless manifolds. These algorithms combine continuous-time dynamics and discrete-time dynamics to achieve robust *global* practical stability of the optimizer of a smooth cost function accessible only via measurements or evaluations. The proposed approach overcomes topological obstructions that prevent the solution of this problem using algorithms modeled by smooth ODEs. We characterized the stability and robustness of the algorithms using tools from the theory of hybrid dynamic inclusions. Future research will explore tracking problems in time-varying optimization settings, as well as the incorporation of dynamic plants in the loop. A completely coordinate-free formulation of the hybrid algorithms, and the development of accelerated dynamics and single-point algorithms, are also future research directions.

Acknowledgements

Chapter 10, in full, is a reprint of the material as it appears in "Robust Global Optimization on Smooth Compact Manifolds via Hybrid Gradient-Free Dynamics" [10]. The dissertation author was the first author of this paper.

Chapter 11

Hybrid Kapitza's Pendulum

Kapitza's pendulum, named after Soviet physicist Pyotr Kapitza, is a deceptively simple mechanical system that has captivated physicists and control theorists since its initial analysis in 1951 [265]. This system, consisting of a pendulum with its pivot point subjected to rapid oscillations (see Figure 11.1), exhibits a counterintuitive behavior: under certain conditions, the typically unstable inverted position becomes stable. This phenomenon exemplifies the rich dynamics that can emerge from nonlinear systems and has found applications ranging from atomic physics to control engineering [266, 267]. Moreover, the study of Kapitza's pendulum has inspired the development of vibrational control techniques for a wide range of systems [268]. This chapter explores Kapitza's pendulum,

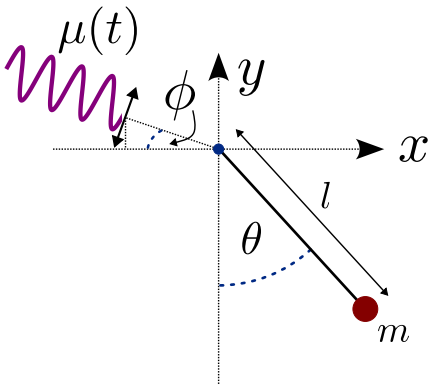


Figure 11.1. Pendulum with oscillating pivot at a fixed angle ϕ

its mathematical foundations, and novel approaches to achieving global stabilization through hybrid control. We begin by examining the equations of motion with a crucial

modification to the regular Kapitza's pendulum: arbitrary directions of oscillations at the pivot. We then discuss the challenges in achieving global asymptotic stability and propose a novel approach to overcome those challenges by using multiple oscillating directions combined with a suitably designed hybrid dynamical system.

11.1 Equations of Motion

In this section, we derive the equations of motion using the Euler-Lagrange formalism. While the traditional analysis of Kapitza's pendulum focuses on vertical oscillations, we extend our analysis to consider multiple directions of oscillation. This generalization allows us to explore a broader range of dynamic behaviors and lays the foundation for the hybrid control approach presented in Section 11.3.

- *Position:* $x(\theta) = l\sin(\theta) + \mu(t)\sin(\phi), \quad y(\theta) = -l\cos(\theta) + \mu(t)\cos(\phi).$
- *Potential Energy:* $U(\theta, t) = mgy(\theta, t) = -mg\left[l\cos(\theta) - \mu(t)\cos(\phi)\right]$
- *Kinetic Energy:*

$$\begin{aligned} K(\theta, \dot{\theta}, t) &= \frac{m}{2}\left(\dot{x}^2(\theta) + \dot{y}^2(\theta, t)\right) \\ &= \frac{m}{2}\left(\left[l\cos(\theta)\dot{\theta} + \dot{\mu}\sin(\phi)\right]^2 + \left[l\sin(\theta)\dot{\theta} + \dot{\mu}\cos(\phi)\right]^2\right) \\ &= \frac{m}{2}\Big(l^2\cos^2(\theta)\dot{\theta}^2 + l^2\sin^2(\theta)\dot{\theta}^2 + \dot{\mu}^2 + \\ &\qquad 2l\dot{\mu}\dot{\theta}\left[\sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi)\right]\Big) \\ &= \frac{m}{2}\left(l^2\dot{\theta}^2 + \dot{\mu}^2 + 2l\sin(\theta + \phi)\dot{\theta}\dot{\mu}\right). \end{aligned}$$

- *Lagrangian:*

$$\mathcal{L}(\theta, \dot{\theta}, t) = K(\theta, \dot{\theta}, t) - U(\theta, t)$$

$$= \frac{m}{2}\left(l^2\dot{\theta}^2 + \dot{\mu}^2 + 2l\sin(\theta+\phi)\dot{\theta}\dot{\mu}\right) + mg\left[l\cos(\theta) - \mu(t)\cos(\phi)\right].$$

- *Disipation potential:*

  We consider the dissipation function

$$D(v) = \frac{\gamma}{2}v^2$$

  where v is the velocity of the bob, i.e., $v = |(\dot{x}, \dot{y})|$, and $\gamma \in \mathbb{R}_{\geq 0}$. Now, we have that

$$\begin{aligned} v^2 &= \dot{x}^2 + \dot{y}^2 \\ &= l^2\dot{\theta}^2 + 2l\sin(\theta+\phi)\dot{\theta}\dot{\mu} + \dot{\mu}^2, \end{aligned}$$

  and, thus, the friction force is given by

$$\begin{aligned} F_{\text{fr}} &= -\frac{\partial D(v)}{\partial \dot{\theta}} \\ &= -\gamma\left(l^2\dot{\theta} + l\sin(\theta+\phi)\dot{\mu}\right) \end{aligned} \tag{11.1}$$

- *Euler-Lagrange Equations (With External Forces - Friction):*

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \dot{\theta}} &= ml^2\dot{\theta} + ml\sin(\theta+\phi)\dot{\mu} \\ \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\theta}} &= ml^2\ddot{\theta} + ml\cos(\theta+\phi)\dot{\theta}\dot{\mu} + ml\sin(\theta+\phi)\ddot{\mu} \\ \frac{\partial \mathcal{L}}{\partial \theta} &= -mgl\sin(\theta) + ml\cos(\theta+\phi)\dot{\theta}\dot{\mu}. \end{aligned}$$

  By using the fact that $\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{\theta}} = \frac{\partial \mathcal{L}}{\partial \theta} + F_{\text{fr}}$, we are now prepared to write down the equations of motion:

$$l\ddot{\theta} = -g\sin(\theta) - \ddot{\mu}\sin(\theta+\phi) - \frac{\gamma}{m}\left(l\dot{\theta} + \sin(\theta+\phi)\dot{\mu}\right)$$

$$\implies \ddot{\theta} = -\frac{g}{l}\sin(\theta) - \frac{1}{l}\sin(\theta+\phi)\ddot{\mu} - \frac{\gamma}{lm}\left(l\dot{\theta} + \sin(\theta+\phi)\dot{\mu}\right). \tag{11.2}$$

To analyze the stability properties of the inverted position $\theta = \pi$ under high-frequency oscillations at the pendulum's pivot, we employ averaging theory for hybrid dynamical systems [269]. Our goal is to transform the system's continuous-time evolution so that it admits a suitable average. To this end, we begin by setting $\mu(t) = a\sin(\omega t)$ and assuming $a/l \ll 1$ and $\omega_0/\omega \ll 1$, where $\omega_0 = \sqrt{g/l}$ is the pendulum's natural frequency. We then introduce the following parameters:

$$\varepsilon \coloneqq a/l, \quad \alpha \coloneqq \omega_0 l/\omega a, \quad \beta \coloneqq \gamma/m\omega_0.$$

Using these definitions, it follows that $\omega_0/\omega = \alpha\varepsilon$. With these substitutions, equation (11.2) transforms into:

$$\begin{aligned}\ddot{\theta} &= -\omega_0^2\sin(\theta) + \varepsilon\omega^2\sin(\omega t+\eta)\sin(\theta+\phi) - \frac{\omega_0}{l}\beta\left(l\dot{\theta} + a\omega\cos(\omega t+\eta)\sin(\theta+\phi)\right)\\ &= -\omega_0^2\sin(\theta) + \varepsilon\omega^2\sin(\omega t+\eta)\sin(\theta+\phi) - \omega_0\beta\dot{\theta} - \varepsilon\omega_0\beta\omega\cos(\omega t+\eta)\sin(\theta+\phi).\end{aligned}$$

Letting $\tau \coloneqq \omega t$, this further reduces to

$$\begin{aligned}\omega^2\frac{d^2\theta}{d\tau^2} &= -\omega_0^2\sin(\theta) + \varepsilon\omega^2\sin(\omega t+\eta)\sin(\theta+\phi) - \omega_0\beta\omega\frac{d\theta}{d\tau}\\ &\qquad - \varepsilon\omega_0\beta\omega\cos(\omega t+\eta)\sin(\theta+\phi)\\ \implies \frac{d^2\theta}{d\tau^2} &= -\frac{\omega_0^2}{\omega^2}\sin(\theta) + \varepsilon\sin(\omega t+\eta)\sin(\theta+\phi) - \frac{\omega_0}{\omega}\beta\frac{d\theta}{d\tau}\\ &\qquad - \varepsilon\frac{\omega_0}{\omega}\beta\cos(\tau+\eta)\sin(\theta)\\ \implies \frac{d^2\theta}{d\tau^2} &= -\alpha^2\varepsilon^2\sin(\theta) + \varepsilon\sin(\tau+\eta)\sin(\theta+\phi)\\ &\qquad - \alpha\beta\varepsilon\frac{d\theta}{d\tau} - \alpha\beta\varepsilon^2\cos(\tau+\eta)\sin(\theta+\phi).\end{aligned} \tag{11.3}$$

Now, by applying the near-identity transformation defined by:

$$x_1 = \theta, \quad x_2 = \frac{1}{\varepsilon}\frac{d\theta}{d\tau} + \cos(\tau+\eta)\sin(\theta+\phi),$$

from (11.3) we obtain

$$\begin{aligned}
\frac{dx_1}{d\tau} &= \frac{d\theta}{d\tau} \\
&= \varepsilon\left(x_2 - \cos(\tau+\eta)\sin(x_1+\phi)\right)
\end{aligned}$$

$$\begin{aligned}
\frac{dx_2}{d\tau} &= \frac{1}{\varepsilon}\frac{d^2\theta}{d\tau^2} - \sin(\tau+\eta)\sin(\theta+\phi) + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= \frac{1}{\varepsilon}\Bigg[- \alpha^2\varepsilon^2\sin(\theta) + \varepsilon\sin(\tau+\eta)\sin(\theta+\phi) - \alpha\beta\varepsilon\frac{d\theta}{d\tau} \\
&\qquad - \alpha\beta\varepsilon^2\cos(\tau+\eta)\sin(\theta+\phi)\Bigg] - \sin(\tau+\eta)\sin(\theta+\phi) \\
&\qquad + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= \frac{1}{\varepsilon}\left[-\alpha^2\varepsilon^2\sin(\theta) - \alpha\beta\varepsilon\frac{d\theta}{d\tau} - \alpha\beta\varepsilon^2\cos(\tau+\eta)\sin(\theta+\phi)\right] \\
&\qquad + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= -\varepsilon\alpha^2\sin(\theta) - \alpha\beta\frac{d\theta}{d\tau} - \alpha\beta\varepsilon\cos(\tau+\eta)\sin(\theta+\phi) + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= -\varepsilon\alpha^2\sin(\theta) - \alpha\beta\varepsilon\left[\frac{1}{\varepsilon}\frac{d\theta}{d\tau} + \cos(\tau+\eta)\sin(\theta+\phi)\right] + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= \varepsilon\left(-\alpha^2\sin(x_1) - \alpha\beta x_2\right) + \cos(\tau+\eta)\cos(\theta+\phi)\frac{d\theta}{d\tau} \\
&= \varepsilon\left(-\alpha^2\sin(x_1) - \alpha\beta x_2\right) \\
&\qquad + \cos(\tau+\eta)\cos(\theta+\phi)\varepsilon\left(x_2 - \cos(\tau+\eta)\sin(\theta+\phi)\right) \\
&= \varepsilon\Bigg(- \alpha\beta x_2 - \alpha^2\sin(x_1) + x_2\cos(x_1+\phi)\cos(\tau+\eta) \\
&\qquad - \sin(x_1+\phi)\cos(x_1+\phi)\cos^2(\tau+\eta)\Bigg).
\end{aligned}$$

Therefore, we have

$$\frac{dz}{d\tau} = \begin{pmatrix} \frac{d}{d\tau}x \\ \frac{d}{d\tau}\phi \\ \frac{d}{d\tau}\chi \end{pmatrix} = \tilde{F}(z) \coloneqq \begin{pmatrix} \varepsilon F_x(x,\phi,\chi) \\ 0 \\ F_\chi(\chi) \end{pmatrix} \tag{11.4}$$

with $z \coloneqq (x,\phi,\chi) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{S}^1$, and where the continuous vector fields $F_x : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{S}^1 \to \mathbb{R}^2$ and $F_\chi : \mathbb{S}^1 \to \mathbb{S}^1$ are defined by:

$$F_x(x,\phi,\chi) \coloneqq \begin{pmatrix} x_2 - \chi_1 \sin(x_1+\phi) \\ -\alpha\beta x_2 - \alpha^2 \sin(x_1) + x_2 \cos(x_1+\phi)\chi_1 - \sin(x_1+\phi)\cos(x_1+\phi)\chi_1^2 \end{pmatrix}$$

$$F_\chi(\chi) \coloneqq \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \chi.$$

The flow map $\tilde{F}$ in (11.4) lends itself to averaging in the context of hybrid dynamical systems. As we will explore in Section 11.3, analysis of its first-order average allows us to strategically select oscillating directions $\varphi(q)$ and partitions of the circle $\mathbb{S}^1$ to ensure that, on average, the system behaves like a heavy-ball system seeking the minimum of a carefully designed effective potential function.

The following section focuses on designing this effective potential so that it has a global minimum at the naturally unstable position $\theta = \pi$. This approach transforms the stabilization problem into an optimization challenge over an engineered potential landscape. The key challenge lies in carefully selecting oscillation directions $\varphi(q)$ and defining regional boundaries on $\mathbb{S}^1$ where these directions of oscillation are implemented to create a globally invex effective potential, while respecting physical constraints specified by the equations of motion in (11.4).

11.2 Designing a Physically-Informed Invex Effective Potential

First, let

$$\Theta_q \coloneqq \begin{cases} [-5\pi/8, -3\pi/8] & \text{if } q = -3 \\ [-3\pi/8, -\pi/8] & \text{if } q = -2 \\ [-\pi/8, 0] & \text{if } q = -1 \\ [-\pi, -5\pi/8] \cup [5\pi/8, \pi] & \text{if } q = 0 \\ [0, \pi/8] & \text{if } q = 1 \\ [\pi/8, 3\pi/8] & \text{if } q = 2 \\ [3\pi/8, 5\pi/8] & \text{if } q = 3 \end{cases}, \qquad \varphi(q) \coloneqq \begin{cases} 3\pi/4 & \text{if } q = -3 \\ \pi/2 & \text{if } q = -2 \\ \pi/4 & \text{if } q = -1 \\ 0 & \text{if } q = 0 \\ 3\pi/4 & \text{if } q = 1 \\ \pi/2 & \text{if } q = 2 \\ \pi/4 & \text{if } q = 3 \end{cases}$$

$$U_q(\theta) \coloneqq -\alpha^2 \cos(\theta) + \frac{1}{4}\sin^2(\theta + \varphi(q)) + c_q, \tag{11.5}$$

where $c_q \in \mathbb{R}$ and $q \in Q \coloneqq [-3, 3] \cap \mathbb{Z}$. Moreover, define the global potential energy function by:

$$U(x_1, q) \coloneqq \sum_{p \in Q} U_p(x_1)\mathbf{1}_{\{x_1 \in \Theta_p\}} - U_q(x_1)\mathbf{1}_{\{x_1 \in \vartheta_q\}}, \tag{11.6}$$

where

$$\vartheta_p \coloneqq \begin{cases} \{0\} & \text{if } p = 0 \\ \Theta_p \cap \Theta_{(p+\text{sign}(p)) \bmod 4} & \text{otherwise} \end{cases},$$

and where the set of constants $\{c_q\}_{q \in Q}$ is chosen such that $U(\theta) \coloneqq \sum_{q \in Q} U_q(\theta)\mathbf{1}_{\{\theta \in \Theta_q\}}$ is continuous and satisfies $U(\pm\pi) = 0$. The negative term on the right of (11.6) is used to

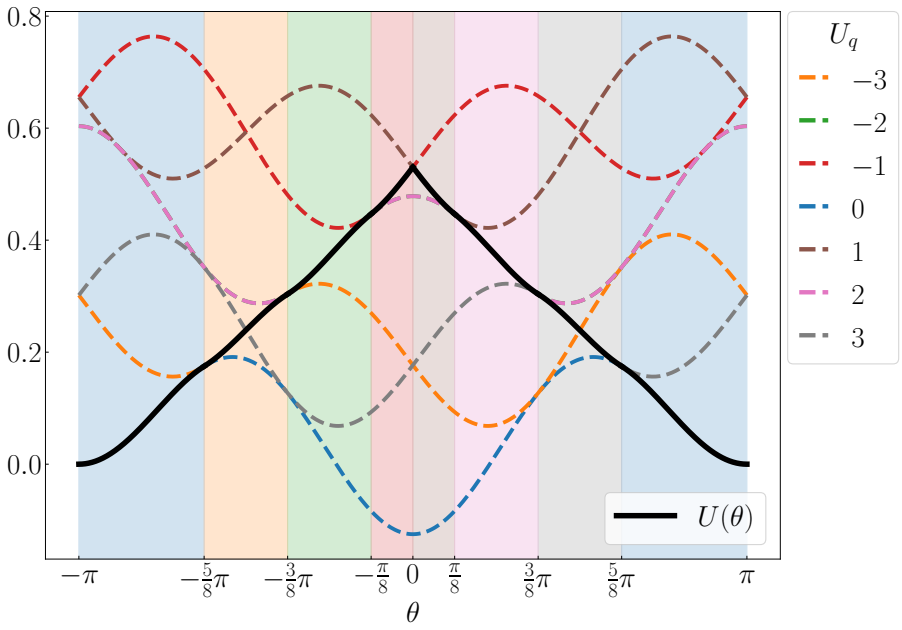


Figure 11.2. Plot of the Designed Effective Potential

remove duplicates values from the sum whenever $x_1 \in \{\vartheta_q\}_{q\in Q}$. Now, note that

$$\nabla U(x_1, q) = \nabla U_q(x) \quad \forall (x, q) \in \bigcup_{q\in Q} \Theta_q \times \{q\}\,.$$

We now introduce a sufficient condition on α to guarantee that the potential defined $U(x_1, q)$, with the directions of oscillation introduced in (11.5), is an invex potential.

Lemma 11.1. If $0 < \alpha^2 < \frac{1}{2}\cos(3\pi/8)$, then

$$\nabla U(x_1, q) = \nabla U_q(x_1) = 0 \iff (\theta, q) \in \{-\pi, \pi\} \times \{0\}\,, \text{ and}$$
$$\langle \nabla U(x_1, q),\, x_1 \rangle < 0 \quad \forall (x_1, q) \in \bigcup_{q\in Q} (\Theta_q \setminus \{-\pi, 0, \pi\}) \times \{q\}\,.$$

Via Lemma 11.1 we obtain that $U(x_1, q) > U(\pm\pi, 0)$ for all (x_1, q) in

$$\bigcup_{q\in Q} (\Theta_q \setminus \{-\pi, \pi\}) \times \{q\} \tag{11.7}$$

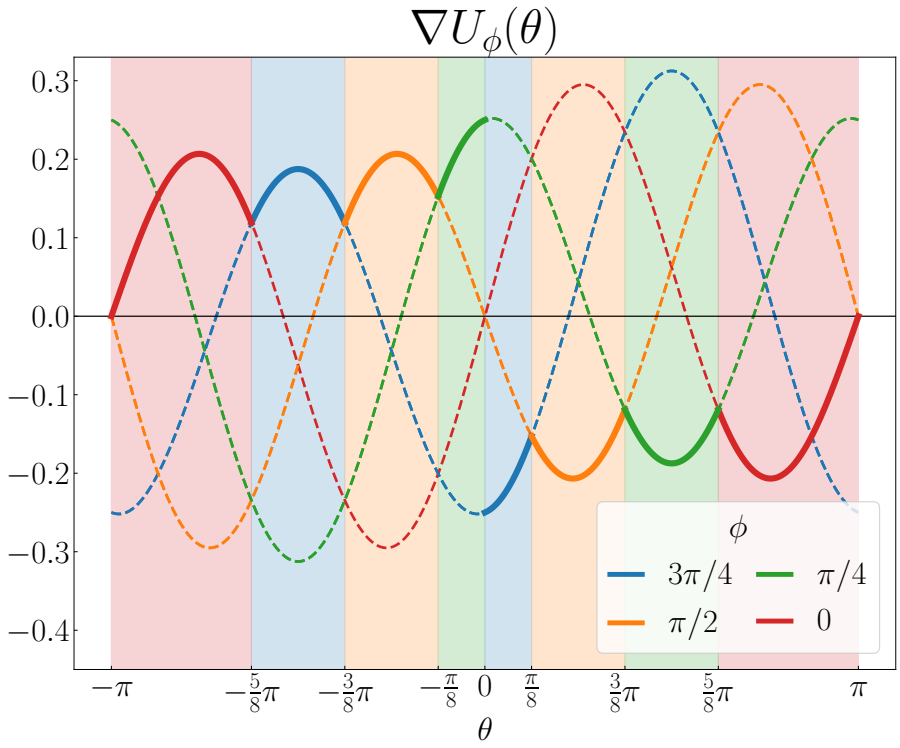


Figure 11.3. Plot of the Designed Gradient Vector Field

Indeed, we have that

$$\begin{aligned} U(x_1, q) &= U(-\pi, 0) + \int_{-\pi}^{\theta} \nabla U_q(\vartheta) \mathbf{1}_{\{\vartheta \in \Theta_q\}} d\vartheta \quad \theta \in (-\pi, 0] \\ &> U(-\pi, 0) \quad \text{for all } x_1 \in (-\pi, 0) \\ U(\theta) &= U(\pi, 0) + \int_{\pi}^{\theta} \nabla U_q(\vartheta) \mathbf{1}_{\{\vartheta \in \Theta_q\}} d\vartheta \quad \theta \in [0, \pi) \\ &= U(\pi, 0) - \int_{\theta}^{\pi} \nabla U_q(\vartheta) \mathbf{1}_{\{\vartheta \in \Theta_q\}} d\vartheta \\ &> U(\pi) \quad \text{for all } \theta \in (0, \pi). \end{aligned}$$

Figures 11.2 and 11.3 present a visualization of the designed potential $U(x_1, q)$ and its gradient $\nabla_{x_1} U(x_1, q)$.

11.3 Hybrid Vibrational Controller

The design of the invex potential in the previous section provides a theoretical foundation for global stabilization of the Kapitza pendulum. However, implementing a control strategy that robustly achieves this stabilization faces a fundamental challenge rooted in the topology of the system's state space. The state space of Kapitza's pendulum is given by $\mathbb{S}^1 \times \mathbb{R}$, where $\mathbb{S}^1$ corresponds to the angular position and $\mathbb{R}$ to the angular velocity. This non-contractible state space poses a fundamental obstacle to global asymptotic stability via traditional control methods. For continuous, discontinuous, or time-varying feedback systems, the basin of attraction of an asymptotically stable equilibrium must be contractible [270, 271]—a condition incompatible with the pendulum's state space.

To overcome these topological limitations, we introduce a hybrid vibrational control approach that addresses these challenge by combining continuous-time vibrational control with discrete switching logic. It implements the oscillation directions defined by the invex potential design across different regions of $\mathbb{S}^1$. The strategy integrates continuous-time vibrational stabilization with discrete-time jumps, ensuring proper switching between oscillation modes as the pendulum traverses these regions. Crucially, it provides mechanisms to avoid Zeno behavior without relying on traditional dwell-time constraints, which are impractical in this context due to the potential for arbitrarily small flow intervals.

To explain the hybrid mechanism, we begin by considering the four specific oscillating directions presented in the design of the invex potential:

$$\phi_1 = 0, \quad \phi_2 = \frac{\pi}{4}, \quad \phi_3 = \frac{\pi}{2}, \quad \phi_4 = \frac{3\pi}{4} \tag{11.8}$$

We formulate our control strategy as a hybrid dynamical system in the τ-continuous time scale. The state of this system is given by $z \coloneqq (x, q, k, \chi)$, where $x = (x_1, x_2)$ represents the pendulum's angular position and velocity, respectively, $q \in Q \coloneqq [-3, 3] \cap \mathbb{Z}$ is a

discrete variable encoding the current region of operation, $k \in \mathcal{K} \coloneqq \mathbb{Z} \cap [-4, 4]$ stores information about the sign of the angular velocity upon entering a region, and χ represents the amplitude of the oscillation. The flow set C, and jump set D of the HDS are given by:

$$C = \bigcup_{q=-3}^{3} \Theta_q \times \mathbb{R} \times \{q\} \times K_q^{\uparrow}, \qquad K_q^{\uparrow} \coloneqq \begin{cases} \{-3, -1\} & \text{if } q = -3 \\ \{-2, 0\} & \text{if } q = -2 \\ \{-3, -1\} & \text{if } q = -1 \\ \{-4, 4\} & \text{if } q = 0 \\ \{1, 3\} & \text{if } q = 1 \\ \{0, 2\} & \text{if } q = 2 \\ \{3, -1\} & \text{if } q = 3 \end{cases}, \tag{11.9a}$$

$$D = \bigcup_{p \in Q} \vartheta_p \times \left(D_p^{\uparrow} \cup D_p^{\downarrow} \right), \tag{11.9b}$$

$$D_p^{\uparrow} \coloneqq \begin{cases} \mathbb{R}_{\geq 0} \times \{-1\} \times \{-3\} & \text{if } p = 0 \\ \{x_2 \in \mathbb{R} \ : \ x_2 \text{sign}(p) \geq 0\} \times \{p\} \times K_p^{\uparrow} & \text{otherwise} \end{cases}, \tag{11.9c}$$

$$D_p^{\downarrow} \coloneqq \begin{cases} \mathbb{R}_{\leq 0} \times \{1\} \times \{3\} & \text{if } p = 0 \\ \{x_2 \in \mathbb{R} \ : \ x_2 \text{sign}(p) \leq 0\} \times \{(p + \text{sign}(p)) \mod 4\} \times K_p^{\downarrow} & \text{otherwise} \end{cases}, \tag{11.9d}$$

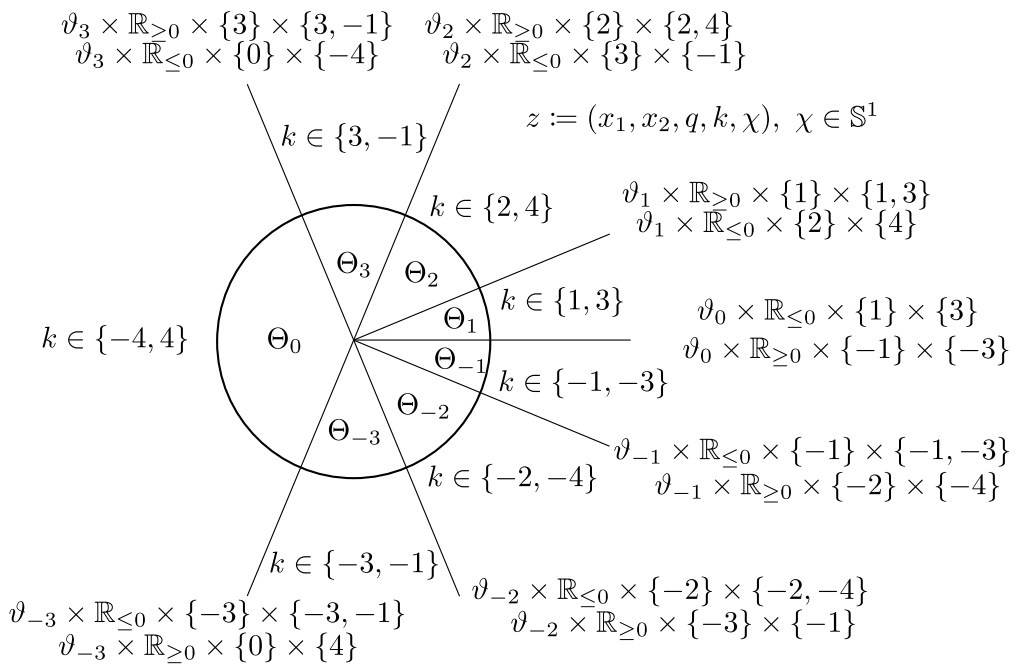


Figure 11.4. Depiction of jump and flow sets for the HDS $\mathcal{H}$ presented in (11.9)

$$K_p^{\downarrow} \coloneqq \begin{cases} \{4\} & \text{if } p = -3 \\ \{-1\} & \text{if } p = -2 \\ \{-4\} & \text{if } p = -1 \\ \{4\} & \text{if } p = 1 \\ \{-1\} & \text{if } p = 2 \\ \{-4\} & \text{if } p = 3 \end{cases}. \tag{11.9e}$$

Figure 11.4 shows a representation of the introduced flow and jump sets. The flow map F of the HDS is given by

$$F(z) = \begin{pmatrix} \varepsilon F_x(x, \varphi(q), \chi) \\ 0 \\ 0 \\ F_\chi(\chi) \end{pmatrix} \quad \forall z \in C \times \mathbb{S}^1, . \tag{11.9f}$$

Finally, the jump map G is defined as follows:

$$G(z) = \begin{cases} \begin{pmatrix} x_1 \\ x_2 \\ \overline{\operatorname{sign}}(x_2) \\ -\operatorname{sign}(q) \\ \chi \end{pmatrix} & \text{if } z \in (D \cap (\{\vartheta_0\} \times \mathbb{R} \times Q \times \mathcal{K})) \times \mathbb{S}^1 \\ \begin{pmatrix} x_1 \\ x_2 \\ \left(q + \overline{\operatorname{sign}}(x_2)\right) \mod 4 \\ q + \operatorname{sign}(q) \\ \chi \end{pmatrix} & \text{if } z \in \begin{pmatrix} D \cap \Big(\Big\{ x_1 \in \bigcup_{q \in Q \setminus \{0\}} \vartheta_q \Big\} \\ \times \mathbb{R} \times [Q \setminus \{0\}] \times \mathcal{K} \Big) \Big) \times \mathbb{S}^1 \end{pmatrix} \\ \begin{pmatrix} x_1 \\ x_2 \\ \overline{s}_+(x_2) \\ q + \operatorname{sign}(q) \\ \chi \end{pmatrix} & \text{if } z \in (D \cap (\{\vartheta_3\} \times \mathbb{R} \times \{0\} \times \mathcal{K})) \times \mathbb{S}^1 \\ \begin{pmatrix} x_1 \\ x_2 \\ \overline{s}_-(x_2) \\ q + \operatorname{sign}(q) \\ \chi \end{pmatrix} & \text{if } z \in (D \cap (\{\vartheta_{-3}\} \times \mathbb{R} \times \{0\} \times \mathcal{K})) \times \mathbb{S}^1 \end{cases}, \quad (11.9g)$$

where

$$\overline{s}_+(x_2) = \begin{cases} 3 & \text{if } x_2 < 0 \\ \{0,3\} & \text{if } x_2 = 0. \\ 0 & \text{if } x_2 > 0 \end{cases}$$

$$\overline{s}_-(x_2) = \begin{cases} -3 & \text{if } x_2 > 0 \\ \{0,-3\} & \text{if } x_2 = 0. \\ 0 & \text{if } x_2 < 0 \end{cases} \cdot$$

Remark 11.1. Note that for $q \in Q \setminus \{0\}$ we have that $\text{sign}(q) = \text{sign}\left(-\nabla U_q(x_1)\right)$ by design.

Lemma 11.2. The HDS $\mathcal{H} = (C \times \mathbb{S}^1, F, D \times \mathbb{S}^1, G)$, where C, F, D, and G are defined in (11.9), is well posed.

Lemma 11.3. Solutions to $\mathcal{H}$ experience at most 2 consecutive jumps between intervals of flow.

We will use averaging techniques for hybrid dynamical systems [272] to prove the main results of this paper. To this end, recall that the system

$$\dot{\chi} = F_\chi(\chi)$$

generates sinusoid components $(\chi_1, \chi_2) = \chi$ with shifted phase $\pi/2$ and initial phase η depending on the initial conditions $(\chi_1(0), \chi_2(0))$. Thus, $\mathcal{H}$ admits the HDS $\mathcal{H}_{av} :=$

(F_{av}, C, G_{av}, D) as its average system, with G_{av} as defined in (I.1),

$$F_{av}(x,q) \coloneqq \begin{pmatrix} x_2 \\ -\alpha\beta x_2 - \alpha^2 \sin(x_1) + \frac{1}{2}\sin(x_1+\varphi(q))\cos(x_1+\varphi(q)) \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} x_2 \\ -\nabla U_q(x_1) - \alpha\beta x_2 \\ 0 \end{pmatrix}, \tag{11.10}$$

and where U_q was defined in (11.5). Next, we introduce two lemmas that characterize the solutions of $\mathcal{H}_{av}$.

Lemma 11.4. Every maximal solution of $\mathcal{H}_{av}$ is complete.

Lemma 11.5. For any $p \in Q \setminus \{-3, 3\}$ and any solution $y : \mathrm{dom}(y) \to C \cup D$ to $\mathcal{H}_{av}$ with

$$y(0,0) \in G_{av}\left(\vartheta_p \times D_p^{\uparrow}\right)$$

there exists $T > 0$ such that $y(T,0) \in \vartheta_{p^\uparrow} \times D_{p^\uparrow}^{\uparrow}$, with $p^\uparrow = p + \mathrm{sign}(p)$.

The following Lemma follows directly by the same ides of the proof of Lemma 11.5.

Lemma 11.6. For any solution $y : \mathrm{dom}(y) \to C \cup D$ to $\mathcal{H}_{av}$ with

$$y(0,0) \in G_{av}\left(\vartheta_0 \times D_0^{\downarrow}\right),$$

there exists $T > 0$ such that $y(T,0) \in \vartheta_{-1} \times D_{-1}^{\uparrow}$.

Remark 11.2. The name for the subset $\vartheta_0 \times D_0^{\downarrow}$ can be misleading. Even though we have used the symbol $\downarrow$, after a solution hits $\vartheta_0 \times D_0^{\downarrow}$ and jumps, it is no longer going "downwards". On the contrary, by the design of the global effective potential, the pendulum will start going "upwards". A possible alternative for the edge case of $x_1 = \vartheta_0$ could be changing the notation to $\vartheta_0 \times \left(D_{0^+}^{\uparrow} \cup D_{0^-}^{\uparrow}\right)$.

The following two lemmas examine the behavior of solutions originating at the boundaries of the region surrounding the inverted position $\theta = \pi$ $(q = 0)$.

Lemma 11.7. For any $p \in \{-3, 3\}$ and any solution $y : \mathrm{dom}(y) \to C \cup D$ to $\mathcal{H}_{av}$ with

$$y(0,0) \in G_{av}\left(\vartheta_p \times D_p^{\uparrow}\right),$$

one of the following holds:

- There exists $T > 0$ such that $y(T, 0) \in \vartheta_{p^{\uparrow}} \times D_{-p}^{\downarrow}$.
- $\lim_{t\to\infty} y(t, 0) \in \mathcal{A}_0$.

where $\mathcal{A}_0 \coloneqq \{-\pi.\pi\} \times \{0\} \times \{0\} \times \{-4, 4\}$.

Lemma 11.8. For any $p \in Q \setminus \{0\}$ and any solution $y : \mathrm{dom}(y) \to C \cup D$ to $\mathcal{H}_{av}$ with

$$y(0,0) \in G_{av}\left(\vartheta_p \times D_p^{\downarrow}\right),$$

one of the following holds:

- $y(0,0) \in \left(\vartheta_p \times D_p^{\uparrow}\right)$, $y(0,0) \notin \left(\Theta_{p^{\uparrow}} \times \mathbb{R} \times \left\{p^{\uparrow}\right\} \times K_{p^{\uparrow}}^{\uparrow}\right)$, and solutions will necessarily jump afterwards.

- There exists $T > 0$ such that $y(T, 0) \in \vartheta_q \times D_q^{\uparrow}$.
- There exists $T > 0$ such that $y(T, 0) \in \vartheta_{p^{\downarrow}} \times D_{p^{\downarrow}}^{\downarrow}$, where $p^{\downarrow} = p - \text{sign}(p)$.

In words, Lemma 11.8 can be interpreted as: Whenever the hybrid dynamical system enforces a jump outside of the "final" region ($q = 0$) and the solution is going downwards, the following scenarios are successfully captured by the dynamics:

- Either the solution has zero speed and the jump map will adjust its logical values such that it will immediately start flowing upwards.
- The solution will loose enough kinetic energy during the next interval of flow and the velocity x_2 will change sign during flows. After that, the solution will start flowing upwards and eventually jump in the upwards direction.
- The solution does not loose enough kinetic energy during the next interval of flow and hits the next downward guard. This scenario can happen a successive finite amount of times until x_2 becomes 0 and any of the two previous scenarios occurs.

With the above lemmas, we can ensure that no solution to the hybrid dynamical system is Zeno.

Lemma 11.9. No maximal solution of $\mathcal{H}_{av}$ with $(x(0, 0), q(0, 0)) \in C \cup D$ is Zeno.

We are now prepared to present the main result of this section, which establishes suitable global asymptotic stability guarantees for the naturally unstable position $\theta = \pi$.

Theorem 11.1. Assume that $\alpha^2 < \frac{1}{2}\cos(3\pi/8)$. Then, the HDS $\mathcal{H}$ renders the set $\mathcal{A} \coloneqq \mathcal{A}_0 \times \mathbb{S}^1$ globally practically asymptotically stable.

Remark 11.3. The proofs for the different results above rely on analyzing the continuous-time behavior of solutions for $\mathcal{H}_{av}$ at different modes determined by the

state q, as well as studying how solutions are allowed to jump between those modes. These proofs make extensive use of the "next" mode $p^{\uparrow}$, and the "previous" mode $p^{\downarrow}$, from an "initial" mode p. The flow and jump sets in (11.10) can be explicitly written in terms of these quantities; this potential modification can improve the clarity of the explanation. Also, the definition of $C_p := \Theta_p \times \mathbb{R} \times \{p\} \times K_p^{\uparrow}$ and $C = \bigcup_{p \in Q} C_p$.

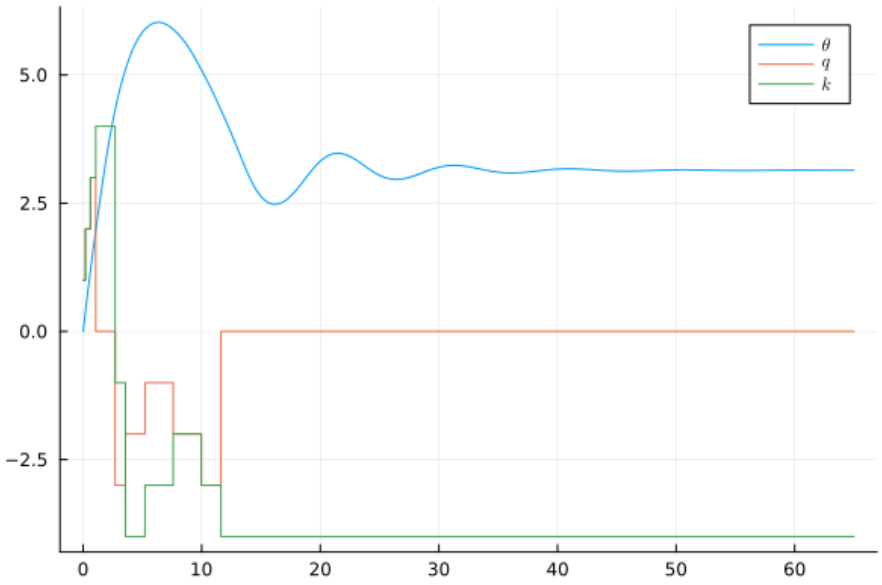


Figure 11.5. Convergence to the naturally unstable equilibrium under the $\mathcal{H}_{av}$.

To validate Theorem 11.1, Figure 11.5 presents numerical simulations of the average system for the Kapitza pendulum under the proposed hybrid vibrational control strategy. The figure illustrates a trajectory converging to the upright equilibrium $\theta = \pi$. We also present the evolution of the logical variables q and k, capturing the hybrid control's transitions between oscillation modes across the partitioned state space.

Remark 11.4. The successful emulation of the heavy-ball optimization dynamics over the global potential $U(x_1, q)$, obtained by "stitching" together several local potentials, relies on the carefully designed jump and flow sets presented in (11.9). This approach appears to be difficult to extend to a different setup than the one

studied in this manuscript. An approach that relies on a hybrid dynamical model of an automata that allows consecutive jumps without intervals of flow in-between, could prove to be useful for extensions to higher dimensional settings.

Acknowledgements

Chapter 11 contains unpublished material "On Hybrid Vibrational Control: Robust Global Stabilization of the Kapitza's Pendulum" coauthored with Professor J. I. Poveda [11]. The dissertation author is one of the primary investigators of this material.

Part V

Analysis and Proofs

Appendix A

Proofs of Chapter 3

A.1 Proofs of Section 3.2

Proof of Theorem 3.2

Gradient of Critic Error-Function in Deviation Variables

First, using (3.20) together with $H(x, u^*(x), \nabla V^*) = 0$ for all x, we obtain:

$$\psi(x, u^*(x))^\top \theta_c^* + Q(x) + R\left(u^*(x)\right) = \epsilon_{\mathrm{HJB}}(x). \tag{A.1}$$

Thus, using (3.19) and (A.1), we can rewrite the gradient of $e(\theta_c, x, u)$ as follows:

$$\nabla_{\theta_c} e(\theta_c, x, u) = \Theta(x, u)\left(\theta_c - \theta_c^*\right) + \upsilon_\epsilon(x, u) + \chi(x, u), \tag{A.2}$$

where

$$\Theta(x, u) \coloneqq \rho_i \Psi(x, u)\Psi(x, u)^\top + \rho_d \Lambda, \tag{A.3}$$

and

$$\upsilon_\epsilon(x, u) \coloneqq \rho_i \frac{\psi(x, u)\epsilon_{\mathrm{HJB}}(x)}{\left(1 + |\psi(x, u)|^2\right)^2} + \rho_d \sum_{k=1}^{N} \frac{\psi(x_k, u^*(x_k))\epsilon_{\mathrm{HJB}}(x_k)}{\left(1 + |\psi(x_k, u^*(x_k))|^2\right)^2} \in \mathbb{R}^{l_c}, \tag{A.4}$$

$$\chi(x,u) \coloneqq \frac{\rho_i \psi(x,u) \left[\frac{\partial \phi_c(x)}{\partial x} g(x) \left(u - u^*(x)\right)\right]^\top \theta_c^*}{\left(1 + |\psi(x,u)|^2\right)^2} + \frac{\rho_i \psi(x,u) \left[R(u) - R(u^*(x))\right]}{\left(1 + |\psi(x,u)|^2\right)^2} \in \mathbb{R}^{l_c}, \tag{A.5}$$

which, by using the fact that $\frac{r}{(1+r^2)^2} \leq \frac{3\sqrt{3}}{16}, \forall r \in \mathbb{R}_{\geq 0}$, satisfy:

$$|\upsilon_\epsilon(x,u)| \leq \frac{3\sqrt{3}}{16} \overline{\epsilon_{\text{HJB}}} \left(\rho_i + N\rho_d\right), \tag{A.6a}$$

$$|\chi(x,u)| \leq \rho_i \frac{3\sqrt{3}}{16} \left(\overline{g} \left(\overline{d\phi_c} \left[1 + |\theta_c^*|\right] + \overline{d\epsilon_c} \right) |u - u^*(x)| + \lambda_{\max}\left(\Pi_u\right) |u - u^*(x)|^2 \right). \tag{A.6b}$$

The following Lemma will be instrumental for our results.

Lemma A.1. If the data is $\underline{\lambda}$-sufficiently-rich, then there exist $\overline{\Theta}, \underline{\Theta} \in \mathbb{R}_{>0}$ such that

$$\underline{\Theta} I_n \preceq \Theta(x,u) \preceq \overline{\Theta} I_n \qquad \forall x \in \mathbb{R}^n,\ \forall u \in \mathbb{R}^m.$$

Proof. Let $\theta \in \mathbb{R}^{l_c}$ be arbitrary. Since, by assumption, the data is $\underline{\lambda}$-SR it follows that:

$$\begin{aligned} \theta^\top \Theta(x,u) \theta &= \theta^\top \rho_i \Psi(x,u) \Psi(x,u)^\top \theta + \theta^\top \rho_d \Lambda \theta \\ &\geq \rho_d \underline{\lambda} |\theta|^2 \\ &\implies \Theta(x,u) \succeq \underline{\Theta} I_{l_c},\ \forall (x,u) \in \mathbb{R}^n \times \mathbb{R}^m, \end{aligned} \tag{A.7}$$

where $\underline{\Theta} \coloneqq \rho_d \underline{\lambda}$. On the other hand, using the fact that $\left|aa^\top\right| = |a|^2,\ \forall a \in \mathbb{R}^n$, we obtain that:

$$\left|\Psi(x,u)\Psi(x,u)^\top\right| = |\Psi(x,u)|^2 \leq 1,\ \forall (x,u) \in \mathbb{R}^n \times \mathbb{R}^m,$$

we obtain:

$$\begin{aligned}\theta^\top \Theta(x,u)\theta &= \theta^\top \rho_i \psi(x,u)\psi(x,u)^\top \theta + \theta^\top \rho_d \Lambda \theta \\ &\leq \left(\rho_i + \rho_d \lambda_{\max}(\Lambda)\right) |\theta|^2 \\ &\implies \Theta(x,u) \preceq \overline{\Theta} I_{l_c}, \quad \forall (x,u) \in \mathbb{R}^n \times \mathbb{R}^m,\end{aligned}$$

where $\overline{\Theta} \coloneqq \rho_i + \rho_d \lambda_{\max}(\Lambda)$. ■

Lyapunov-Based Analysis

Recall from Section 3.2 that $y = (\theta_c, p, \tau)$, suppose that the assumptions of Theorem 3.2 hold and consider the Lyapunov candidate function $V_c : \mathbb{R}^{l_c} \times \mathbb{R}^{l_c} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ given by:

$$V_c(y) \coloneqq \frac{|p-\theta_c|^2}{4} + \frac{|p-\theta_c^*|^2}{4} + k_c \rho_d \tau^2 \frac{(\theta_c - \theta_c^*)\top \Lambda (\theta_c - \theta_c^*)}{2}, \tag{A.8}$$

where Λ was defined in Assumption 3.1 and which satisfies:

$$\begin{aligned}\underline{c}|y|_{\mathcal{A}_c}^2 \leq V_c(y) \leq \overline{c}|y|_{\mathcal{A}_c}^2, \\ \underline{c} \coloneqq \min\left\{\frac{1}{4}, \frac{k_c \rho_d T_0^2 \underline{\lambda}}{2}\right\}, \quad \overline{c} \coloneqq \left\{\frac{3}{4}, \ \frac{1}{2}\left(1 + k_c \rho_d T^2 \overline{\lambda}\right)\right\},\end{aligned} \tag{A.9}$$

where $\overline{\lambda} \coloneqq \lambda_{\max}(\Lambda)$. Now, let $u \in \mathcal{U}_V$, and consider the time derivative of V_c along the continuous-time evolution of the critic subsystem, i.e., $\dot{V}_c = \nabla_y V_c(y)^\top \dot{y}$. Then, by using (A.2) and Lemma A.1, and some algebraic manipulation, $\dot{V}_c$ can be shown to satisfy

$$\dot{V}_c \leq -\frac{k_c \tau}{2} \begin{pmatrix} |p-\theta_c| & |\theta_c - \theta_c^*| \end{pmatrix} M(\tau) \begin{pmatrix} |p-\theta_c| \\ |\theta_c - \theta_c^*| \end{pmatrix} + 2\sqrt{2} k_c y_{\mathcal{A}_c}\big(|\upsilon_\epsilon(x)| + |\chi(x,u(x))|\big), \tag{A.10}$$

where

$$M(\tau) \coloneqq \begin{pmatrix} \frac{2}{k_c \tau^2} & -\frac{\rho_i}{2} \\ -\frac{\rho_i}{2} & \underline{\Theta} \end{pmatrix}, \tag{A.11}$$

and $\mathcal{A}_c$ was defined in Section 3.2. Since $2\rho_d \underline{\lambda} > \rho_i$ and $T^2 < \frac{8\rho_d \underline{\lambda}}{k_c \rho_i^2}$ by means of Asssumption 3.2, and $\tau(t,j) \in [T_0, T],\ \forall (t,j) \in \mathrm{dom}\,(y)$ by construction of the critic update dynamics (3.21), it follows that $M(\tau) \succeq \tilde{r}$ with $\tilde{r} \coloneqq \underline{\Theta} - \frac{\rho_i}{2}$. Hence, defining $\underline{r} \coloneqq \frac{k_c T_0}{2}\tilde{r}$, from (A.10) and using (A.6), we obtain that:

$$\dot{V}_c \leq -\underline{r}|y|^2_{\mathcal{A}_c} + |y|_{\mathcal{A}_c}\Big(\gamma_\nu\left(\overline{\epsilon_{\mathrm{HJB}}}\right) + \gamma_\chi\left(|u(x) - u^*(x)|\right)\Big), \tag{A.12}$$

where $\gamma_\nu, \gamma_\chi \in \mathcal{K}_\infty$ are given by:

$$\begin{aligned} \gamma_\nu(r) &\coloneqq \frac{3\sqrt{6}}{8}\left(\rho_i + N\rho_d\right) r, \quad \gamma_\chi(r) \coloneqq c_\chi(r + r^2), \\ c_\chi &\coloneqq \frac{3\sqrt{6}}{8}\rho_i \max\left\{\overline{g}\left(\overline{d\phi_c}\left[1 + |\theta_c^*|\right] + \overline{d\epsilon_c}\right),\ \lambda_{\max}\left(\Pi_u\right)\right\}. \end{aligned}$$

Thus, letting $D_0^c \in (0,1)$, and using (A.9), (A.12):

$$\dot{V}_c \leq -\frac{\underline{r}(1 - D_0^c)}{\overline{c}} V_c(y), \quad \forall |y|_{\mathcal{A}_c} \geq \frac{1}{D_0^c}\Big(\gamma_\nu\left(\overline{\epsilon_{\mathrm{HJB}}}\right) + \gamma_\chi\left(|u(x) - u^*(x)|\right)\Big). \tag{A.13a}$$

On the other hand, the change of V_c during the jumps in the update dynamics for the critic (3.21), satisfies:

$$V_c\left(y^+\right) - V_c(y) \leq -\eta V_c(y), \tag{A.14}$$

with $\eta \coloneqq 1 - \frac{T_0^2}{T^2} - \frac{1}{2k_c\rho_d\underline{\lambda}T^2}$ which satisfies $\eta \in (0,1)$ by means of Assumption 3.2. Together, (A.13) and (A.14), in conjuction with the quadratic bounds of (A.9), imply the results of Theorem 3.2 via [273, Prop 2.7] and the fact that $|\theta_c(t,j) - \theta_c^*| \leq |y(t,j)|_{\mathcal{A}_c} \leq |(\theta_c(t,j), p(t,j))|_{\mathcal{A}_{\theta_c,p}}$ for all $(t,j) \in \mathrm{dom}\,(y)$. ■

Proof of Theorem 3.3

Gradient of Actor Error-Function in Deviation Variables

First, note that we we can write (3.29) as:

$$\nabla_{\theta_u}\varepsilon_a(x,\theta_c,\theta_u) = \Omega(x)\left(\theta_u - \theta_c^* - (\theta_c - \theta_c^*)\right),$$

and consider the following Lemma, instrumental for our results.

Lemma A.2. There exists $\overline{\Omega}, \underline{\Omega} \in \mathbb{R}_{>0}$ such that

$$\underline{\Omega} I_{l_c} \preceq \Omega(x) \preceq \overline{\Omega} I_{l_c}.$$

Proof. Let $\theta \in \mathbb{R}^{l_c}$ be arbitrary. Then, by the definition of $\Omega : \mathbb{R}^n \to \mathbb{R}^{l_c \times l_c}$ in (3.30), it follows that:

$$\theta^\top \Omega(x)\theta = \alpha_1 \frac{\left|\omega(x)^\top \theta\right|^2}{1 + \mathrm{Tr}\{\omega(x)^\top \omega(x)\}} + \alpha_2 |\theta|^2 \geq \alpha_2 |\theta|^2 \implies \Omega(x) \succeq \underline{\Omega} I_{l_c}, \quad \forall x \in \mathbb{R}^n,$$

where $\underline{\Omega} \coloneqq \alpha_2$. On the other hand, we obtain:

$$\theta^\top \Omega(x)\theta = \left(\alpha_1 \frac{|\omega(x)|^2}{1 + |\omega(x)|_F^2} + \alpha_2\right) |\theta|^2 \leq \overline{\Omega} |\theta|^2 \implies \Omega(x) \preceq \overline{\Omega} I_{l_c}, \quad \forall x \in \mathbb{R}^n,$$

where $\overline{\Omega} \coloneqq \alpha_1 + \alpha_2$, $|A|_F$ represents the Frobenius norm and where we used $|A| \leq |A|_F$, $\forall A \in \mathbb{R}^{l_c \times l_c}$ and $\frac{r^2}{1+r^2} \leq 1$ $\forall r \in \mathbb{R}$. ■

Now, consider the Lyapunov function

$$\mathcal{V}(z) \coloneqq V_o(x) + V_c(y) + V_a(\theta_u), \tag{A.15a}$$

$$V_o(x) \coloneqq V^*(x), \quad V_a(\theta_u) \coloneqq \frac{1}{2}|\theta_u - \theta_c^*|^2, \tag{A.15b}$$

where V_c was defined in (A.8) and where we recall that $z = (x, y, \theta_u)$. By [22, Lemma 4.3], and since $V_o = V^*$ is a continuous and positive definite function in $\mathbb{R}^n$, there exist $\underline{\gamma}_o, \overline{\gamma}_o \in \mathcal{K}$ such that $\underline{\gamma}_o(|x|) \leq V_o(x) \leq \overline{\gamma}_o(|x|)$. Hence, using (A.9), and the fact that sum of class $\mathcal{K}$ is in turn of class $\mathcal{K}$, there exist $\underline{\gamma}_{\mathcal{V}}, \overline{\gamma}_{\mathcal{V}} \in \mathcal{K}$ such that:

$$\underline{\gamma}_{\mathcal{V}}(|z|_{\mathcal{A}}) \leq \mathcal{V}(z) \leq \overline{\gamma}_{\mathcal{V}}(|z|_{\mathcal{A}}) \tag{A.16}$$

Now, the time derivative of $\dot{V}_o = \nabla V_o(x)^\top \dot{x}$ along the trajectories of (3.32) satisfies:

$$\dot{V}_o \leq -Q(x) + \frac{\overline{g}^2 \lambda_{\max}(\Pi_u^{-1})}{2} \left(\overline{d\phi_c}|\theta_c^*| + \overline{d\epsilon_c}\right)\left(\overline{d\phi_c}|\theta_u - \theta_c^*| + \overline{d\epsilon_c}\right). \tag{A.17}$$

On the other hand, making use of Lemma 6.2, for the time derivative of $\dot{V}_a = \nabla_{\theta_u} V_a(\theta_u)^\top \theta_u$ we obtain:

$$\dot{V}_a \leq -k_u \alpha_2 |\theta_u - \theta_c^*|^2 + k_u \overline{\Omega} |\theta_u - \theta_c^*||\theta_c - \theta_c^*|. \tag{A.18}$$

Hence, using (A.10), (A.17), and (A.18), together with the upper bounds in (A.6), we obtain that the time derivative of $\mathcal{V}$ along the trajectories of the closed-loop system satisfies:

$$\begin{aligned}\dot{\mathcal{V}} \leq & -Q(x) - \underline{r}|y|_{\mathcal{A}_c}^2 - k_u \alpha_2 |\theta_u - \theta_c^*|^2 \\ & + c_y |y|_{\mathcal{A}_c} + c_u |\theta_u - \theta_c^*| + c_{yu} |\theta_u - \theta_c^*||y|_{\mathcal{A}_c} \\ & + c_{yu^2} |y|_{\mathcal{A}_c} |\theta_u - \theta_c^*|^2 + c_0,\end{aligned} \tag{A.19}$$

where

$$c_y := \frac{3\sqrt{6}}{8} k_c \left(\overline{\epsilon_{\text{HJB}}} (\rho_i + N\rho_d) \right.$$

$$+\frac{1}{2}\overline{g}^2\rho_i\left[\lambda_{\max}\left(\Pi_u^{-1}\right)\left(\overline{d\phi_c}\left[1+|\theta_c^*|\right]+\overline{d\epsilon_c}\right)\overline{d\epsilon_c}+\lambda_{\max}\left(\Pi_u\right)\lambda_{\max}\left(\Pi_u^{-1}\right)^2\overline{d\epsilon_c}^2\right]\Bigg),$$

$$c_u \coloneqq \frac{1}{2}\left(\overline{d\phi}_c|\theta_c^*|+\overline{d\epsilon_c}\right)\overline{g}^2\lambda_{\max}\left(\Pi_u^{-1}\right)\overline{d\phi_c},$$

$$c_{yu} \coloneqq \frac{3\sqrt{6}k_c}{16}\left(2k_u\overline{\Omega}+\overline{g}^2\rho_i\lambda_{\max}\left(\Pi_u^{-1}\right)\left(\overline{d\phi_c}\left[1+|\theta_c^*|\right]+\overline{d\epsilon_c}\right)\overline{d\phi_c}\right),$$

$$c_{yu^2} \coloneqq \frac{3\sqrt{6}}{16}k_c\overline{g}^2\rho_i\lambda_{\max}\left(\Pi_u\right)\lambda_{\max}\left(\Pi_u^{-1}\right)^2\overline{d\phi_c}^2,$$

$$c_0 \coloneqq \frac{1}{2}\left(\overline{d\phi}_c|\theta_c^*|+\overline{d\epsilon_c}\right)\overline{g}^2\lambda_{\max}\left(\Pi_u^{-1}\right)\overline{d\epsilon_c}.$$

Then, for all $|\theta_u-\theta_c^*|\leq\frac{c_{yu}}{c_{yu^2}}$, by using $Q(x)=x^\top\Pi_x x$ and letting $d_1\in(0,1)$, from (A.19), $\dot{\mathcal{V}}$ can be further upper bounded as:

$$\begin{aligned}\dot{\mathcal{V}} \leq & -\lambda_{\min}\left(\Pi_x\right)|x|^2-(1-d_1)\left(\underline{r}|y|_{\mathcal{A}_c}^2+k_u\alpha_2|\theta_u-\theta_c^*|^2\right)\\ & +c_y|y|_{\mathcal{A}_c}+c_u|\theta_u-\theta_c^*|+c_0\\ & -\begin{pmatrix}|y|_{\mathcal{A}_c} & |\theta_u-\theta_c^*|\end{pmatrix}\begin{pmatrix}d_1\underline{r} & -c_{yu}\\ -c_{yu} & d_1k_u\alpha_2\end{pmatrix}\begin{pmatrix}|y|_{\mathcal{A}_c}\\ |\theta_u-\theta_c^*|\end{pmatrix}.\end{aligned}\tag{A.20}$$

Now, pick a set of tunable parameters (ρ_i,ρ_d,k_c,k_u) such that $\underline{r}\geq\frac{c_{yu}^2}{d_1^2k_u\alpha_2}$ so that from (A.20), we obtain:

$$\dot{\mathcal{V}}\leq-(1-d_2)d_z|z|_{\mathcal{A}}^2,\quad\forall|z|_{\mathcal{A}}\geq\max\left\{\frac{c_0}{2d_{yu}},\ \frac{2d_{yu}}{d_2d_z}\right\},\ |\theta_u-\theta_c^*|\leq\frac{c_{yu}}{c_{yu^2}},\tag{A.21a}$$

with

$$\begin{aligned}d_z &\coloneqq \min\left\{\lambda_{\min}\left(\Pi_x\right),\ (1-d_1)\underline{r},\ k_u\alpha_2\right\},\\ d_{yu} &\coloneqq \max\left\{2c_{yu},\ c_0\right\},\quad d_2\in(0,1).\end{aligned}$$

Notice that for every compact set K_θ of initial conditions for θ_u we can pick suitable

$\rho_i, \rho_d, \alpha_1, \alpha_2, k_c, k_u$ to satisfy $K_\theta \subset \frac{c_{yu}}{cyu^2}\mathbb{B}$ such that (A.21) holds for every trajectory with $\theta_u(0,0) \in K_\theta$. Now, during jumps x and θ_u do not change, and hence $\mathcal{V}$ satisfies:

$$\mathcal{V}(z^+) - \mathcal{V}(z) = V_c(y^+) - V_c(y) \leq -\eta V_c(y). \tag{A.22}$$

The result of the theorem follows by using the strong-decrease of $\mathcal{V}$ during flows outside a neighborhood of $\mathcal{A}$ described in (A.21), the non-increase of $\mathcal{V}$ during jumps given in (A.22), by noting that, by design, the closed-loop dynamics are a well-posed HDS which experiences periodic jumps followed by intervals of flow of length $T - T_0 > 0$ (c.f. [6]), and by following the same arguments of [33, Prop 3.27] and [273, Prop. 2.7]. ■

Appendix B

Proofs of Chapter 4

B.1 Proofs of Section 4.3

For the sake of clarity in the presentation of the proofs, in this section, we assume that $n = 1$, i.e., that the action of each player q_i is one-dimensional. The proofs for the general case where $n \geq 1$ follow analogously, with the appropriate incorporation of Kronecker products to handle the higher-dimensional action spaces.

Proof of Lemma 4.1

1) Well-posedness follows directly by [16, Thm. 6.30], since F_1 is continuous, C_1 and D_1 are closed sets, and G_1 is outer-semicontinuous (OSC) and locally bounded (LB) in D_1 by construction.

2) To rule out finite escape times it suffices to study the behavior of the states (q, p). Indeed, by using the ℓ-global Lipschitz assumption on $\mathcal{G}$, leveraging the structure of the flow map F_1 in (4.9c), and using the fact that $\mathcal{G}(q^*) = 0$, it follows that

$$|\dot{q}| \leq \frac{2}{T_0}|p - q|, \text{ and } |\dot{p}| \leq 2T\ell|q - q^*|,$$

for all $q^* \in \mathcal{A}_{NE}$, which implies that $|(\dot{q}, \dot{p})| \leq \tilde{\ell}|(q, p) - (q^*, q^*)|$, with $\tilde{\ell} :=$

$2\sqrt{3}\max\{\frac{1}{T_0}, T\ell\}$. Then, by using the fact that

$$\frac{d|u(t)|}{dt} \leq \left|\frac{du(t)}{dt}\right|,$$

for any differentiable function $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^N$ and any $t \in \mathrm{dom}\,(u)$, it follows that

$$\frac{d|(q-q^*, p-q^*)|}{dt} \leq \tilde{\ell}|(q-q^*, p-q^*)| \quad \forall q^* \in \mathcal{A}_{NE}.$$

Then, the Gronwall-Bellman inequality implies that during flows

$$|(q-q^*, p-q^*)| \leq |(q(0)-q^*, p(0)-q^*)|e^{\tilde{\ell}t}$$

for all $q^* \in \mathcal{A}_{NE}$ and all $(t, \cdot) \in \mathrm{dom}((q,p))$. Thus, the flow map (4.9c) does not generate finite escape times.

Additionally, since after jumps we have that $\tau^+ \in \{T_0, T\}^N$, it follows that $G_1(D) \subset C_1 \cup D_1$, which in turn implies that solutions do not stop due to jumps leaving the set $C_1 \cup D_1$. Next, note that the synchronization mechanism that governs the state τ is decoupled from the dynamics of the states (q,p), and can be written as a the following HDS:

$$\tau \in C_\tau \coloneqq [T_0, T]^N, \qquad \dot{\tau} = \eta \mathbf{1}_N, \tag{B.1a}$$

$$\tau \in D_\tau \coloneqq \left\{\tau \in C_\tau : \max_i \tau_i = T\right\}, \quad \tau^+ \in G_\tau(\tau), \tag{B.1b}$$

where $G_\tau(\tau)$ is the projection of G_1 into the τ-component, which is independent of (p,q). This hybrid system is well-posed by construction, and by [51, Thm. 1] it renders uniformly fixed-time stable the set $\mathcal{A}_{\text{sync}}$, with a convergence bound T^*

given by

$$T^* \coloneqq \frac{1}{\eta}(T - T_0) + N, \tag{B.2}$$

for all $\tau(0,0) \in [T_0, T]^N$. Moreover, each solution has at most N jumps in any interval of length $L \coloneqq \frac{1}{\eta}(T - T_0)$, and, for any pair of hybrid times $(t,j), (s,i) \in \mathrm{dom}\,(\tau)$ with $t+j \geq s+i \geq T^*$ the following dwell-time condition holds $L + t - s \geq \left\lfloor \frac{j-i}{N} \right\rfloor L$, where $\lfloor \cdot \rfloor$ denotes the floor function. Thus, any solution τ of system (B.1a) is complete and also satisfies $|\tau(t,j)|_{\mathcal{A}_{\mathrm{sync}}} = 0$ for all $t + j \geq T^*$ such that $(t,j) \in \mathrm{dom}(\tau)$. Since the states (q,p) of $\mathcal{H}_1$ evolve in $\mathbb{R}^N \times \mathbb{R}^N$, for each $\tau(0,0) \in [T_0, T]^N$ the hybrid time domains of system (4.9a)-(4.9f) are the same hybrid time domains of system (B.1a). This equivalence, plus the above properties, establish the result. ■

The previous Lemma directly implies the following:

Lemma B.1. Let $\nu > 0$ and consider the HDS $\mathcal{H}_1$ with restricted flow and jump sets given by:

$$C_\nu \coloneqq \Big\{ x \in \mathbb{R}^{2nN+N} : (p,q) \in \{(q^*, q^*)\} + \nu\mathbb{B},\ \tau \in [T_0, T]^N \Big\},$$
$$D_\nu \coloneqq \Big\{ x \in \mathbb{R}^{2nN+N} :\ x \in C_\nu,\ \ \max_{i \in \mathcal{V}} \tau_i = T \Big\},$$

and jump map G_1 with values intersected with the set $C_\nu \cup D_\nu$. Then, the restricted system $\mathcal{H}_\nu = \{F_1, C_\nu, G_\nu, D_\nu\}$ renders UGFxS the set $\mathcal{A}_\nu \coloneqq (\{(q^*, q^*)\} + \nu\mathbb{B}) \times \mathcal{A}_{\mathrm{sync}}$.

Proof of Theorem 4.1

First, by using Lemmas 4.1 and B.1, we note that we can analyze the HDS $\mathcal{H}_1$ by instead studying the HDS $\mathcal{H}_v$ with data intersected with the set $\mathcal{A}_\nu$. We denote this new restricted HDS as $\mathcal{H}_s \coloneqq \{F_s, C_s, G_s, D_s\}$, and we note that any compact set $\mathcal{A}' \subset \mathbb{R}^N \times \mathbb{R}^N$ such that $\mathcal{A}' \times \mathcal{A}_{\mathrm{sync}}$ is UGAS for this system will also be UGAS for $\mathcal{H}_\nu$ thanks to the hybrid reduction principle [16, Cor. 7.24]. Moreover, since ν is arbitrary and independent

of any parameter, and $\mathcal{H}_1$ has no finite escape times, the set $\mathcal{A}' \times \mathcal{A}_{\text{sync}}$ will also be UGAS for $\mathcal{H}_1$. Our subsequent analysis, introduces multiple lemmas that characterize the stability properties of $\mathcal{A}$ under $\mathcal{H}_s$.

Lemma B.2. Under the conditions of Theorem 4.1-(i_1), system $\mathcal{H}_s$ renders UGAS the set $\mathcal{A}$ given by (4.14). □

Proof. By using the potential function P we define the error $\tilde{P}(q) \coloneqq P(q) - P(\mathcal{A}_{\text{NE}})$, and we consider the Lyapunov function

$$V = V_1 + V_2 + V_3, \tag{B.3}$$

where the smooth functions V_i are defined as follows:

$$V_1(x) \coloneqq \frac{1}{4}|p-q|^2,\ V_2(x) \coloneqq \frac{1}{4}|p|^2_{\mathcal{A}_{NE}},\ V_3(x) \coloneqq \frac{|\tau|^2}{N}\tilde{P}(q). \tag{B.4a}$$

where $|z|^2_{,\mathcal{A}_{NE}} = \min_{s\in\mathcal{A}_{NE}} |z-s|^2$ and $|z|^2 = z^\top z$. By the definition of potential-games, and the construction of V_1 and V_2, the function V is radially unbounded and positive definite with respect to the compact set $\mathcal{A} \cap (C_s \cup D_s)$.

Now, during flows in C_s the time derivative of V satisfies:

$$\dot{V}(x) = \frac{\partial V(x)}{\partial x}\dot{x} = \frac{\partial V(x)}{\partial q}\dot{q} + \frac{\partial V(x)}{\partial p}\dot{p} + \frac{\partial V(x)}{\partial \tau}\dot{\tau}. \tag{B.5}$$

The first term of (B.5) is given by:

$$\begin{aligned}\frac{\partial V(x)}{\partial q}\dot{q} &= \left(-\frac{1}{2}(p-q) + \frac{|\tau|^2}{N}\nabla P(q)\right)^\top \dot{q} \\ &= -(p-q)^\top \mathcal{T}^{-1}\cdot(p-q) \\ &\quad + 2\frac{|\tau|^2}{N}\nabla P(q)^\top \mathcal{T}^{-1}\cdot(p-q),\end{aligned}$$

$$= -\frac{1}{\tau_s}(p-q)^\top(p-q) + 2\tau_s(p-q)^\top \nabla P(q),$$

where in the last equality we used the fact that in the set $C_s \backslash D_s$ we have $\tau = \tau_s \mathbf{1}_N$ with $\tau_s \in [T_0, T]$, and that points in $C_s \cap D_s$ lead to solutions that cannot flow. Similarly,

$$\begin{aligned}
\frac{\partial V(x)}{\partial p}\dot{p} &= \left(\frac{1}{2}(p-q) + \frac{1}{2}\big(p - \Pi_{\mathcal{A}_{NE}}(p)\big)\right)^\top \dot{p} \\
&= -(p-q)^\top \mathcal{T}\mathcal{G}(q) \\
&\quad - (p - \Pi_{\mathcal{A}_{NE}}(p))^\top \mathcal{T}\mathcal{G}(q), \\
&= -\tau_s \Big((p-q) + (p - \Pi_{\mathcal{A}_{NE}}(p))\Big)^\top \mathcal{G}(q),
\end{aligned}$$

where $\Pi_{\mathcal{A}_{NE}}(p)$ is the projection of p on $\mathcal{A}_{NE}$. Moreover, it follows that

$$\frac{\partial V(x)}{\partial \tau}\dot{\tau} = \left(\tilde{P}(q)\frac{\tau}{N}\right)^\top \dot{\tau} = \tilde{P}(q)\frac{\tau^\top \mathbf{1}_N}{N} = 2\tau_s \tilde{P}(q)\eta.$$

Combining the above inequalities, and using the definition of $\mathcal{G}(q) = \nabla P(q)$ and $0 < \eta \leq 1/2$, we obtain:

$$\dot{V}(x) \leq -\frac{1}{\tau_s}|p-q|^2 - \tau_s \left(\big(q - \Pi_{\mathcal{A}_{NE}}(p)\big)^\top \mathcal{G}(q) - \tilde{P}(q)\right). \tag{B.6}$$

Now, since P is a convex function with ℓ-Lipschitz gradient $\mathcal{G}$ the following standard inequality is satisfied [274, Thm. 5.8]:

$$P(y) \geq P(x) + (y-x)^\top \mathcal{G}(x) + \frac{1}{2\ell}|\mathcal{G}(x) - \mathcal{G}(y)|^2, \quad \forall x, y \in \mathbb{R}^N. \tag{B.7}$$

Then, letting $x = q$ and $y = \Pi_{\mathcal{A}_{NE}}(p)$ for $q, p \in \mathbb{R}^n$, and using the fact that

$\mathcal{G}\left(\Pi_{\mathcal{A}_{NE}}(p)\right) = 0$, it follows from (B.7) that

$$\begin{aligned}\frac{1}{2\ell}|\mathcal{G}(q)|^2 &\leq -\Big(\Pi_{\mathcal{A}_{NE}}(p) - q\Big)^\top \mathcal{G}(q) + P\left(\Pi_{\mathcal{A}_{NE}}(p)\right) - P(q) \\ &= \Big(q - \Pi_{\mathcal{A}_{NE}}(p)\Big)^\top \mathcal{G}(q) - \tilde{P}(q). \end{aligned} \tag{B.8}$$

Thus, from (B.6), we obtain during flows that

$$\dot{V}(x) \leq -\frac{1}{\tau_s}|p-q|^2 - \frac{\tau_s}{2\ell}|\mathcal{G}(q)|^2. \tag{B.9}$$

Since $|\mathcal{G}(q)| = 0$ if and only if $q \in \mathcal{A}_{\text{NE}}$, during flows we have $\dot{V}(x) < 0$ for all $x \in C_s \backslash \mathcal{A}$.

Similarly, during each jump the change of V, denoted $\Delta_j^{j+1}V(x) \coloneqq V(x(t,j+1)) - V(x(t,j))$, satisfies $\Delta_j^{j+1}V(x) = \Delta_j^{j+1}V_3(x)$ and

$$\begin{aligned}\Delta_j^{j+1}V_3(x) &= \frac{\tilde{P}(q)}{N}\Big(|\tau^+|^2 - |\tau|^2\Big) = \frac{\tilde{P}(q)}{N}\sum_{i=1}^{N}(\tau_i^{2+} - \tau_i^2), \\ &= -\frac{\tilde{\varepsilon}}{N}\tilde{P}(q) \leq 0, \end{aligned} \tag{B.10}$$

for some $\tilde{\varepsilon} > 0$, where the last equality follows by the definition of G_1 in (4.9a) and the two following facts: first, if $x \in D_s$, we have two possible cases for all players $i \in \mathcal{V}$: a) if $\tau_i = T_0$, then $\tau_i^+ = T_0$; b) if $\tau_i = T$ then $\tau_i^+ \in \{T_0, T\}$; second, if $x \in D_s$, we have that in each jump one and only one player i satisfies $\tau_i = T$ and $\tau_i^+ = T_0$. Therefore, since $T > T_0$ there exists $\tilde{\varepsilon} > 0$ such that $T_0^2 - T^2 = -\tilde{\varepsilon}$, leading to (B.10). This implies that V does not increase during each reset triggered by a player. Given that the hybrid time domains of $\mathcal{H}_s$ are intervals of flow of duration $\frac{1}{\eta}(T - T_0)$,

followed by n consecutive jumps, we can use (B.10) n times to obtain:

$$\Delta_j^{j+N} V(z) = \sum_{k=1}^{N} \Delta_{j+k-1}^{j+k} V(z) = -\varepsilon \tilde{P}(q) \leq 0, \quad \forall\, x \in D_s.$$

By [16, Prop. 3.27], the periodic strong decrease of V during flows, and its non-increase during jumps, imply that $\mathcal{H}_s$ renders UGAS the set $\mathcal{A}$. ■

Lemma B.3. Under the conditions of Theorem 4.1-(i_2), system $\mathcal{H}_s$ renders UGES the set $\mathcal{A}$.

Proof. The κ-strong convexity of $\tilde{P}$ and the ℓ-Lipschitz continuity of $\mathcal{G}$, yields the following quadratic bounds on V:

$$\min\left\{\frac{1}{4}, \frac{\kappa T_0^2}{2}\right\} |x|_{\mathcal{A}}^2 \leq V(x) \leq \frac{1}{2} \max\left\{1 + \frac{\ell^2}{\kappa},\ \frac{3}{2}\right\} |x|_{\mathcal{A}}^2, \tag{B.11}$$

where we have used the fact that in such cases $\mathcal{A}_{\mathrm{NE}} = \{q^*\}$. Additionally, it follows that $\Pi_{\mathcal{A}_{NE}}(p) = q^*$. Then, from (B.6), we obtain

$$\dot{V}(x) \leq -\frac{1}{\tau_s}|p - q|^2 - \tau_s \frac{\kappa}{2}|q - q^*|^2. \tag{B.12}$$

Using the quadratic bounds (B.11) in (B.12) yields

$$\dot{V}(x) \leq -\lambda V(x), \tag{B.13}$$

where

$$\lambda \coloneqq \frac{2}{3T} \frac{\min\{1, 0.25 T_0 T \kappa\}}{\max\{1, 2T^2 \ell\}} \approx \frac{1}{12T} \frac{T_0}{T} \frac{\kappa}{\ell}, \tag{B.14}$$

and the approximation holds when T_0 is sufficiently small, and T is sufficiently large

(but finite). Then, during each interval of flow, V satisfies the t-time bound

$$V(t,j) \leq V(t_j,j)e^{-\lambda(t-t_j)}, \tag{B.15}$$

for all $(t,j) \in \operatorname{dom}(x)$ such that $j = k \cdot N$ for some $k \in \mathbb{N}$. To characterize the behavior of V during jumps, let Θ and I be the set of indices of players who implement $\alpha_i = 0$, and $\alpha_i = 1$, respectively. It follows that after the n consecutive jumps that proceed the intervals of flow of every solution of $\mathcal{H}_s$, we have

$$\begin{aligned}
\Delta_j^{j+N} V(x) \leq & -\frac{1}{4}\left(|p-q|^2 + |p-q^*|^2\right) \\
& + \frac{1}{4}\sum_{i\in I}\left[(p_i - q_i)^2 + (p_i - q_i^*)^2\right] \\
& + \frac{1}{4}\sum_{i\in\Theta}[q_i - q_i^*]^2 - \frac{\kappa}{2}(\tau_s^2 - T_0^2), \\
\leq & -\frac{1}{4}\sum_{i\in\Theta}\left((p_i - q_i)^2 + (p_i - q_i^*)^2\right) \\
& - \frac{1}{2}\left(\kappa(\tau_s^2 - T_0^2) - \frac{1}{2}\right)\sum_{i\in\Theta}(q_i - q_i^*)^2 \\
& - \frac{\kappa}{2}(\tau_s^2 - T_0^2)\sum_{i\in I}(q_i - q_i^*)^2 \implies \Delta_j^{j+N} V(x) \leq 0,
\end{aligned} \tag{B.16}$$

where we used the fact $\tilde{P}$ is strongly convex, and the condition $T^2 - T_0^2 > \frac{1}{2\kappa}$ given by (4.16) with $\rho_J = \kappa^{-1}$. Therefore, it follows that $V(t_j,j) \leq V(t_j, j-N)e^{-\lambda(t-t_j)}$ for all $j \geq n$ and $t \geq t_j$. Since each interval of flow has length $L = (T - T_0)/\eta$, it follows that $\tilde{V}(t_j,j) \leq \tilde{V}(t_{j-N} + L, j-N)e^{-\lambda L}e^{-\lambda(t-t_j)}$. Iterating, and using (B.15):

$$V(t,j) \leq V(0,0)e^{-\lambda\left(\left\lfloor \frac{j}{N}\right\rfloor - 1\right)L}e^{-\lambda(t-t_j)}. \tag{B.17}$$

Thus, from (B.17), and using the quadratic bounds on V, we obtain:

$$|x(t,j)|_{\mathcal{A}} \leq c|x(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{2}(t-(T-\max\tau(0,0))/\eta)}, \tag{B.18}$$

with $c > 0$. Moreover, using the structure of the hybrid time domains, all hybrid times $(t,j) \in \mathrm{dom}\,(x)$ satisfy

$$-\frac{\lambda}{2}t \leq -\frac{1}{3N}\lambda\,(t+j) + \frac{\lambda L}{3}, \tag{B.19}$$

for all $\lambda > 0$. Hence, we obtain:

$$|x(t,j)|_{\mathcal{A}} \leq \hat{c}_s|x(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{3N}(t+j)}, \tag{B.20}$$

where $\hat{c}_s \coloneqq ce^{\lambda L\left(\frac{1}{3}+\frac{1}{2\eta}\right)}$. Inequality (B.20) implies that $\mathcal{A}$ is UGES under $\mathcal{H}_s$. ■

Lemma B.4. Under the conditions of Theorem 4.1-(i_3), system $\mathcal{H}_s$ renders UGES the set $\mathcal{A}$.

Proof. Using the Lyapunov function V given by (B.3), and the fact that $\mathcal{A}_{\mathrm{NE}} = \{q^*\}$, we obtain again inequality (B.13) during flows. Since now $\alpha = \mathbf{0}_n$, during jumps we have

$$\begin{aligned}\Delta_j^{j+1}V(x) \leq -V_1(z) - V_2(z) - \frac{\tilde{P}(q)}{N}\sum_{i=1}^{N}(\tau_i^2 - \tau_i^{2+}) \\ + \frac{1}{4}\left(\sum_{i=1}^{N}\left((p_i^+ - q_i^+) + (p_i^+ - q_i^{*+})\right)\right).\end{aligned}$$

For each solution x and each $(t,j) \in \mathrm{dom}(x)$ such that $x(t,j) \in D_1$ and $x(t,j+1) \in G_1(x(t,j))$, consider the sets $\mathcal{I}_{j+1}(t), \mathcal{I}_{j+1}^+(t) \subset \{1,2,\ldots,n\}$, which satisfy the following: 1) $\mathcal{I}_{j+1}(t) \cap \mathcal{I}_{j+1}^+(t) = \{\emptyset\}$, $\mathcal{I}_{j+1}(t) \cup \mathcal{I}_{j+1}^+(t) = \{1,2,\ldots,n\}$; and,

letting $\text{card}(A)$ denote the cardinality of a set $A \subset \mathbb{Z}_{\geq 0}$, 2) they are generated recursively as $\text{card}(\mathcal{I}_{j+1}(t)) = \text{card}(\mathcal{I}_j(t)) - 1$, and $\text{card}(\mathcal{I}^+_{j+1}(t)) = \text{card}(\mathcal{I}^+_j(t)) + 1$, with $\text{card}(\mathcal{I}_{\underline{j}}(t)) = N$ and $\text{card}(\mathcal{I}^+_{\underline{j}}(t)) = 0$, where $\underline{j} \coloneqq \min\{j \in \mathbb{Z}_{\geq 0} : (t,j) \in \text{dom}(x), x(t,j) \in D_1\}$. Since the reset policy $\alpha_i = 0$ for all $i \in \mathcal{V}$ corresponds to resets of the form $\tau_i^+ = T_0$, $p_i^+ = q_i$, $\forall\ i \in \mathcal{V}$, and since every jump corresponds to one and only one reset, the change of the Lyapunov function at each of the j^{th} jumps that follow each interval of flow in the hybrid time domain of a solution to $\mathcal{H}_s$ satisfies

$$\begin{aligned}\Delta_j^{j+1} V(x) \leq -V_1(x) - V_2(x) - (\tau_s^2 - T_0^2)\frac{\tilde{P}(q)}{N} \\ + \frac{1}{4}\sum_{i \in \mathcal{I}_{j+1}(t)} \left((p_i - q_i)^2 + (p_i - q_i^*)^2\right) \\ + \frac{1}{4}\sum_{i \in \mathcal{I}^+_{j+1}(t)} (q_i - q_i^*)^2. \end{aligned} \tag{B.21}$$

It follows that after N consecutive jumps, the change of the Lyapunov function satisfies

$$\Delta_j^{j+N} V(x) \leq -V_1(x) - V_2(x) - \frac{1}{2}(\tau_s^2 - T_0^2)\tilde{P}(q) + \frac{1}{4}|q - q^*|^2. \tag{B.22}$$

By strong convexity of $\tilde{P}$, we can further bound (B.22) as:

$$\begin{aligned}\Delta_j^{j+N} V(x) &\leq -V_1(x) - V_2(x) - \gamma\left(\kappa^{-1}\right) V_3(x) \\ &\leq -\gamma\left(\kappa^{-1}\right) V(x), \end{aligned} \tag{B.23}$$

where $\gamma(\cdot)$ is given by (4.19), and satisfies $\gamma\left(\kappa^{-1}\right) \in (0,1)$ under (4.16). Thus, by [223, Thm. 1], inequalities (B.13) and (B.23), and the quadratic upper and lower bounds of V, we obtain that $\mathcal{H}_s$ renders UGES the set $\mathcal{A}$. ■

Now, by using Lemmas B.2-B.4, which establish suitable stability properties for the "synchronized" HDS $\mathcal{H}_s$, we proceed to proof the three main items of Theorem 4.1 for the original hybrid dynamics $\mathcal{H}_1$.

(a) Stability: By the hybrid reduction principle [16, Cor. 7.24], UGAS of $\mathcal{A}$ for system $\mathcal{H}_s$ (established in Lemmas B.2, B.3 and B.4), and UGFxS of $\mathcal{A}_\nu$ for system $\mathcal{H}_\nu$, imply that $\mathcal{A}$ is UGAS for system $\mathcal{H}_\nu$. Moreover, since the choice of $\nu > 0$ is arbitrary, and has no effect on the dynamics of the system, and since the trajectories of the original HDS $\mathcal{H}_1$ are complete and bounded, the compact set $\mathcal{A}$ is also UGAS for system $\mathcal{H}_1$. This establishes UGAS of $\mathcal{A}$ under the conditions of items (i_1), (i_2) and (i_3). For items (i_2) and (i_3), UGES follows by the exponential convergence bounds of Lemmas B.3-B.4 and the fixed-time synchronization of τ. UGAS and UGES follow directly by robustness results of well-posed HDS, specifically by [16, Thm. 7.21].

(b) Convergence Bounds: For any solution x and all $(t, j) \in \mathrm{dom}\,(x)$ such that $t + j \geq (T - T_0)/\eta + N =: T^*$ we have that $|\tau(t, j)|_{\mathcal{A}_{\mathrm{sync}}} = 0$. Thus, for such times the trajectories of $\mathcal{H}_1$ satisfy the Lyapunov inequalities established in Lemmas B.2-B.4. To establish (4.17), we use inequality (B.9), which implies that for each $(t, j), (s, j) \in \mathrm{dom}\,(x)$, such that $t > s$, and $s + j \geq T^*$, we have $V(t, j) \leq V(s, j)$. Since $V_3(x) \leq V(x)$ for all $x \in \mathbb{R}^{2nN+N}$, and using $s_j \coloneqq \min\{t \in \mathbb{R}_{\geq 0} \ : \ (t, j) \in \mathrm{dom}\,(x)\,,\ t + j \geq T^*\}$, we obtain

$$\tilde{P}(q(t, j)) \leq \frac{N}{\tau^\top \tau} V(s_j, j) = \frac{c_j}{\tau_s^2}, \quad \forall t \ > s_j, \tag{B.24}$$

where $c_j \coloneqq V(s_j, j)$. Using the fact V is non-increasing during flows and jumps, and also converges to zero, we obtain that $\{c_j\}_{j=0}^{\infty} \to 0^+$. To obtain the convergence bound of item (i_2), we first note that from the proof of Lemma 4.1 it follows that $\frac{d|x(t,j)|_{\mathcal{A}}}{dt} \leq \tilde{\ell}|x(t + j)|_{\mathcal{A}}$ for all $(t, j) \in \mathrm{dom}\,(x)$, where $\tilde{\ell} = 2\sqrt{2}\max\left\{\frac{1}{T_0}, T\ell\right\}$. In particular, this implies that

$$|x(t_s, j_s)|_{\mathcal{A}} \leq |x(0, 0)|_{\mathcal{A}} e^{\tilde{l}(T - \max \tau(0,0))/\eta}, \tag{B.25}$$

where (t_s, j_s) are the smallest times for which $|\tau(t,j)|_{\mathcal{A}_{sync}} = 0$ for all $t + j \geq t_s + j_s$. Note that $x(t_s, j_s) \in C_s \cup D_s$, and hence (B.18) holds with $|x(0,0)|_{\mathcal{A}}$ replaced by $|x(t_s, j_s)|_{\mathcal{A}}$, i.e., $|x(t,j)|_{\mathcal{A}}$ satisfies:

$$|x(t,j)|_{\mathcal{A}} \leq c|x(t_s, j_s)|_{\mathcal{A}} e^{-\frac{\lambda}{2}(t-(T-\max \tau(0,0))/\eta)}, \tag{B.26}$$

for all $t + j \geq t_s + j_s$. Therefore, inequality (B.26), together with (B.25), implies:

$$|x(t,j)|_{\mathcal{A}} \leq \tilde{c}|x(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{2}t}, \tag{B.27}$$

for all $(t,j)\text{dom}\,(x)$, and where $\tilde{c} = ce^{\left(\frac{\lambda}{2}+\tilde{l}\right)L}$. Using (B.19) and (B.27) gives:

$$|x(t,j)|_{\mathcal{A}} \leq \hat{c}|x(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{3N}(t+j)}, \tag{B.28}$$

with $\hat{c} = \tilde{c}e^{\lambda L/3}$ which establishes the bound in (4.18). This also implies that $\mathcal{H}_1$ renders $\mathcal{A}$ UGES under the conditions of Theorem 4.1-(i_2). Finally, to establish the convergence bound of item (i_3), we note that (B.23) implies $V(x(t, j+N)) \leq (1 - \gamma\,(\kappa^{-1}))V_3(x(t,j))$. Since $V_3(x) \leq V(x)$ for all $(t,j) \in \text{dom}\,(x)$ such that $t + j \geq T^*$, V does not increase during flows, and using the periodicity of the hybrid time domains, we obtain:

$$V_3(t, j_s + kN) \leq (1 - \gamma\left(\kappa^{-1}\right))^k V_3(t_s, j_s), \quad \forall\ k \in \mathbb{Z}_{\geq 0}, \tag{B.29}$$

for all $t \in (t_s + (k-1)L, t_s + kL)$, where (t_s, j_s) denotes the first hybrid time after which the timers τ flow synchronized. By Lemma 4.1, such times are uniformly bounded as $0 \leq t_s + j_s \leq 2T^*$. Using (B.29), the definition of V_3, as well as strong convexity and smoothness of $\tilde{P}$, we obtain:

$$|q(t, j_s + kN) - q^*| \leq \sigma_r \sqrt{\frac{\ell}{\kappa}} (1 - \gamma\left(\kappa^{-1}\right))^{\frac{k}{2}} |q(t_s, j_s) - q^*|, \tag{B.30}$$

for all $k \in \mathbb{Z}_{\geq 0}$. Finally, since by Lemma 4.1 all solutions are bounded, it follows that for each compact set of initial conditions K_0 there exists $M_0 > 0$ such that $|x(t,j)|_{\mathcal{A}} \leq M_0$ for all $(t,j) \in \mathrm{dom}(x)$ such that $0 \leq t \leq t_s$ and $0 \leq j \leq j_s$. This bound, combined with (B.30), implies the bound of the theorem via the change of variable $j = j_s + kN$ and the upper bound $n \leq j_s \leq 2n$. ■

B.2 Proofs of Section 4.4

Proof of Lemma 4.2

First, by Assumption (4.1)-(i_1), the system of equations characterizing Nash flows (4.22) reduces to:

$$Az_q^f + b + \delta \cdot q + \mu \mathbf{1}_N - \lambda = 0, \tag{B.31a}$$

$$z_q^f \in \Delta, \quad \lambda_i z_{i,q}^f = 0, \quad \lambda_i \geq 0 \qquad \forall i \in \mathcal{V}, \tag{B.31b}$$

where $\lambda \coloneqq (\lambda_1, \cdots, \lambda_N)$. Moreover, by Assumption (4.1)-(i_2), it follows that $\lambda_i = 0 \; \forall i \in \mathcal{V}$. Thus, we have that (B.31) can be written as

$$\begin{pmatrix} A & \mathbf{1}_N \\ \mathbf{1}_N^\top & 0 \end{pmatrix} \begin{pmatrix} z_q^f \\ \mu \end{pmatrix} = \begin{pmatrix} -(b + \delta \cdot q) \\ 1 \end{pmatrix}. \tag{B.32}$$

By solving (B.32) for z_q^f we obtain that

$$\begin{aligned} z_q^f = &- \left(I - A^{-1} \frac{\mathbf{1}_N \mathbf{1}_N^\top}{\mathbf{1}_N^\top A^{-1} \mathbf{1}_N} \right) A^{-1}(b + \delta \cdot q) \\ &+ \frac{A^{-1}\mathbf{1}_N}{\mathbf{1}_N^\top A^{-1} \mathbf{1}_N}. \end{aligned}$$

Letting $\mathcal{O}(q) \coloneqq z_q^f$ for all $q \in \mathbb{R}^N$ obtains the result. ■

Proof of Lemma 4.3

We solve for an optimal incentive that makes a Nash flow state maximize the social welfare of the system in the sense of (4.23). To do so, consider the KKT conditions characterizing the socially optimal flow (4.24), and replace z^* with the Nash flow $\mathcal{O}(q^*)$ at the optimal incentive q^*:

$$-\frac{\partial c_i}{\partial z_i}(\mathcal{O}(q^*)) \cdot [\mathcal{O}(q^*)]_i - c_i([\mathcal{O}(q^*)]_i) + \tilde{\mu} - \tilde{\lambda}_i = 0 \tag{B.33a}$$

$$\mathcal{O}(q^*) \in \Delta, \quad \tilde{\lambda}_i [\mathcal{O}(q^*)]_i = 0, \quad \tilde{\lambda}_i \geq 0 \quad \forall i \in \mathcal{V}. \tag{B.33b}$$

Note that $\mathcal{O}(q^*) \in \text{relint}(\Delta)$ via Assumption 4.1-(i_2) and the fact that $\mathcal{O}(q^*)$ is a Nash flow satisfying (4.22). Therefore, from (B.33b), we obtain that $\tilde{\lambda}_i = 0, \ \forall i \in \mathcal{V}$. This fact, together with Assumption-(i_1), reduces (B.33a) to

$$2A\mathcal{O}(q^*) + b = \tilde{\mu}\mathbf{1}_N, \quad \tilde{\mu} \in \mathbb{R}, \tag{B.34}$$

where we have used the fact that $\frac{\partial c}{\partial q}(\tilde{q}) = A, \ \forall \tilde{q} \in \mathbb{R}^n$. On the other hand, let $L \in \mathbb{R}^{N \times N}$ be an arbitrary Laplacian matrix. It follows that $L\mathbf{1}_N = 0$ and thus that

$$\begin{aligned} LA\mathcal{O}(q) &= -\left(L - L\mathbf{1}_N \frac{\mathbf{1}_N^\top A^{-1}}{\mathbf{1}_N^\top A^{-1}\mathbf{1}_N}\right)(b + \delta \cdot q) \\ &= -L(b + \delta \cdot q), \end{aligned} \tag{B.35}$$

for all $q \in \mathbb{R}^N$, and where we have used the definition of Q in (4.26). Therefore, left-multiplying (B.34) by Q, recalling that Q is a Laplacian matrix, and using (B.35), we obtain that

$$-2Q(b + \delta \cdot q^*) + Qb = Q\mu\mathbf{1}_N$$

$$\implies Qq^* = Q\left(\mu \mathbf{1}_N - \frac{b}{2\delta}\right),\ \mu \in \mathbb{R}$$
$$\implies q^* = -\frac{b}{2\delta} + \mu \mathbf{1}_N,\ \mu \in \mathbb{R},$$

which obtains the result. ■

Proof of Lemma 4.4

We first rewrite $\mathcal{G}^{\mathcal{O}}(q)$ using the fact that $c(q) = Aq + b$ by Assumption 4.1, together with the result of Lemma 4.2:

$$\mathcal{G}^{\mathcal{O}}(q) = 2A\mathcal{O}(q) + b \tag{B.36}$$
$$= 2A\left(-Q\left(b + \delta \cdot q\right) + \alpha\right) + b \tag{B.37}$$

In particular, (B.37) implies that $\mathcal{G}^{\mathcal{O}}(q)$ is $\ell_{\mathcal{G}^{\mathcal{O}}}$-Lipschitz with $\ell_{\mathcal{G}^{\mathcal{O}}}$ satisfying

$$\ell_{\mathcal{G}^{\mathcal{O}}} \leq 2\delta\left(1 + \frac{n\underline{a}^{-1}}{\sum_{i=1}^{N}\frac{1}{a_i}}\right) \tag{B.38}$$

where $a \in \mathbb{R}^n$ is such that $A = \mathrm{diag}(a)$, and $\underline{b} \coloneqq \min_{i\in\mathcal{V}}(b_i)$ for $b \in \mathbb{R}^N$. Moreover, since Q is a Laplacian matrix, from (B.37) it follows that $\mathcal{G}^{\mathcal{O}}\left(q^{\square}\right) = 0$ for all $q^{\square} \in \mathcal{B}_q$, where

$$\mathcal{B}_q \coloneqq \Bigg\{q \in \mathbb{R}^n\ :\ q = \frac{1}{\delta}\left[Q^{\dagger}\left(A^{-1}\frac{b}{2} + \alpha\right) - \frac{b}{2} + \mu\mathbf{1}_N\right],\ \mu \in \mathbb{R}\Bigg\},$$

and $Q^{\dagger}$ is the pseudoinverse of Q.

Step 1 (Absence of finite escape times): Follows directly by noting that the flow map of the HOPD is globally Lipschitz and via the application of the comparison principle.

Step 2 (Completeness of Solutions): By [275], the dynamics of the timers always generate

complete solutions. On the other hand, by the properties of the Laplacian $\mathcal{L}$, we have that $\mathbf{1}_N^\top \dot{p} = 0$. Thus, the state p always remains in $\ker(\mathcal{L})^\perp$, which is unbounded. Since the state q evolves in $\mathbb{R}^N$, and there are no finite escape times, every solution of the HDS is complete.

Step 3 (Synchronization in finite time): Achieved by the results of [275] for strongly connected graphs. The synchronization state is reached in finite time such that $\tau(t,j) = \tau_s \mathbf{1}_N,\ \tau_s \in [T_0, T]$ for all $t + j \geq T^*$ and where $T^* \coloneqq 2\,(T - T_0) + N$.

Step 4 (Fixed-Time Stability of $\mathcal{A}_{sync}$): First, note that since the timer state of the system satisfies $\tau(i,j) = \tau_s \mathbf{1}_N$ with $\tau_s \in [T_0, T]$, it follows that $\mathcal{L}\mathcal{T} = \tau_s \mathcal{L}$. Moreover, we have that

$$\begin{aligned}
\frac{d\left(\mathbf{1}_N^\top p(t,j)\right)}{dt} &= \mathbf{1}_N^\top \dot{p} \\
&= \mathbf{1}_N^\top \left(2\tau_s \mathcal{L}\mathcal{G}^{\mathcal{O}}(q(t,j))\right) \\
&= 2\tau_s \mathbf{1}_N^\top \mathcal{L}\mathcal{G}^{\mathcal{O}}(q(t,j)) \\
&= 0,
\end{aligned} \tag{B.39}$$

$$\begin{aligned}
\frac{d\left(\mathbf{1}_N^\top q(t,j)\right)}{dt} &= \mathbf{1}_N^\top \dot{q} \\
&= \frac{2}{\tau_s} \mathbf{1}_N^\top \left(p(t,j) - q(t,i)\right) \\
&= \frac{2}{\tau_s} \mathbf{1}_N^\top \left(p(t,j) - q(t,j)\right) \\
&= -\frac{2}{\tau_s} \mathbf{1}_N^\top q(t,j) \\
&\implies \mathbf{1}_N^\top q(t,j) = e^{-2\ln\left(\frac{2t}{\tau_s(0,0)}+1\right)} \mathbf{1}_N^\top q(t_j, j),
\end{aligned} \tag{B.40}$$

where $t_j \coloneqq \min\{t \in \mathbb{R}_{>0} \ :\ (t,j) \in \mathrm{dom}\,(x)\}$. Due to $p(0,0) \in \ker(\mathcal{L})^\perp$, together with the fact that $(q,p)^+ = (q,p)$ during jumps, we have that: 1) by using (B.39) that $\ker(\mathcal{L})^\perp$ is strongly forward invariant for p, and 2) after the first interval of flow followed by n successive jumps, from (B.40), that during $\mathbf{1}_N^\top q(t,j) = e^{-2\ln\left(\frac{2t}{\tau_s(0,0)}+1\right)} \mathbf{1}_N^\top p(t_j, j) = 0$ for

all $(t, j) \in \operatorname{dom}(x)$ such that $j \geq N$, and thus that q converges to $\ker(\mathcal{L})^{\perp}$ in fixed time. Therefore, since we have that the set of initial conditions is arbitrary for q, it follows that in fact $\mathcal{A}_{\text{sync}}$ is UGFxS under the hybrid dynamics of $\mathcal{H}_1^{\text{HOPD}}$. ■

Proof of Theorem 4.2

Consider the restricted hybrid dynamical system

$$\mathcal{H}_{1,s}^{\text{HOPD}} \coloneqq \left(C_1^{\text{HOPD}} \cap \mathcal{A}_{\text{sync}}, F_1^{\text{HOPD}}, D_1^{\text{HOPD}} \cap \mathcal{A}_{\text{sync}}, G_1^{\text{HOPD}}\right). \tag{B.41}$$

We first rewrite the $\dot{p}$ vector field in terms of the deviation variable $\tilde{q} \coloneqq q - \Pi_{\mathcal{A}_q}(p)$, where $\Pi_{\mathcal{A}_q}(p)$ denotes the projection of p onto $\mathcal{A}_q$, and by using the definition of $\mathcal{O}(q)$ in (4.26):

$$\begin{aligned}
\dot{p} &= 2\tau_s \mathcal{L}\left(2A\mathcal{O}(q) + b\right) \\
&= 2\tau_s \mathcal{L}\left(-2AQ(\delta \cdot q + b) + 2A\alpha + b\right) \\
&= 2\tau_s \left(-2\mathcal{L}\left(\delta \cdot q + b\right) + 2\mathcal{L}A\alpha + \mathcal{L}b\right) \\
&= 2\tau_s \left(-2\mathcal{L}\left(\delta \cdot q + b\right) + 2\mathcal{L}\frac{\mathbf{1}_N}{\mathbf{1}_N^{\top} A^{-1} \mathbf{1}_N} + \mathcal{L}b\right) \\
&= 2\tau_s \left(-2\mathcal{L}\left(\delta \cdot q + b\right) + \mathcal{L}b\right) \\
&= 2\tau_s \left(-2\mathcal{L}\left(\delta \cdot [q^* + \tilde{q}] + b\right) + \mathcal{L}b\right) \\
&= 2\tau_s \left(-2\mathcal{L}\left(\delta \cdot \left[-\frac{b}{2\delta} + \tilde{q}\right] + b\right) + \mathcal{L}b\right) \\
&= -2\tau_s \cdot 2\delta \mathcal{L}\tilde{q},
\end{aligned}$$

where we have used the definitions of α and q^* in Lemmas 4.2 and 4.3, respectively, together with the fact that $\mathcal{L}$ is a Laplacian matrix and thus that $\mathcal{L}AQ = \mathcal{L}$ since $\mathcal{L}\mathbf{1}_N = 0$.
Now, consider the Lyapunov function given by

$$V(x) = \frac{1}{4}|p - q|^2 + \frac{1}{4}|p|_{\mathcal{A}_q}^2 + \delta \frac{|\tau|^2}{N} \tilde{q}^{\top} \mathcal{L}\tilde{q}.$$

Using the fact that $\nabla_p \left(\frac{1}{2}|p|_{\mathcal{A}_q}\right)^2 = p - \Pi_{\mathcal{A}_q}(p) =: \tilde{p}$, the time derivative of V along the flows of the HOPD satisfies

$$\begin{aligned}\dot{V} =& -\frac{1}{\tau_s}(p-q)^\top(p-q) + 4\tau_s(p-q)^\top\delta\mathcal{L}\tilde{q} \\ & - \tau_s\gamma\Big((p-q)+(p-\Pi_{\mathcal{A}_q}(p))\Big)^\top 2\delta\mathcal{L}\tilde{q} + \tau_s\delta\tilde{q}^\top\mathcal{L}\tilde{q} \\ \leq& -\frac{1}{T}|\tilde{p}-\tilde{q}|^2 - \gamma\delta T_0\tilde{q}^\top\mathcal{L}\tilde{q}, \end{aligned} \tag{B.42}$$

where $\tilde{p} \coloneqq p - p^*$, and where we have used the fact that $\hat{q}^\top\mathcal{L} = \hat{q}^\top\mathcal{L}^\top = q^*,\ \forall \hat{q} \in \mathcal{A}_q$ and $\mathcal{L} = \mathcal{L}^\top$ since $\mathcal{L}$ is the Laplacian matrix of a connected undirected graph. Now, during jumps we have that

$$\begin{aligned} V(x^+) - V(x) =& -\frac{|p-q|^2}{4} - \frac{\delta}{2}\left(T^2 - T_0^2\right)\tilde{q}^\top\mathcal{L}\tilde{q} \\ =& -\left(1 - \frac{T_0^2}{T^2} - \frac{1}{2T^2\gamma\delta\sigma_2\left(\mathcal{L}\right)}\right)V(x), \end{aligned} \tag{B.43}$$

where we have used the fact that $\sigma_2\left(\mathcal{L}\right)|q|^2 \leq q^\top\mathcal{L}q \leq \sigma_N\left(\mathcal{L}\right)|q|^2$ for all $q \in \ker\left(L\right)^\perp$, together with the fact that $\ker\left(\mathcal{L}\right)^\perp$ is invariant under $\dot{\tilde{q}}$ and $\dot{\tilde{p}}$, where $\tilde{p} \coloneqq p - q^*$. Using similar arguments to the proof of [1, Thm 1-(i_3)], UGES of $\mathcal{A}_q$ and the convergence bound of the theorem, follow by the fact that $T^2 - T_0^2 > \frac{1}{2\gamma\delta\sigma_2(\mathcal{L})}$ together with (B.43), the decrease during flows of (B.42) in $\ker\left(\mathcal{L}\right)^\perp$, and the invariance of $\ker\left(\mathcal{L}\right)^\perp$ for p and q. ■

B.3 Proofs of Section 4.5

The convergence and stability analysis of the HDS $\mathcal{H}_A$ starts with the following lemma. The proof is almost identical to the proofs of [51, Prop. 1] and [275, Thm. 1], and it is presented here only for the sake of completeness.

Lemma B.5. Consider the hybrid dynamical system

$$\tau \in C_\tau \coloneqq [T_0,\ T]^N\,, \qquad \dot{\tau} = \frac{1}{2}\mathbf{1}_N, \tag{B.44a}$$

$$\tau \in D_\tau \coloneqq \left\{\tau \in C_\tau : \max_i \tau_i = T\right\}, \qquad \tau^+ \in \overline{G}_\tau(\tau), \tag{B.44b}$$

where $\overline{G}_\tau$ is the osc hull of the set-valued mapping resulting from projection of G_1^{HARDD} into the τ-component, which is independent of (q, p). If r_i satisfies $r_i \in \left(0, \frac{T-T_0}{N}\right)$ for all $i \in \mathcal{V}$, then every solution is complete and uniformly non-Zeno, the set $\mathcal{A}_\tau$ is UGAS, and every solution satisfies $|\tau(t,j)|_{\mathcal{A}_\tau} = 0$, for all $(t,j) \in \text{dom}(\tau)$ such that $t + j \geq 2(T - T_0) + N$.

Proof. Absence of finite escape times follows by compactness of the flow set and jump set. Being uniformly non-Zeno follows by the fact that after n jumps the system is necessarily synchronized and the timers satisfy $\tau \in [T_0, T]^N \backslash D_\tau$, which implies that the system has to flow. Since the intervals of flow have a maximum duration of $T - T_0$, it follows that there can be at most N consecutive jumps in any interval of length $2(T - T_0)$. This also implies completness of solutions. To show UGAS of $\mathcal{A}_\tau$ we define a Lyapunov function $V : [T_0, T]^N \to \mathbb{R}_{\geq 0}$ to be the the infimum of the lengths of all arcs that touch all timers (see Figure B.1 for an illustration), where the points T_0 and T in the interval $[T_0, T]$ are identified to be the same to form a circle. Since all the timers have the same frequency, during the flows the Lyapunov function does not change, i.e., $\dot{V}(\tau) = 0$. Moreover, during jumps the Lyapunov function cannot increase its value since jumps only happen whenever one or more timers satisfy the condition $\tau_i = T$, which either leaves the timers in the same position of the circle, or forces some of the timers to go to T. In both cases, $V(\tau^+)$ does not increase. To show that V converges to zero in a fixed-time, we note that for any initial condition $\tau(0,0) \in [T_0, T]^N$ the Lyapunov function V

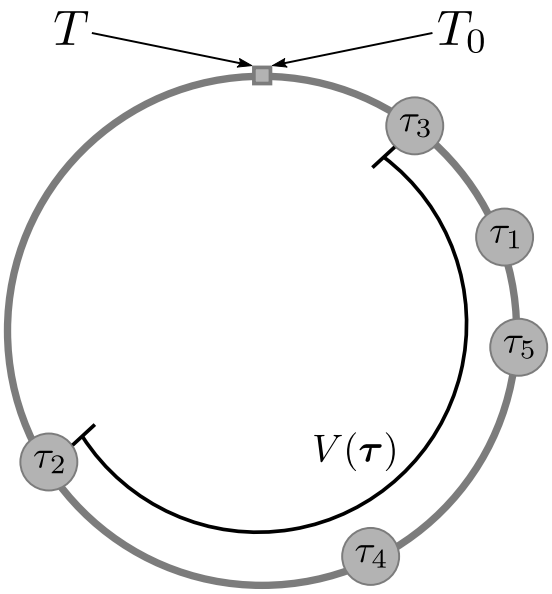


Figure B.1. Representation of the Lyapunov function used in proof of Lemma B.5

always satisfies $V(\tau) \leq T - T_0\left(1 - \frac{1}{n}\right)$. Since all timers have the same frequency, and since $r_i \in (0, \frac{T-T_0}{N})$ for all i, there will exist a time $0 \leq t < 2T - T_0$ and some $j \in \{0, 1, \ldots, n\}$ such that $\tau_i(t, j) > r_i$ for all $i \in \mathcal{V}$. From this point, since the graph is connected and undirected, any jump induced by an agent j satisfying $\tau_j = T$ will be followed by at most $n - 1$ jumps after which all timers will be synchronized at the position $\tau_i = T_0$, which implies $V(\tau) = 0$. From this point, the system remains synchronized. Since no complete solution keeps V equal to a non-zero constant, UGAS of the set $\mathcal{A}_\tau$ follows now directly by the Hybrid Invariance Principle [33, Thm. 8.8]. ■

Proof of Theorem 4.3

We divide the proof in seven main steps. Refer to Figure B.2 for an overview and visualization of the main aspects of the proof.

Step 1: Absence of Finite Escape Times

First, note that the function F_1^{HARDD} is continuous in C_1^{HARDD}. Also, by item (b) in

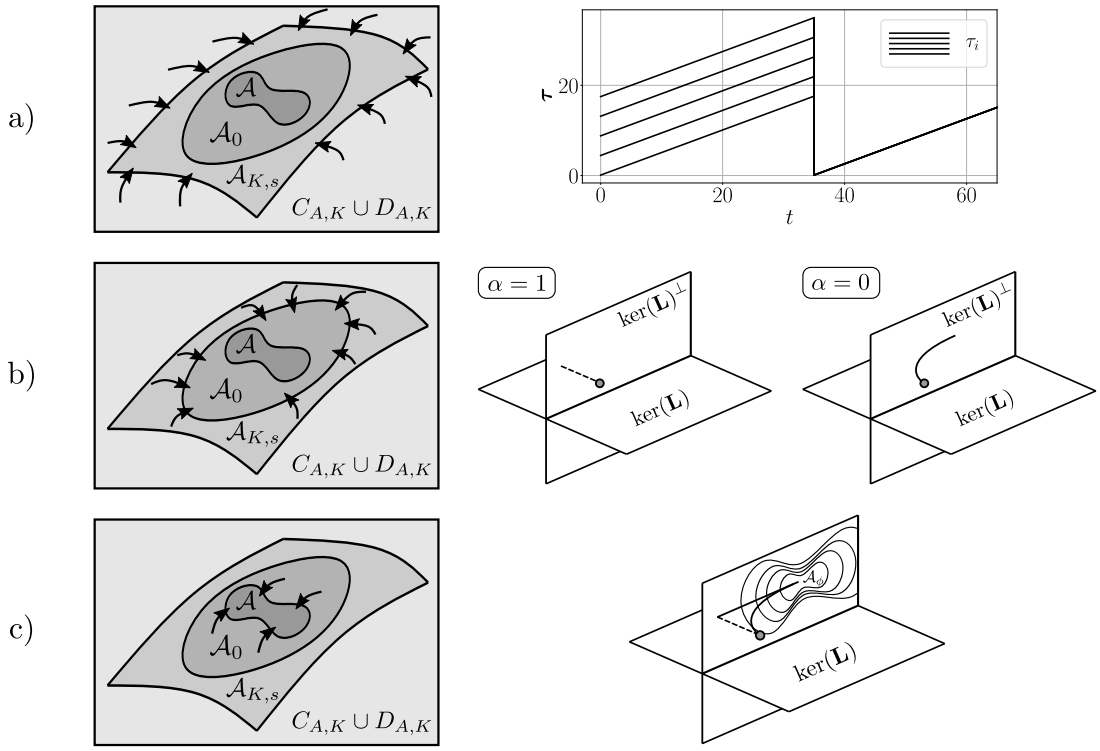


Figure B.2. To prove UGAS of the set $\mathcal{A}$:
a) First, UGAS of the set $\mathcal{A}_{K,s}$ from $C_{1,K} \cup D_{1,K}$ is proven by the fixed-time synchronization of the timers.

b) Second, when $\nabla F(z)$ is globally Lipschitz ($\alpha = 1$), the set $\ker(\mathbf{L})^\perp$ is UGAS via the jump map $G_1^{\text{HARDD}}(p)$. When $\alpha = 0$, UGAS is achieved via the flow map (4.48). This fact, together with strong forward invariance of $\ker(\mathbf{L})^\perp$ and A_τ, for p and τ respectively, render the set $\mathcal{A}_0$ UGAS from $A_{K,s}$.
c) Third, with the Lyapunov function in (B.47), UGAS of $\mathcal{A}$ from $\mathcal{A}_0$ is guaranteed under the HDS $\mathcal{H}_1^{\text{HARDD}}$.
d) Last, by the nested application of the Hybrid Reduction Principle [33, Cor. 7.24] UGAS of $\mathcal{A}$ is guaranteed.

Lemma 4.5, the gradient $\nabla\phi$ is globally Lipschitz. It then follows that since $T_0 \leq \tau_i \leq T$ for all i, we have:

$$|\dot{p}| \leq 2\gamma(T)|\mathbf{L}z^*(\mathbf{L}q)| \leq 2\gamma(T)\ell_\phi|q - p^*|,$$
$$|\dot{q}| \leq 2|\mathcal{T}^{-1}(p - q)| \leq \frac{2}{T_0}|p - q|,$$

for any $p^* \in \mathcal{A}_\phi$. Combining these inequalities, and the Comparison Lemma [22] we obtain

that the flows (4.48) always generate bounded signals (q, p), which rules out finite escape times.

Step 2: Completeness of Solutions

We show that solutions cannot stop due to flows or jumps leaving $C_1^{\text{HARDD}} \cup D_1^{\text{HARDD}}$. By Lemma B.5, the dynamics of the timers always generate complete solutions. On the other hand, by the properties of the Laplacian L, we have that $\mathbf{1}_{nN}^\top \dot{p} = 0$. Thus, the state y always remains in $\ker(\mathbf{L})^\perp$, which is unbounded. Since the state x evolves in $\mathbb{R}^{np}$, every solution of the HDS is complete and the hybrid time domains of the solutions are generated by the hybrid time domains of the HDS (B.44). This establishes properties (P.1) and (P.2).

Step 3: Fixed-Time Synchronization of Restarting Mechanisms

Let $k > 0$ and define the compact set $K := \mathcal{A}_\phi + k\mathbb{B}$. Let us restrict the data of the original HDS $\mathcal{H} := \{C_1^{\text{HARDD}}, F_1^{\text{HARDD}}, D_1^{\text{HARDD}}, G_1^{\text{HARDD}}\}$ by intersecting with K the (q, p)-components of the flow set and the jump set. The resulting HDS has data $\mathcal{H}_K := \{C_{1,K}, F_1^{\text{HARDD}}, D_{1,K}, G_1^{\text{HARDD}}\}$, where

$$C_{1,K} := K \times (\ker(\mathbf{L})^\perp \cap K) \times [T_0, T]^N \tag{B.45a}$$

$$D_{1,K} := K \times (\ker(\mathbf{L})^\perp \cap K) \times D_\tau. \tag{B.45b}$$

Since the dynamics of the timers τ are independent of x and y, by the definition of UGAS and by Lemma B.5 the restricted HDS $\mathcal{H}_K$ renders UGAS the compact set

$$\mathcal{A}_{K,s} := K \times (\ker(\mathbf{L})^\perp \cap K) \times \mathcal{A}_\tau. \tag{B.46}$$

Step 4: Asymptotic Stability of Feasible Set

Let us now further restrict the flow and the jump sets of the HDS $\mathcal{H}_K$ with the set $\mathcal{A}_{K,s}$.

We denote this new restricted HDS as $\mathcal{H}_{K,s} := \{C_{A,K,s}, F_1^{\text{HARDD}}, D_{A,K,s}, G_1^{\text{HARDD}}\}$, where

$$C_{A,K,s} = K \times (\ker(\mathbf{L})^\perp \cap K) \times ([T_0, T]^n \cap \mathcal{A}_\tau)\,,$$
$$D_{A,K,s} = K \times (\ker(\mathbf{L})^\perp \cap K) \times (D_\tau \cap \mathcal{A}_\tau)$$

In this HDS, during flows the timers satisfy $\tau = \alpha \mathbf{1}_N$ where $\alpha \in [T_0, T]$.

Lemma B.6. For the HDS $\mathcal{H}_{K,s}$ the compact set $\mathcal{A}_0 = (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \mathcal{A}_\tau$ is UGAS. Moreover, if $\alpha = 0$, then every complete solution satisfies $q(t,j) \in \ker(\mathbf{L})^\perp$ for all $(t,j) \in \text{dom}(x)$ such that $t + j \geq 2N + 4(T - T_0)$.

Proof. We show that $\mathcal{A}_0$ is strongly forward invariant and globally uniformly attractive, which implies UGAS via [33, Prop. 7.5]. Let $x(0,0) \in \mathcal{A}_0$. By Lemma B.5 the set $\mathcal{A}_\tau$ is strongly forward invariant under the dynamics of τ. By Step 2, the set $\ker(\mathbf{L})^\perp$ is strongly forward invariant for the state p. Moreover, by definition of the flow set $C_{A,K,s}$, during flows $\tau(t,j) = \alpha \mathbf{1}_N$ with $\alpha \in [T_0, T]$. Thus, $\mathbf{1}_{np}^\top \dot{q} = 0$ which implies that $\mathbf{1}_{nN}^\top q(t,0) = 0$ for all $(t,0) \in \text{dom}(x)$ since $q(0,0) \in \ker(\mathbf{L})^\perp$. Because $q^+ = q$ ($\alpha = 0$) or $q^+ = p$ ($\alpha = 1$), the state q remains in $\ker(\mathbf{L})^\perp$ also during jumps. To show that $\mathcal{A}_0$ is uniformly attractive it suffices to show that every complete trajectory of q converges to $\ker(\mathbf{L})^\perp$. Let $\alpha = 0$ and consider the auxiliary variable $\tilde{q} = \mathbf{1}_{nN}^\top q$, with dynamics $\dot{\tilde{q}} = 2\mathbf{1}_{nN}^\top \mathcal{T}^{-1}(p - q) = \frac{2}{\alpha}\mathbf{1}_{nN}^\top (p - q)$, where we used again the fact that $\tau = \alpha \mathbf{1}_N$ during flows. Since $\mathbf{1}_{nN}^\top p = 0$ always hold, we obtain $\dot{\tilde{q}} = -2\frac{1}{\alpha}\tilde{q}$, which implies that $\tilde{q}$ converges exponentially fast to zero, i.e., $q(t,j)$ converges to $\ker(\mathbf{L})^\perp \cap K$ exponentially fast during flows. When $\alpha = 1$, convergence of q to $\ker(\mathbf{L})^\perp$ happens in finite time after the n consecutive jumps induced by the synchronized timers τ, and the fact that $q_i = p_i$ after each jump. ■

Step 5: Asymptotic Stability of the Optimal Set

Having established UGAS of $\mathcal{A}_0$ for the HDS $\mathcal{H}_{K,s}$, we now proceed to further restrict the data of $\mathcal{H}_{K,s}$ using the set $\mathcal{A}_0$. In particular, we consider a HDS $\mathcal{H}_{K,s,0} := \{C_{A,K,s,0}, F_1^{\text{HARDD}}, D_{A,K,s,0}, G_1^{\text{HARDD}}\}$ with flow and jump sets

$$C_{A,K,s,0} := (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \big([T_0, T]^n \cap \mathcal{A}_\tau\big),$$
$$D_{A,K,s,0} := (\ker(\mathbf{L})^\perp \cap K) \times (\ker(\mathbf{L})^\perp \cap K) \times \big(D_\tau \cap \mathcal{A}_\tau\big).$$

For this restricted HDS, the following two lemmas establish UGAS of the set $\mathcal{A}$, and suitable acceleration properties.

Lemma B.7. Let $\alpha = 0$. Then, the HDS $\mathcal{H}_{K,s,0}$ renders UGAS the set $\mathcal{A}$, and the dual function is minimized at a rate of $\mathcal{O}(1/\tau^2)$ during flows.

Proof. Consider the Lyapunov function

$$V(x) := \frac{1}{4}|q-p|^2 + \frac{1}{4}|p|^2_{\mathcal{A}_\phi} + \gamma \frac{\tau^\top \tau}{N}(\phi(q) - \phi^*) \tag{B.47}$$

where $\phi^* = \phi(\mathcal{A}_\phi)$. This is a modified version of the "centralized" Lyapunov function considered in [24]. By construction, this function is positive definite with respect to the compact set $\mathcal{A}$ in $C_{A,K,s,0} \cup D_{A,K,s,0}$, and also radially unbounded due to Assumption 4.4. Since during flows $\tau(t) = \tau_s(t)\mathbf{1}_N$, with $\dot{\tau}_s(t) = 0.5$, the gradient of V satisfies

$$\nabla V(x) = \begin{pmatrix} \frac{1}{2}(q-p) + \gamma \frac{\tau^\top \tau}{N} \nabla \phi(q) \\ \frac{1}{2}(p-q) + \frac{1}{2}(p-p^*) \\ \gamma \frac{2\tau}{N}(\phi(q) - \phi^*) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}(q-p) + \gamma\tau_s^2 \nabla\phi(q) \\ y - \frac{1}{2}(q+p^*) \\ 2\gamma\frac{\tau_s}{N}(\phi(q)-\phi^*)\mathbf{1}_N \end{pmatrix} \tag{B.48}$$

where p^* is the Euclidean projection of p in $\mathcal{A}_\phi$. Therefore, the derivative of V along the trajectories of the system satisfies

$$\begin{aligned} \dot{V}(x) &= -\frac{|p-q|^2}{\tau_s} + 2\gamma\tau_s(p-q)^\top\nabla\phi(q) - 2\gamma\tau_s y^\top\nabla\phi(q) \\ &\qquad + \gamma\tau_s(q+p^*)^\top\nabla\phi(q) + \gamma\tau_s(\phi(q)-\phi^*) \\ &= -\frac{|p-q|^2}{\tau_s} - \gamma\tau_s\left((q-p^*)^\top\nabla\phi(q) - (\phi(q)-\phi^*)\right). \end{aligned} \tag{B.49}$$

Since, by assumption ϕ is convex, we have that $(q-p^*)^\top\nabla\phi(q) - (\phi(q)-\phi^*) \geq 0$. Moreover, due to the fact that $\tau_s \in [T_0, T]$, and by Assumption 4.4 from (B.49) we obtain that $\dot{V} < 0$ for all $p \in C_{A,K,s,0}\backslash\mathcal{A}$. On the other hand, jumps occur whenever $\tau_1 = \tau_2 = \ldots = \tau_n = T$. This condition will trigger N consecutive jumps, after which the system will flow again. Thus, we evaluate the Lyapunov function after the N jumps and obtain:

$$V(q^{+^n}) = \frac{1}{4}|q-p|^2 + \frac{1}{4}|p|^2_{\mathcal{A}_\phi} + \gamma T_0^2(\phi(q)-\phi^*).$$

Therefore, the difference $\Delta(p^{+^n})V := V(p^{+^n}) - V(x)$ satisfies

$$\begin{aligned} \Delta V(q^{+^n}) &= \frac{1}{4}|q-p|^2 + \frac{1}{4}|p|^2_{\mathcal{A}_\phi} + \gamma T_0^2(\phi(q)-\phi^*) \\ &\qquad - \frac{1}{4}|q-p|^2 - \frac{1}{4}|p|^2_{\mathcal{A}_\phi} - \gamma(T)^2(\phi(q)-\phi^*) \\ &= -\gamma(\phi(q)-\phi^*)\left((T-T_0)^2 + 2T_0(T-T_0)\right) \\ &\leq 0, \end{aligned}$$

for all $x \in D_{A,K,s,0}$. Therefore, the Lyapunov function V has strict decrease during flows and does not increase during jumps. Since at most n jumps can happen between each interval of flow of duration $2(T - T_0)$, UGAS follows directly by the invariance principle [33, Thm. 8.8].

To obtain the convergence bound for the dual function, note that $\dot{V} \leq 0$ implies $V(x(t,j)) \leq V(x(s,j))$ for all $(t,j),(s,j) \in \mathrm{dom}(x)$ with $t \geq s$. In turn, this inequality implies that

$$\phi(q(t,j)) - \phi^* \leq \frac{N \cdot V(s_j, j)}{\gamma \tau^\top \tau} = \frac{c_j}{\tau_i^2}, \tag{B.50}$$

during flows, where $s_j = \inf\{t \geq 0 : (t,j) \in \mathrm{dom}(p)\}$, and $c_j := V(s_j, j)/\gamma$. ■

Lemma B.8. Let $\alpha = 1$. Then, the HDS $\mathcal{H}_{K,s,0}$ renders uniformly globally exponentially stable (UGES) the set $\mathcal{A}$, and ϕ is minimized at an exponential rate.

Proof. By Lemma 4.5, the function ϕ is μ_ϕ-strongly convex in $\ker(\mathbf{L})^\perp$, and has a globally ℓ_ϕ-Lipschitz gradient. Thus, $\mathcal{A}_\phi = \{q^*\}$. We consider again the same Lyapunov function $V(x)$, which now satisfies

$$\underline{c}|x|_{\mathcal{A}}^2 \leq V(x) \leq \overline{c}|x|_{\mathcal{A}}^2$$

with $\underline{c} := 0.25 \min\{1, 2\gamma {T_0}^2 \mu_\phi\}$ and $\overline{c} := 0.25 \max\{3, 6\gamma (T)^2 \ell_\phi\}$. Since the continuous-time dynamics are the same of Lemma B.7, we still have $\dot{V} \leq 0$. However, using strong convexity we have

$$(\phi(q) - \phi^*) - (q - q^*)^\top \nabla \phi(q) \leq -\frac{\mu_\phi}{2}|q - q^*|^2$$

to further obtain

$$\begin{aligned}
\dot{V}(x) &\leq -\frac{|p-q|^2}{\alpha} - \gamma\mu_\phi\frac{\alpha}{2}|q-q^*|^2 \\
&\leq -\frac{|p-q|^2}{T_0+T-T_0} - \gamma\mu_\phi\frac{T_0}{2}|q-q^*|^2 \\
&\leq \min\left\{\frac{1}{T_0+T-T_0},\ \frac{\gamma T_0\mu_\phi}{4}\right\}\left(-|p-q|^2-|q-q^*|^2\right) - \gamma\mu_\phi\frac{T_0}{4}|q-q^*| \\
&\leq -\frac{1}{2}\min\left\{\frac{1}{T_0+T-T_0},\ \frac{\gamma T_0\mu_\phi}{4}\right\}\left(|p-q^*|^2+|q-q^*|^2\right) \\
&\leq -\frac{\rho}{\bar{c}}V(x)
\end{aligned}$$

for all $x \in C_{A,K,s,0}$, where $\rho := 0.5\min\{\frac{1}{T}, 0.25\gamma T_0\mu_\phi\}$. On the other hand, after the n consecutive jumps triggered by the condition $\tau_1 = \tau_2 = \ldots = \tau_n = T$, the change in the Lyapunov function is

$$\begin{aligned}
\Delta V(x^{+^n}) = \frac{1}{4}\left|q^{+^n} - p^{+^n}\right|^2 + \frac{1}{4}\left|p^{+^n} - q^*\right|^2 + \gamma T_0^2\left(\phi\left(q^{+^n}\right) - \phi^*\right) \\
- \frac{1}{4}|q-p|^2 - \frac{1}{4}|p-q^*|^2 - \gamma(T)^2(\phi(q)-\phi^*) \\
= \gamma {T_0}^2(\phi(p)-\phi^*) - \frac{1}{4}|q-p|^2 - \gamma(T)^2(\phi(q)-\phi^*).
\end{aligned}$$

Since ϕ is strongly convex in $\ker(\mathbf{L})^\perp$, and $\nabla\phi$ is globally Lipschitz, we have that

$$\phi(p) - \phi^* \leq \frac{1}{2}\ell_\phi|p-q^*|^2,\ \forall\ p \in \ker(\mathbf{L})^\perp, \tag{B.51}$$

and

$$\phi(q) - \phi^* \geq \frac{1}{2}\mu_\phi|q-q^*|^2,\ \forall\ q \in \ker(\mathbf{L})^\perp. \tag{B.52}$$

Using these two inequalities to further upper bound $\Delta V(p^{+^n})$, we obtain:

$$\Delta V(x^{+^n}) \leq \gamma\ell_\phi\frac{T_0^2}{2}|p-q^*|^2 - \frac{1}{4}|q-p|^2 - \gamma\mu_\phi\frac{(T)^2}{2}|q-q^*|^2$$

$$\leq -T_1|q-p|^2 - T_2|q-q^*|^2$$
$$\leq -\underline{T}(|q-p|^2 + |q-q^*|^2),$$

where $T_1 := \frac{1}{4} - \gamma\ell_\phi T_0^2$, $T_2 := \frac{1}{2}\gamma\mu_\phi(T)^2 - \gamma\ell_\phi T_0^2$ and $\underline{T} = \min\{T_1, T_2\}$. The constants T_1 and T_2 are positive provided the following holds: $0 < T_0 < 1/(2\sqrt{\gamma\ell_\phi})$ and $(T)/T_0 > \sqrt{2\kappa_\phi}$, which are precisely conditions **(C.1)** and **(C.2)**.

Finally, since V can be upper bounded as

$$\begin{aligned} V(x) &\leq \overline{c}\left(|q-q^*|^2 + |p-q^*|^2\right) \\ &\leq \overline{c}\left(|q-q^*|^2 + 2|p-q|^2 + 2|q-q^*|^2\right) \\ &\leq 3\overline{c}\left(|q-q^*|^2 + |p-q|^2\right), \end{aligned}$$

we obtain

$$\Delta V(x^{+^n}) \leq -\beta V(x)$$

where $\beta := \underline{T}/3\overline{c}$. These bounds, and the fact that the system is uniformly non-Zeno with N consecutive jumps followed by a constant interval of flow, establish UGES of $\mathcal{A}$ via [223, Thm. 1]. ■

Step 6: Nested Application of the Reduction Principle

We now repeatedly apply the Hybrid Reduction Principle [33, Cor. 7.24] to establish UGAS for the original hybrid system $\mathcal{H}_A$. First, since the set $\mathcal{A}$ is UGAS for the HDS $\mathcal{H}_{K,s,0}$, and the set $\mathcal{A}_0$ is UGAS for the HDS $\mathcal{H}_{K,s}$, by the reduction principle, we obtain that the set $\mathcal{A}$ is UGAS for the HDS $\mathcal{H}_{K,s}$. Moreover, since the compact set $A_{K,s}$ is UGAS for the HDS $\mathcal{H}_K$, it follows again by the reduction principle that $\mathcal{H}_K$ renders UGAS the set $\mathcal{A}$. Finally, since by Step 1, there are no finite escape times in the original HDS $\mathcal{H}_A$, and since the compact set K was arbitrary, we have that the set $\mathcal{A}$ is indeed UGAS for

the HDS $\mathcal{H}_A$ with no restriction. This establishes the stability result of property **(P.3)**. See Figure B.2 for an illustration of the nested application of the reduction principle to guarantee asympotic stability of the set $\mathcal{A}$.

Step 7: Optimal Bounds for the Primal. Finally, we derive convergence bounds for the primal problem (4.41) based on the convergence bounds derived for the dual problem in Steps 1-6.

Let $\alpha = 1$. Since the gradient of ϕ is globally Lipschitz, we have

$$|\nabla\phi(q)|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(q) - \phi^*) \tag{B.53}$$

Using (4.44),

$$|\mathbf{L}z|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(q) - \phi^*) \tag{B.54}$$

Let us decompose $z = \bar{z} + \tilde{z}$, where $\bar{z} \in \ker(\mathbf{L})$ and $\tilde{z} \in \ker(\mathbf{L})^\perp$. We then have

$$|\mathbf{L}\tilde{z}|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}}(\phi(q) - \phi^*) \tag{B.55}$$

and since $\tilde{z} \in \ker(\mathbf{L})^\perp$, which implies that $|\mathbf{L}\tilde{z}|^2 \geq \lambda^+_{\min}(\mathcal{L}^2)|\tilde{z}|^2 s$, we get

$$|\tilde{z}|^2 \leq \frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}(\phi(q) - \phi^*). \tag{B.56}$$

Using (4.44) and the definition of $z = h(\cdot)$ in we get:

$$\langle \mathbf{L}q, z\rangle - F(z) \geq \langle \mathbf{L}q, z^*\rangle - F(z^*), \tag{B.57}$$

where z^* is the minimizer of the primal problem. Now, using the fact that $\mathbf{L}z^* = 0$ we

obtain:

$$\begin{aligned} F(z) &\leq F(z^*) + \langle \mathbf{L}q,\, z - z^* \rangle \\ &= F(z^*) + \langle q,\, \mathbf{L}\tilde{z} \rangle \\ &= F(z^*) + \langle \mathbf{L}q,\, \tilde{z} \rangle \\ &\leq F(z^*) + |\langle \nabla F(z), \tilde{z} \rangle|, \end{aligned}$$

where in the last step we have used that $\mathbf{L}q = \nabla F(z)$ by KKT conditions. Moreover, by the Cauchy-Schwartz inequality we get:

$$F(z) - F(z^*) \leq |\nabla F(z)||\tilde{z}|. \tag{B.58}$$

Combining (B.56) and (B.58), we get

$$F(z) - F(z^*) \leq |\nabla F(z)| \sqrt{\frac{\lambda_{\max}(\mathcal{L}^2)}{\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}(\phi(q) - \phi^*)}. \tag{B.59}$$

By UGAS, for each compact set of initial condition K_0 there exists $M > 0$ such that $|\nabla F(z)| < M$. By exponential stability [223] and property (B.51) we obtain:

$$\phi(q) - \phi^* \leq \frac{1}{2}\ell_\phi c^2 |q(0,0) - q^*|^2 \exp(-2\lambda(t+j)), \tag{B.60}$$

where $c^2 \coloneqq \frac{\overline{c}}{\underline{c}}$ and $\lambda = \frac{\rho}{c}$. Combining this with (B.59) we finally obtain for all initial conditions $x(0,0) \in K_0$ that:

$$F(z) - F(z^*) \leq cM \sqrt{\frac{\lambda_{\max}(\mathcal{L}^2)}{2\overline{\mu}\lambda^+_{\min}(\mathcal{L}^2)}} |q(0,0) - q^*| \exp(-\lambda(t+j)). \tag{B.61}$$

When $\alpha = 0$ we can follow exactly the same steps, using the bound (B.50) instead of using (B.51). ■

Appendix C

Proofs of Chapter 5

C.1 Proofs of Section 5.1

Similar to the proof of Theorem 4.1 in Appendix B.1 we divide the proof in different lemmas.

Lemma C.1. Consider the HDS $\mathcal{H}_s$ under the Assumptions of Theorem 5.1. Then, the set $\mathcal{A}$ is UGAS.

Proof. First, by Assumption 5.3 and strict convexity of ϕ_i on q_i for all $i \in \mathcal{V}$, which follows by strict monotonicity of the pseudo-gradient, existence of the NE is guaranteed via [56, Cor 4.2]. With this at hand, we consider the Lyapunov function $\tilde{V} = V_1 + V_2 + \tilde{V}_3$, where V_1 and V_2 are defined in (B.4), and $\tilde{V}_3$ is given by:

$$\tilde{V}_3(x) \coloneqq c_o \frac{|\tau|^2 |\mathcal{G}(q)|^2}{2n}, \tag{C.1}$$

where c_o corresponds to the cocoercivity constant of $\mathcal{G}$. By construction and Assumption 5.2, we have that V is radially unbounded, and also positive definite with respect to $\mathcal{A} \cap (C_s \cup D_s)$. Using c_o-cocoercivity of $\mathcal{G}$, inequality (B.6) now becomes

$$\dot{\tilde{V}}(x) \leq -\frac{1}{\tau_s}|p-q|^2 - 2\tau_s (p-q)^\top \mathcal{G}(q)$$

$$+ 2c_o\tau_s\mathcal{G}^\top(q)\frac{\partial\mathcal{G}}{\partial q}(q)(p-q) + c_o(1-\eta)\tau_s|\mathcal{G}(q)|^2,$$

which can be written in quadratic form as

$$\dot{\tilde{V}}(x) \leq -\tau_s\tilde{x}^\top M_{1/c_o}(q,\tau_s)\tilde{x}, \tag{C.2}$$

with $\tilde{x} \coloneqq \big((p-q),\mathcal{G}(q)\big)$, and

$$M_{1/c_o}(q,\tau_s) \coloneqq \begin{pmatrix} \frac{1}{\tau_s^2}I_n & I_n - c_o\partial\mathcal{G}(q)^\top \\ I_n - c_o\partial\mathcal{G}(q) & c_o(1-\eta)I_n \end{pmatrix}. \tag{C.3}$$

Since $\eta \leq \frac{1}{2}$ by design, $c_o = 1/\ell$, and $\tau_s \in [T_0, T]$, under the conditions of Theorem 5.1, we have that $M_\ell(q,\tau_s) \succ 0$ for all $\tau_s \in [T_0,T]$ and $q \neq q^*$ whenever

$$0 \prec I_n - \frac{T^2}{\ell(1-\eta)}\big(\ell I_n - \partial\mathcal{G}(q)^\top\big)\big(\ell I_n - \partial\mathcal{G}(q)\big). \tag{C.4}$$

The expression in (C.4) is precisely (5.7) with $\rho_F = \ell$ and $\delta = 0$. Thus, since by assumption $\mathcal{S}_0$ is ℓ-GC, it follows that (C.4) holds. Finally, note that when $q = q^*$ inequality (C.2) reduces to $\dot{\tilde{V}}(x) \leq -\frac{1}{\tau_s}|p-q|^2 \leq 0$.

On the other hand, after the n consecutive jumps that proceed each interval of flow, the change of V is

$$\Delta_j^{j+n}\tilde{V}(z) = \frac{c_o}{2}|\mathcal{G}(q)|^2\left(T_0^2 - T^2\right) \leq 0. \tag{C.5}$$

Finally, we show that no complete solution x of $\mathcal{H}_s$ keeps $\tilde{V}$ in a non-zero level set. In particular, since for all $(q,p,\tau) \in \mathbb{R}^n\backslash\{q^*\} \times \mathbb{R}^n \times [T_0,T]$ we have that $\dot{V} < 0$, it suffices to consider the case $q = q^*$, which leads to $\dot{V} = 0$ only when $p = q$, i.e., when $(p,q) \in \mathcal{A}$. Therefore, no solution that flows can keep $\tilde{V}$ constant in a non-zero level set. Since the flows are periodic UGAS of $\mathcal{A}$ follows now by the hybrid invariance

principle [16, Thm. 8.8]. ■

Proof of Theorem 5.1

(a) Stability Properties: Follows by the same ideas used in the proof of the stability properties of Theorem 4.1-(i_1), but using Lemma C.1 instead of Lemma B.2.

(b) Convergence Bounds: Follows by the same steps used in the proof of Theorem 1-(i_1), substituting (B.24) by

$$|\mathcal{G}(q)|^2 \leq \frac{2\ell n}{\tau^\top \tau}\tilde{V}_3(s_j, j) = \frac{\tilde{c}_j}{\tau_s^2}. \tag{C.6}$$

where $\tilde{c}_j \coloneqq 2\ell\tilde{V}_3(s_j, j)$. ■

Lemma C.2. Consider the HDS $\mathcal{H}_s$ under the Assumptions of Theorem 5.2-(i_4). Then, the set $\mathcal{A}$ is UGES.

Proof. We consider the Lyapunov function V used in the proof of Lemma C.1, with $\tilde{V}_3$ given by (C.1) and $c_o = \kappa/\ell^2$. The time derivative of $\tilde{V}$ now satisfies

$$\dot{\tilde{V}}(x) \leq -\tau_s \tilde{x}^\top M_{\sigma_\phi \ell}(q, \tau_s)\tilde{x},$$

with $\tilde{x} \coloneqq \big((p-q), \mathcal{G}(q)\big)$. By assumption we know that $\mathcal{S}_\delta$ is $(\sigma_\phi\ell)$-GC, which is equivalent to:

$$0 \prec I_n - \left(\frac{T^2}{1-T^2\delta}\right)\frac{\left(\sigma_\phi \ell I_n - \partial\mathcal{G}(q)^\top\right)\left(\sigma_\phi \ell I_n - \partial\mathcal{G}(q)\right)}{\sigma_\phi\ell(1-\eta) - \sigma_\phi^2\ell^2\delta}.$$

In turn, when $0 < \delta < (1-\eta)/\sigma_\phi\ell$ and $0 < \eta \leq 1/2$, the above inequality directly implies that $M_{\sigma_\phi\ell}(q, \tau_s) \succ \delta I_n$, for all $\tau_s \in [T_0, T]$ and all $q \neq q^*$. Thus, for such points, and during flows, we have $\dot{\tilde{V}} \leq -\delta(|p-q|^2 + |\mathcal{G}|(q))$. Using κ-strong-monotonicity

and κ/ℓ^2-cocoercivity of $\mathcal{G}$ we conclude

$$\dot{\tilde{V}}(x) \leq -\lambda \tilde{V}(x), \text{ with } \lambda = \frac{4\delta}{\max\left\{3, 2(\frac{1}{\kappa^2} + \frac{\kappa}{\ell^2}T^2)\right\}}. \tag{C.7}$$

On the other hand, during jumps, using (RC_1), the definition of $\tilde{V}_3$, and the Reset Policy $\alpha \in \{0,1\}^n$, the change of $\tilde{V}$ is

$$\Delta_j^{j+n}\tilde{V} \leq -\frac{1}{4}\sum_{i\in\Theta}\left[(p_i - q_i)^2 + (p_i - q_i^*)^2\right] - \gamma(\sigma_\phi^2\kappa^{-1})\tilde{V}_3(x), \tag{C.8}$$

where $\gamma(\sigma_\phi^2\kappa^{-1}) \in (0,1)$ is given by (5.6), and Θ is defined in the proof of Lemma B.3. Thus, it follows that $\Delta_j^{j+n}\tilde{V} \leq 0$. Moreover, by the κ-strong monotonicity and ℓ-Lipschitz continuity of $\mathcal{G}$, $\tilde{V}$ satisfies the quadratic bounds $\underline{c}|x|_{\mathcal{A}}^2 \leq \tilde{V}(x) \leq \overline{c}|x|_{\mathcal{A}}^2$, where:

$$\underline{c} \coloneqq \min\left\{\frac{1}{4},\ \frac{\kappa T_0^2}{2\sigma_a^2}\right\},\ \overline{c} \coloneqq \max\left\{\frac{3}{4},\ \frac{1}{2} + \frac{\kappa T^2 \ell^2}{2}\right\}. \tag{C.9}$$

The exponential decrease of V during the flows (which are periodic), the non-increase of V during the jumps, and the quadratic upper and lower bounds of $\tilde{V}$, imply that $\mathcal{H}_s$ renders UGES the set $\mathcal{A}$. ■

Lemma C.3. Consider the HDS $\mathcal{H}_s$ under the Assumptions of Theorem 5.2-(i_5). Then, the set $\mathcal{A}$ is UGES.

Proof. Consider the Lyapunov function $\tilde{V}$ used in the proof of Lemma C.1, which still satisfies (C.7). During jumps, the reset policy $\alpha = \mathbf{0}_n$ implies that $\Theta = \mathcal{V}$ in (C.8), leading to $\Delta_j^{j+n}\tilde{V}(x) \leq -V_1(x) - V_2(x) - \gamma(\sigma_\phi^2\kappa^{-1})V_3(x) \leq -\gamma(\sigma_\phi^2\kappa^{-1})\tilde{V}(x)$. Therefore, by [223, Thm. 1], and the quadratic upper and lower bounds of $\tilde{V}$, system $\mathcal{H}_s$ renders UGES the set $\mathcal{A}$. ■

Proof of Theorem 5.2

(a) Stability Properties: Follows by using using Lemmas C.2 and C.3 in conjunction with the same ideas used in the proof of Theorem 1.

(b) Convergence Bounds: We follow the same steps of the proof of Theorem 1, using now $\tilde{V}_3$ instead of V_3. For item (i_4), this leads to the following bound instead of (B.28):

$$|x(t,j)|_{\mathcal{A}} \leq \hat{c}|x(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{3n}(t+j)},$$

where λ are defined in (C.7), $\hat{c} \coloneqq \sqrt{\overline{c}/\underline{c}} \cdot e^{\left(\frac{5}{6}\lambda+\tilde{l}\right)L}$, and $\underline{c}$ and $\overline{c}$ are given in (C.9). Finally, for item (i_5), we obtain the following bound instead of (B.30):

$$|q(t, j_s + kn) - q^*| \leq \sigma_r \sigma_\phi \left(1 - \gamma \left(\sigma_\phi^2 \kappa^{-1}\right)\right)^{\frac{k}{2}} |q(t_s, j_s) - q^*|,$$

from here, the proof follows the exact same steps. ■

Proof of Lemmas 5.1 and 5.2

We first show Lemma 5.2. Using $c_o = \kappa/\ell^2$ we have that (RC_3) can be equivalently written as

$$\tilde{\alpha} > 1 - 2c_o\kappa + c_o^2\ell^2, \tag{C.10}$$

with $\tilde{\alpha} \coloneqq \left(\frac{1}{T^2} - \delta\right)(c_o(1-\eta) - \delta)$. By using the fact that $\mathcal{G}$ is ℓ-Lipschitz continuous, it follows that $\partial\mathcal{G}(q)^\top \partial\mathcal{G}(q) \prec \ell^2 I_n$ [123], where we use $\partial\mathcal{G}$ in place of $\frac{\partial\mathcal{G}}{\partial q}$ to simplify notation. Using this fact, together with monotonicity properties of the pseudogradient, which implies that $\partial\mathcal{G}(q) + \partial\mathcal{G}(q)^\top \succ 2\kappa I_n$ [118, Prop 2.3.2 c)], from (C.10) we obtain

$$0 \prec (\alpha - 1)I + c_o\left(\partial\mathcal{G}(q) + \partial\mathcal{G}(q)^\top\right) - c_o^2 \partial\mathcal{G}(q)^\top \partial\mathcal{G}(q).$$

and hence that

$$0 \prec I - \left(\frac{T^2}{1 - T^2\delta}\right) \frac{\left(I_n - c_o \partial \mathcal{G}(q)^\top\right)\left(I_n - c_o \partial \mathcal{G}(q)\right)}{c_o(1-\eta) - \delta},$$

which implies, whenever $0 \leq \delta < c_0(1-\eta)$, that $\mathcal{S}_\delta$ is $(1/c_o)$-GC. Lemma 2 follows by the same arguments, using $c_o = 1/\ell$ and letting $\kappa \to 0^+$. ■

Proof of Lemma 5.3

To satisfy $(\mathrm{RC})_1$ and $(\mathrm{RC})_3$ we need

$$T_0^2 + \frac{\ell^2}{2\kappa^3} < \frac{\kappa(1-\eta) - \delta\ell^2}{\ell^2 - \kappa^2 + \delta(\kappa(1-\eta) - \delta\ell^2)}.$$

Since $T_0^2 > 0$, it is necessary that

$$\frac{\ell^2}{2\kappa^3} < \frac{\kappa(1-\eta) - \delta\ell^2}{\ell^2 - \kappa^2 + \delta(\kappa(1-\eta) - \delta\ell^2)},$$

which in turn is equivalent to

$$1 < 2\frac{(1-\eta)}{\sigma_\phi^4} + \frac{1}{\sigma_\phi^2}\left(1 - \delta\left(2\kappa + \frac{1-\eta}{\kappa} - \delta\frac{\ell^2}{\kappa^2}\right)\right). \tag{C.11}$$

Since by assumption $\sigma_\phi^4 - \sigma_\phi^2 < 2(1-\eta)$, there exists $\delta > 0$ sufficiently small such that (C.11) holds. ■

Proof of Lemma 5.4:

The convergence bound of Theorem 5.2-(i_5) implies the slightly looser bound $\forall\ (t,j) \in \mathrm{dom}(x)$:

$$|q(t,j) - q^*| \leq \sigma_r \sigma_\phi \left(1 - \gamma\left(\rho_J\right)\right)^{\frac{1}{2}\left\lfloor \frac{j-n}{n} \right\rfloor} M_0. \tag{C.12}$$

Since $(0,0) \in \mathrm{dom}\,(x)$, by using the structure of the hybrid time domain, with $(i,s) = (0,0)$, we obtain

$$\left\lfloor \frac{j-n}{n} \right\rfloor \leq \frac{t}{L},$$

where $L = (T - T_0)/\eta$. Hence, using (C.12), $|q - q^*|$ satisfies

$$|q(t,j) - q^*| \leq \sigma_\phi \sigma_r \left(1 - \gamma\left(\sigma_\phi^2 \kappa^{-1}\right)\right)^{\frac{\eta}{2}\frac{t}{T-T_0}} M_0. \tag{C.13}$$

By minimizing the right hand side of (C.13) with respect to the restarting parameter T, we find that the minimum value is achieved when $(1 - \gamma(\sigma_\phi \kappa^{-1})) = \frac{1}{e^2}$. Solving for T, we obtain precisely T^{opt}. Using this restarting parameter, for any $\nu > 0$ the error $|q(t,j) - q^*| \leq \nu$ is obtained when $\sigma_\phi \sigma_r e^{-\eta \frac{t}{T^{\mathrm{opt}} - T_0}} M_0 \leq \nu$. This inequality holds precisely for all $t \geq t_\nu^{\mathrm{opt}}$. Finally, using $(1 - \gamma(\sigma_\phi \kappa^{-1})) = \frac{1}{e^2}$ in (C.13):

$$|q(t,j) - q^*| \leq \sigma_\phi \sigma_r e^{-\eta \frac{t}{T^{\mathrm{opt}} - T_0}} M_0. \tag{C.14}$$

As $T_0 \to 0^+$, and using the value of T^{opt}:

$$|q(t,j) - q^*| \leq \sigma_\phi \sigma_r e^{-t \frac{\eta\sqrt{2\kappa}}{e\sigma_\phi}} M_0, \tag{C.15}$$

which gives the convergence bound of order $\mathcal{O}(e^{-\sqrt{\kappa}/\sigma_\phi})$.

Proof of Proposition 5.1

We consider the Lyapunov function $\hat{V} = V_1 + V_2 + \hat{V}_3$, where V_1 and V_2 are given by (B.4a), and $\hat{V}_3$ is defined as $\hat{V}_3 = \kappa \frac{|\tau|^2}{2n} |q - q^*|^2$. During flows we now have

$\dot{\hat{V}}(x) \leq -\tau_s \hat{x}^\top M_Q(\tau_s)\hat{x}$, where $\hat{x} = (p-q, q-q^*)$ with M_Q given by

$$M_Q(\tau_s) = \begin{pmatrix} \frac{1}{\tau_s^2} I_n & (A - \kappa I_n) \\ \left(A^\top - \kappa I_n\right) & \kappa(1-\eta) I_n \end{pmatrix}. \tag{C.16}$$

This matrix is positive definite if and only if the following matrix inequality holds for all $\tau_s \in [T_0, T]$:

$$0 \prec \frac{\kappa(1-\eta)}{\tau_s^2} I_n - (\kappa I_n - A)\left(\kappa I_n - A^\top\right). \tag{C.17}$$

Hence, it suffices to verify the condition $0 \prec I_n - \frac{T^2}{\kappa(1-\eta)}(\kappa I_n - A)\left(\kappa I_n - A^\top\right)$, which is equivalent to κ-GC of $\mathcal{S}_0$. It follows that $M_Q(\tau_s) \succ 0$ for all $\tau \in [T_0, T]$, and $\dot{\hat{V}} \leq -\tau_s c|\hat{z}|^2$ during flows, for some $c > 0$. On the other hand, during jumps, the policy $\alpha = \mathbf{0}_n$ leads to:

$$\Delta_j^{j+n} V(z) \leq -V_1(z) - V_2(z)\left(\frac{\kappa}{2}(\tau_s^2 - T_0^2) - \frac{1}{4\kappa^2}\right)|q - q^*|^2.$$

Using (RC_2), we obtain $\Delta_j^{j+n} V(z) \leq -V_1(z) - V_2(z) - \gamma(\kappa^{-1})V_3(z)$, with γ as in (5.6). UGES of $\mathcal{A}$ follows by the same arguments of the proof of Theorem 3-(i_5). ■

Proofs of Theorem 5.3

To prove Theorem 5.3, we present two auxiliary lemmas:

Lemma C.4. Consider the assumptions of Theorem 5.3, and let $\mathcal{H}_{2,s} = \{C_{2,s}, F_{2,s}, D_{2,s}, G_{2,s}\}$ be obtained by intersecting the data of $\mathcal{H}_2$ with $\mathcal{A}_{2,\nu} \coloneqq \mathcal{A}_\nu \times (\mathcal{Q}(\mathbf{1}_n \otimes q^*) + \nu\mathbb{B})$, where $\mathcal{A}_\nu = (\{(q^*, q^*)\} + \nu\mathbb{B}) \times \mathcal{A}_{\text{sync}}$. Then $\mathcal{H}_{2,s}$ renders UGAS the set $\mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*)\}$.

Proof. Consider the change of variable $\theta = \hat{q} - h(q)$, with $h(q) \coloneqq \mathcal{Q}(\mathbf{1}_n \otimes q)$ and let

$$W(q, \theta, \varepsilon) \coloneqq -\mathcal{Q}\mathbf{L}\mathcal{Q}^\top\theta - \varepsilon\mathcal{Q}\left(\mathbf{1}_n \otimes 2D(\tau)^{-1}(p-q)\right)$$

$$+ \varepsilon \mathcal{Q}\left(\mathbf{1}_n \otimes \mathcal{P}\mathbf{L}\mathcal{Q}^\top \theta\right). \tag{C.18}$$

This change of coordinates leads to a HDS $\mathcal{H}_\vartheta$ with state $\vartheta \coloneqq (x, \theta)$, where $x = (q, p, \tau)$, and data $\mathcal{H}_\vartheta = (C_{2,\vartheta}, F_{2,\vartheta}, D_{2,\vartheta}, G_{2,\vartheta})$, where $C_{2,\vartheta}, D_{2,\vartheta}$ and $G_{2,\vartheta}$ are obtained directly from (5.16), (5.17), and (5.18) respectively via the change of coordinates, and where the flow map is defined by $F_{2,\vartheta}(\vartheta) \coloneqq (U(x, \theta + h(q)),\ W(q, \theta, \varepsilon)/\varepsilon)$ where:

$$U(x, \theta + h(q)) = \begin{pmatrix} 2D(\tau)^{-1}(p - q) - \mathcal{P}\mathbf{L}\mathcal{Q}^\top \theta \\ -2D(\tau)\hat{\mathcal{G}}(\mathbf{1}_n \otimes q + \mathcal{Q}^\top \theta) \\ \eta \mathbf{1}_n \end{pmatrix}. \tag{C.19}$$

Let $\mathcal{H}_{\vartheta,s}$ be the HDS that results from intersecting the data of $\mathcal{H}_\vartheta$ with $\mathcal{A}_\nu \times (\nu\mathbb{B})$, with $\nu > 0$. Note that studying the stability of $\mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*\}$ under $\mathcal{H}_{2,s}$, is is equivalent to analyzing the stability properties of the compact set $\mathcal{A}_{\mathbb{G},\theta} = \mathcal{A} \times \{0\}^{n^2-n}$ under $\mathcal{H}_{\vartheta,s}$. For this last system, we consider the Lyapunov function

$$V_{\mathbb{G}}(\vartheta) = (1 - d)\tilde{V}(x) + d \cdot V_\theta(\theta), \text{ with } d \in (0, 1), \tag{C.20}$$

where $\tilde{V}$ is defined as in Lemma C.1, and $V_\theta(\theta) \coloneqq \frac{1}{2}|\theta|^2$. By using the proof of Lemma C.1, and noting that $\hat{\mathcal{G}}(\mathbf{1} \otimes q) = \mathcal{G}(q)$, we obtain:

$$\frac{\partial \tilde{V}(x)}{\partial x} U(x, h(q)) \leq -\tau_s \tilde{x}^\top M_\ell(q, \tau_s)\tilde{x}, \tag{C.21}$$

with $\tilde{x} \coloneqq \big((p - q), \mathcal{G}(q)\big)$ and M_ℓ given by (C.3) with $c_o = \frac{1}{\ell}$. Under the assumptions of Theorem 5.3 we know that

$$0 \prec I_n - \left(\frac{T^2}{1 - T^2\delta}\right) \frac{\left(\ell I_n - \partial \mathcal{G}(q)^\top\right)\left(\ell I_n - \partial \mathcal{G}(q)\right)}{\ell(1 - \eta) - \ell^2\delta},$$

and thus that $M_\ell(q,\tau_s) \succ \delta I_n \ \forall \tau_s \in [T_0, T]$. Hence, from (C.21) and letting $\xi(x) \coloneqq \left(|p-q|^2 + |q-q^*|^2\right)^{1/2}$ we obtain that

$$\frac{\partial \tilde{V}(x)}{\partial x}\dot{x} \leq -T_0\delta \min\{1,\zeta\}\,\xi^2(x), \tag{C.22}$$

where we have also used the bound of Assumption 5.4. Additionally, it also follows that

$$\begin{aligned}
&\frac{\partial \tilde{V}}{\partial x}\left(U(x,\theta+h(q)) - U(x)\right) \leq c_1\left(|p-q| + |q-q^*|\right)|\theta|, \\
&c_1 \coloneqq \frac{T^2\lambda_{\max}(\mathcal{L})}{\sqrt{2}}\max\left\{\frac{1}{T^2} + \frac{4\ell}{T\lambda_{\max}(\mathcal{L})}, 2 + \frac{2\ell}{T\lambda_{\max}(\mathcal{L})}\right\}.
\end{aligned} \tag{C.23}$$

On the other hand, by the fact that the underlying communication graph is undirected and connected, it follows that $\mathcal{Q}\mathbf{L}\mathcal{Q}^\top$ is positive definite [107, Lemma 6], and, moreover

$$\frac{\partial V_\theta}{\partial \theta}W(q,\theta,0) \leq -\frac{\lambda_2(\mathcal{L})}{n}|\theta|^2. \tag{C.24}$$

We also have that

$$\left(\frac{\partial V_\theta}{\partial x} - \frac{\partial V_\theta}{\partial \theta}\frac{\partial h}{\partial x}\right)U(x,\theta+h(q)) \leq c_2\psi(x)|\theta| + c_3|\theta|^2, \tag{C.25}$$

where $c_2 \coloneqq 2\sqrt{2n}/T_0$ and $c_3 \coloneqq 2\sqrt{n}\lambda_{\max}(\mathcal{L})$. Hence, using (C.22)-(C.25) it follows that the time derivative of $V_{\mathbb{G}}$ satisfies $\dot{V}_{\mathbb{G}} \leq -(\xi(x),\theta)^\top \Lambda_\varepsilon(\xi(x),\theta)$ with

$$\Lambda_\varepsilon \coloneqq \begin{pmatrix} (1-d)T_0\epsilon\min\{1,\zeta^2\} & -\frac{1}{2}(1-d)c_1 - \frac{1}{2}c_2 \\ -\frac{1}{2}(1-d)c_1 - \frac{1}{2}c_2 & d\left(\frac{\lambda_2(\mathcal{L})}{\varepsilon n} - c_3\right) \end{pmatrix},$$

which is positive definite provided that $\varepsilon \in (0, \varepsilon_\delta^*)$ where ε_δ^* is as defined in (5.20).

Note moreover, that if ε satisfies this condition there exists $k_\varepsilon > 0$ such that

$$\dot{V}_{\mathbb{G}} \leq -k_\varepsilon \left(|p-q|^2 + |q-q^*|^2 + |\theta|^2 \right). \tag{C.26}$$

Leveraging the results regarding the change of the Lyapunov function $\tilde{V}$ during jumps presented in the proofs of Lemmas C.1, C.2 and C.3, given that (RC_1) is satisfied with $\rho_J = 0$ by assumption, and since $V_\theta^+(\theta) = V_\theta(\theta)$ for all θ whenever $\vartheta \in D_{2,\vartheta}$, it follows that $\Delta_j^{j+n} V_{\mathbb{G}}(\vartheta) \leq 0$ for any resetting policy $\alpha \in \{0,1\}^n$. This inequality and (C.26) imply that $\mathcal{H}_{\vartheta,s}$ renders the set $\mathcal{A}_{\mathbb{G},\theta}$ UGAS via [16, Prop. 3.27]. The stability results for $\mathcal{H}_{2,s}$ follow directly by the change of cooordinates $\hat{q} = \theta + h(q)$ and the described result for $\mathcal{H}_{\vartheta,s}$. ■

Lemma C.5. Every solution of $\mathcal{H}_2$ is complete.

Proof. Since τ is restricted to a compact set, it suffices to study the behavior of the states $(q,p,\hat{q})$ or equivalently of (q,p,θ). Hence, considering the dynamics in (C.18) and (C.19) it follows that $|\dot{q}| \leq \tilde{\ell}_q \left(|p-q| + |\theta| \right)$, $|\dot{p}| \leq \tilde{\ell}_p \left(|q-q^*| + |\theta| \right)$, and $\left|\dot{\theta}\right| \leq \tilde{\ell}_\theta \left(|p-q| + |\theta| \right)$, where $\tilde{\ell}_q \coloneqq \max\left\{ \frac{2}{T_0}, \lambda_{\max}(\mathcal{L}) \right\}$, $\tilde{\ell}_p \coloneqq 2\ell T\sqrt{N}$ and $\tilde{\ell}_\theta \coloneqq \max\left\{ \lambda_{\max}(\mathcal{L}) \left(\frac{1}{\varepsilon} + \sqrt{N} \right), \frac{\sqrt{N}}{T_0} \right\}$. Using these inequalities we obtain:

$$\frac{d|(q-q^*,\ p-p^*,\ \theta)|}{dt} \leq \left| (\dot{q}, \dot{p}, \dot{\theta}) \right| \leq \hat{\ell} |(q-q^*, p-q^*, \theta)|,$$

with $\hat{\ell} = 2\sqrt{3} \max\left\{ \tilde{\ell}_q, \tilde{\ell}_p, \tilde{\ell}_\theta \right\}$, which by the Gronwall-Bellman inequality implies that the continuous time dynamics of $\mathcal{H}_2$ do not generate finite escape times. Since $G_2(D_2) \subseteq C_2 \cup D_2$, solutions do not stop due to jumps. Therefore, every maximal solution of $\mathcal{H}_2$ is complete. ■

We are now prepared to present the proof of Theorem 5.3.

(a) Let $\mathcal{H}_{2,\nu}$ be defined from $\mathcal{H}_2$ by following the same procedure described in the statement

of Lemma B.1. Since the addition of the state $\hat{q}$ and its associated dynamics do not affect the synchronization dynamics, $\mathcal{H}_{2,\nu}$ renders UGFxS the set $\mathcal{A}_{2,\nu}$, where $\mathcal{A}_{2,\nu}$ is as defined in the previous Lemma. Therefore, by the hybrid reduction principle [16, Cor. 7.24], UGAS of $\mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*)\}$ for system $\mathcal{H}_{2,s}$, established in Lemma C.4, implies that $\mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*)\}$ is UGAS for system $\mathcal{H}_{2,\nu}$. Since the choice of $\nu > 0$ is arbitrary and since solutions of $\mathcal{H}_2$ are complete and bounded, using Lemma C.5 we obtain that the compact set $\mathcal{A} \times \{\mathcal{Q}(\mathbf{1}_n \otimes q^*)\}$ is also UGAS for system $\mathcal{H}_2$.

(b) Let $\nu > 0$, and $K_0 \coloneqq K_x \times K_{\hat{q}} \subset \mathbb{R}^{3n} \times \mathbb{R}^{n^2-n}$ be an arbitrary compact set. Moreover, define $\overline{v} \coloneqq \max_{\vartheta \in K_0} V_{\mathbb{G}}(\vartheta)$, where $V_{\mathbb{G}}$ is as defined in (C.20). Notice that $\overline{v}$ exists since $V_{\mathbb{G}}$ is continuous and K_0 is compact by assumption. It follows that $K_0 \subseteq L_{V_{\mathbb{G}}}(\overline{v})$, where $L_f(c)$ represents the c-sublevel set of the function $f : \mathbb{R}^m \to \mathbb{R}$. Since $V_{\mathbb{G}}$ is radially unbounded by construction and Assumption 5.2, it follows that $L_{V_{\mathbb{G}}}(\overline{v})$ is compact. Let $K^V \coloneqq L_{V_{\mathbb{G}}}(\overline{v})$ and define the HDS $\mathcal{H}_{2,K} = (F_2,\ C_2 \cap K^V,\ G_2,\ D_2 \cap K^V)$. Notice that under $\mathcal{H}_{2,K}$, $\hat{q}$ evolves in a compact set. Moreover, by the arguments presented in the proof of item (a), $\mathcal{H}_{2,K}$ renders K^V strongly forward invariant for any $\varepsilon \in (0, \varepsilon_\delta^*)$. Hence, using Lemma C.5, it follows that, given any arbitrary compact set $\tilde{K}_x \times \tilde{K}_{\hat{q}} \subset K^V$, every solution to $\mathcal{H}_{2,K}$ with $(x(0,0), \hat{q}(0,0)) \in \tilde{K}_x \times \tilde{K}_{\hat{q}}$ is complete. Therefore, by [30, Thm. 1], for any pair $\hat{t}, \hat{j} > 0$ there exists $\tilde{\varepsilon} \in (0, \varepsilon_\delta^*)$ such that for each $\varepsilon \in (0, \tilde{\varepsilon}]$ and each solution z to $\mathcal{H}_{2,K}$, with $z(0,0) \in K_x \times K_{\hat{q}}$, there exists a solution x to $\mathcal{H}_1$such that x and z are $(\hat{t}, \hat{j}, \nu)-$close. The result follows by using $\varepsilon^{**} = \min\{\tilde{\varepsilon}, \varepsilon_\delta^*\}$. ■

Proof of Theorem 5.4

The proof uses tools recently developed for hybrid extremum seeking control [111, 25]. Specifically, we show that all the assumptions needed to apply [25] are satisfied. In particular, using a Taylor expansion of the form $\phi_i(q+\varepsilon_a\tilde{\mu})\tilde{\mu}_i = \tilde{\mu}_i\phi_i(q)+\varepsilon_a\tilde{\mu}_i\tilde{\mu}^\top\nabla\phi_i(q)+\tilde{\mu}_i\mathcal{O}(\varepsilon_a^2)$, and the fact that $|\tilde{\mu}_i| \leq 1$ for all $i \in \mathcal{V}$, and that $\frac{1}{\tilde{L}}\int_0^{\tilde{L}} \tilde{\mu}_i(t)\tilde{\mu}(t)^\top dt = e_i$, where $\tilde{L} = 2\pi\text{LCM}\{1/\varsigma_1, 1/\varsigma_2, \ldots, 1/\varsigma_n\}$ and LCM denotes the least common multiple, the

average dynamics of $\mathcal{H}_3$ are precisely given by $\mathcal{H}_3^A = (C_1, F_1^A, D_1, G_1)$, where G_1, C_1 and D_1 are given by (4.9a), (4.9e), and (4.9f), respectively, and F_1^A is given by:

$$F_1^A(x) = \begin{pmatrix} 2\mathcal{D}(\tau)^{-1}(p-q) \\ -2\mathcal{D}(\tau)\left(\mathcal{G}(q) + \mathcal{O}(\varepsilon_a)\right) \\ \eta \mathbf{1}_n \end{pmatrix}. \tag{C.27}$$

It follows that, on compact sets, we have

$$F_1^A(x) \in \overline{\text{con}} F_1(x + k\varepsilon_a \mathbb{B}) + k\varepsilon_a \mathbb{B}, \tag{C.28}$$

for some $k > 0$, where F_1 was defined in (4.9c). Thus, any solution of the average dynamics $\mathcal{H}_3^A$ is also a solution of an inflated HDS generated from $\mathcal{H}_1$. By [16, Thm. 7.21], we conclude that, under the Assumptions of Theorems 4.1-5.2, system $\mathcal{H}_3^A$ renders SGPAS as $\varepsilon_a \to 0^+$ the compact set $\mathcal{A}$. Since $\mathcal{H}_3^A$ and $\mathcal{H}_1$ are nominally well-posed, all the assumptions needed to apply [25, Thm.7] are satisfied, and we can conclude that $\mathcal{H}_3$ renders SGPAS as $(\varepsilon_p, \varepsilon_a) \to 0^+$ the compact set $\mathcal{A} \times \mathbb{T}^n$. Item (b) follows directly by [25, Prop. 6]. ■

C.2 Proofs of Section 5.2

Proof of Proposition 5.2

For the purpose of clarity, we divide the proof of Proposition 5.2 into multiple lemmas.

> **Lemma C.6.** Suppose that Assumption 5.6 holds; then, there exists a unit vector $q \in \mathbb{R}^N$ such that items (a), (b), and (c) of Proposition 5.2 hold.

Proof. Items (a)-(b) follow directly by [276, Prop. 1]. To show item (c), we use the expressions in (5.51) and (5.52), and by direct substitution we obtain:

$$\begin{aligned}\boldsymbol{\Sigma}+\boldsymbol{\Omega} &= k_r\mathbf{Q}\boldsymbol{\Delta}+\frac{k_{\mathrm{c}}}{2}\left(\mathbf{Q}\mathbf{L}+\mathbf{L}^\top\mathbf{Q}\right)+\frac{k_{\mathrm{c}}}{2}\left(\mathbf{Q}\mathbf{L}-\mathbf{L}^\top\mathbf{Q}\right)\\ &= k_r\mathbf{Q}\boldsymbol{\Delta}+k_{\mathrm{c}}\mathbf{Q}\mathbf{L}.\end{aligned}$$

Applying a left-multiplication by $\mathbf{Q}^{-1}$ and a right-multiplication by $\tilde{\theta}$ leads to

$$\mathbf{Q}^{-1}\left(\boldsymbol{\Sigma}+\boldsymbol{\Omega}\right)\tilde{\theta} = k_r\boldsymbol{\Delta}\tilde{\theta}+k_{\mathrm{c}}\mathbf{L}\tilde{\theta}, \tag{C.29}$$

and since $\tilde{\theta}=\theta-\mathbf{1}_N\otimes\theta^\star$, and $\mathbf{L}(\mathbf{1}_N\otimes\theta^\star)=0$, we obtain:

$$\mathbf{Q}^{-1}\left(\boldsymbol{\Sigma}+\boldsymbol{\Omega}\right)\tilde{\theta} = k_r\boldsymbol{\Delta}\tilde{\theta}+k_{\mathrm{c}}\mathbf{L}\theta.$$

Finally, we show that $\boldsymbol{\Phi}(\theta,0) = \boldsymbol{\Delta}\tilde{\theta}$. Indeed, since $\boldsymbol{\Phi}(\theta,0) = (\Phi_1(\theta_1,0),\ldots,\Phi_N(\theta_N,0))$ and $\Phi_i(\theta_1,0)$ is given by (5.33), we have:

$$\begin{aligned}\Phi_i(\theta_i,0) &= \sum_{k=1}^{\bar{k}_i}\phi_i(t_{i,k})\left(\phi_i(t_{i,k})^\top\theta_i-\phi_i(t_{i,k})^\top\theta^\star\right)\\ &= \sum_{k=1}^{\bar{k}_i}\phi_i(t_{i,k})\phi_i(t_{i,k})^\top\left(\theta_i-\theta^\star\right)=\Delta_i\tilde{\theta}_i,\quad\forall\, i\in\mathcal{V},\end{aligned}$$

which implies $\boldsymbol{\Phi}(\theta,0)=\mathbf{diag}(\{\Delta_1,\ldots,\Delta_N\})\tilde{\theta}=\boldsymbol{\Delta}\tilde{\theta}$. ■

Lemma C.7. There exists $\chi\in\mathcal{K}_\infty$ such that (5.53) holds.

Proof. Consider the following matrix:

$$\mathbf{W}(t) \coloneqq \Big[\boldsymbol{\Omega}+k_t\mathbf{Q}\mathbf{A}(t)\Big]\Big[\boldsymbol{\Omega}+k_t\mathbf{Q}\mathbf{A}(t)\Big]^\top, \tag{C.30}$$

and recall that for any symmetric matrix $A \in \mathbb{R}^{n\times n}$ we have $A \preceq \lambda_{\max}(A) I_n$ [277, Cor. 10.4.2] and $\lambda_{\max}(A) \leq \sigma_{\max}(A) = \|A\|$ [277, Fact 7.12.9], where $\lambda_{\max}(A)$ and $\sigma_{\max}(A)$ are the maximum eigenvalue and the maximum singular value of A, respectively. By using these facts, together with the sub-multiplicativity of the matrix norm, we obtain that:

$$\begin{aligned}\mathbf{W}(t) &= \mathbf{\Omega}\mathbf{\Omega}^\top + k_t \left(\mathbf{\Omega}\mathbf{A}(t)^\top \mathbf{Q} + \mathbf{Q}\mathbf{A}(t)\mathbf{\Omega}^\top\right) \\ &\quad + k_t^2 \mathbf{Q}\mathbf{A}(t)\mathbf{Q}\mathbf{A}(t)^\top \\ &\preceq \left(\overline{\sigma}_{\mathbf{\Omega}}^2 + 2\overline{\sigma}_{\mathbf{\Omega}}\overline{\sigma}_{\mathbf{Q}}\|\mathbf{A}(t)\| k_t + \overline{\sigma}_{\mathbf{Q}}^2 \|\mathbf{A}(t)\|^2 k_t^2\right) I_{Nn}. \end{aligned} \tag{C.31}$$

Since $\phi_i(\cdot)$ is uniformly bounded, there exists $\overline{\phi} > 0$ such that $\phi_i(t) < \overline{\phi}$ for all $i \in \mathcal{V}$ and all $t \in \mathbb{R}$. Combining this fact with the diagonal structure of $\mathbf{A}(t)$ leads to $\|\mathbf{A}(t)\| \leq (\overline{\phi})^2$. By using this bound in (C.31) we obtain:

$$\mathbf{W}(t) \preceq \left(\overline{\sigma}_{\mathbf{\Omega}}^2 + 2\overline{\sigma}_{\mathbf{\Omega}}\overline{\sigma}_{\mathbf{Q}}\overline{\phi}^2 k_t + \overline{\sigma}_{\mathbf{Q}}^2 \overline{\phi}^4 k_t^2\right) I_{Nn}.$$

The result follows using $\chi(k_t) \coloneqq \sqrt{2\overline{\sigma}_{\mathbf{\Omega}}\overline{\sigma}_{\mathbf{Q}}\overline{\phi}^2 k_t + \overline{\sigma}_{\mathbf{Q}}^2 \overline{\phi}^4 k_t^2}$, which is clearly a class-$\mathcal{K}_\infty$ function. ■

Proof of Proposition 5.3

We divide the proof into two lemmas:

Lemma C.8. Under Assumption 5.6, item (d) of Proposition 5.2 holds, i.e., $\mathbf{\Sigma}$ is positive definite.

Proof. We present the proof step-by-step.

(a) First, note that $\mathbf{Q}\mathbf{\Delta} = \mathbf{\Delta}\mathbf{Q}$ since $\mathbf{Q} = \mathcal{Q} \otimes I_n = \mathbf{diag}\left(\{q_1 I_n, \ldots, q_N I_n\}\right)$, $\mathbf{\Delta} = \operatorname{diag}\left(\{\Delta_1, \ldots, \Delta_N\}\right)$, with $\Delta_i \coloneqq \sum_{k=1}^{\overline{k}_i} \phi(t_{i,k})\phi(t_{i,k})^\top \in \mathbb{R}^{n\times n}$, and $q_i I_n \Delta_i = \Delta_i q_i I_n$

trivially. Then, since $\mathbf{Q} \succ 0$ and $\mathbf{\Delta} \succeq 0$ it follows that $\mathbf{Q\Delta} \succeq 0$. (b) Let the eigenvalues of the matrix $\mathcal{L}^\top \mathcal{Q} + \mathcal{Q}\mathcal{L}$ be organized as $0{=}\lambda_1{<}\lambda_2{\leq}\cdots{\leq}\lambda_N$, and let $v_i \in \mathbb{R}^N$ be the eigenvector that corresponds to the eigenvalue λ_i and satisfies $|v_i| = 1$. It follows that $v_1 = \frac{1}{\sqrt{N}}\mathbf{1}_N$. (c) Let $\mathbf{M} := \mathbf{L}^\top\mathbf{Q} + \mathbf{QL}$, and let

$$\begin{aligned}\mathbf{E} &:= \frac{1}{\sqrt{N}}\left[\mathbf{1}_N \otimes e_1, \cdots, \mathbf{1}_N \otimes e_n\right] \\ \mathbf{U} &:= \left[v_2 \otimes e_1, \cdots, v_2 \otimes e_n, \cdots, v_N \otimes e_1, \cdots, v_N \otimes e_n\right]\end{aligned}$$

where the vectors e_i denote the standard basis in $\mathbb{R}^n$. Note that the matrices $\mathbf{E} \in \mathbb{R}^{Nn\times n}$ and $\mathbf{U} \in \mathbb{R}^{Nn\times(N-1)n}$ characterize the null space and the range space of $\mathbf{M}$, respectively.

(d) Let $\hat{\mathbf{x}} \in \mathbb{R}^{Nn}$ be a unit vector, which we can write as

$$\hat{\mathbf{x}} = \mathbf{Eb} + \mathbf{Uc} \tag{C.32}$$

where $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{c} \in \mathbb{R}^{(N-1)n}$ satisfy $|\mathbf{b}|^2 + |\mathbf{c}|^2 = 1$.

(e) Since $\mathbf{E}$ can be written as $\mathbf{E} = \frac{1}{\sqrt{N}}\mathbf{1}_N \otimes I_n$, and $\mathbf{Q\Delta} = \mathbf{diag}\left(\{q_1\Delta_1, \ldots, q_N\Delta_N\}\right)$, we have that

$$\mathbf{Q\Delta E} = \frac{1}{\sqrt{N}}\begin{pmatrix} q_1\Delta_1 \\ \vdots \\ q_N\Delta_N \end{pmatrix},$$

which leads to

$$\mathbf{E}^\top\mathbf{Q\Delta E} = \frac{1}{N}\sum_{i=1}^{N} q_i\Delta_i. \tag{C.33}$$

Using (C.32) and (C.33), we obtain

$$\hat{\mathbf{x}}^\top\mathbf{Q\Delta}\hat{\mathbf{x}} \geq \mathbf{b}^\top\mathbf{E}^\top\Delta\mathbf{Eb} + 2\mathbf{b}^\top\mathbf{E}^\top\mathbf{Q\Delta Uc}$$

$$\geq \underline{\sigma}_{\mathbf{Q}}\bar{\alpha}|\mathbf{b}|^2 + 2\mathbf{b}^\top \mathbf{E}^\top \mathbf{Q}\boldsymbol{\Delta}\mathbf{U}\mathbf{c},$$

where $\bar{\alpha} := \alpha/N$, α is given by Assumption 5.6, $\underline{\sigma}_{\mathbf{Q}}$ and $\overline{\sigma}_{\mathbf{Q}}$ are defined in (5.50), and $\overline{\sigma}_{\boldsymbol{\Delta}} := |\boldsymbol{\Delta}|$. Moreover, since $|2\mathbf{b}^\top \mathbf{E}^\top \mathbf{Q}\boldsymbol{\Delta}\mathbf{U}\mathbf{c}| \leq 2|\mathbf{b}||\mathbf{c}|\overline{\sigma}_{\mathbf{Q}}\overline{\sigma}_{\boldsymbol{\Delta}}$, and using $|\mathbf{c}| = \sqrt{1-|\mathbf{b}|^2}$, we obtain:

$$\hat{\mathbf{x}}^\top \mathbf{Q}\boldsymbol{\Delta}\hat{\mathbf{x}} \geq \underline{\sigma}_{\mathbf{Q}}\bar{\alpha}|\mathbf{b}|^2 - 2\overline{\sigma}_{\mathbf{Q}}\overline{\sigma}_{\boldsymbol{\Delta}}|\mathbf{b}|\sqrt{1-|\mathbf{b}|^2} =: \xi_1(\mathbf{b}). \tag{C.34}$$

(f) On the other hand, we have that

$$\hat{\mathbf{x}}^\top \mathbf{M}\hat{\mathbf{x}} \geq \lambda_2|\mathbf{c}|^2 = \lambda_2(1-|\mathbf{b}|^2) =: \xi_2(\mathbf{b}). \tag{C.35}$$

Since by the construction of $\boldsymbol{\Sigma}$ in (5.52) we have $\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}\hat{\mathbf{x}} = k_r \hat{\mathbf{x}}^\top \mathbf{Q}\boldsymbol{\Delta}\hat{\mathbf{x}} + \frac{k_c}{2}\hat{\mathbf{x}}^\top \mathbf{M}\hat{\mathbf{x}}$, the above bounds imply that $\boldsymbol{\Sigma} \succeq \underline{\sigma}_{\boldsymbol{\Sigma}} I_{Nn}$, where

$$\underline{\sigma}_{\boldsymbol{\Sigma}} \;\geq\; \min_{0\leq\nu\leq 1} \max\left\{k_r\xi_1(\nu), \frac{k_c}{2}\xi_2(\nu)\right\}, \tag{C.36}$$

with ξ_1 given by (C.34) and ξ_2 given by (C.35).

(g) Next, we study (C.36) and we show that this lower bound is indeed positive. Since, by item (a), $\mathbf{Q}\boldsymbol{\Delta} \succeq 0$, without loss of generality we can assume that the first term in the brackets in (C.36) is non-negative. Indeed, suppose by contradiction that such term is negative. Then, since $\mathbf{Q}\boldsymbol{\Delta} \succeq 0$, we can take $\xi(\mathbf{b})$ as a non-negative lower bound for $\hat{\mathbf{x}}^\top \boldsymbol{\Sigma}\hat{\mathbf{x}}$, and since $\xi(\mathbf{b}) = 0$ only if $\mathbf{b} = 1$, we obtain that for such $\mathbf{b}$ the first term in the brackets is indeed positive.

(h) To get a closed form of the expression in (C.36), let $\nu = \sin(\theta),\ \theta \in [0, \pi/2]$. In

the θ variable, (C.36) becomes:

$$\min_{0\le\theta\le\frac{\pi}{2}} \max\Big\{k_1(1-\cos(2\theta))-k_2\sin(2\theta), k_3(1+\cos(2\theta))\Big\},$$

where the constants $k_1, k_2, k_3 > 0$ are given by $k_1 \coloneqq \frac{k_r\underline{\sigma}_{\mathbf{Q}}\bar{\alpha}}{2}$, $k_2 \coloneqq k_r\overline{\sigma}_{\mathbf{\Delta}}\overline{\sigma}_{\mathbf{Q}}$, $k_3 \coloneqq \frac{k_c}{4}\lambda_2$. Further simplifying, we obtain

$$\begin{aligned}\min_{0\le\theta\le\frac{\pi}{2}} \max\Bigg\{& k_1 - \sqrt{k_1^2+k_2^2}\sin\left(2\theta+\tan^{-1}\left(\frac{k_1}{k_2}\right)\right), \\ & k_3 + k_3\cos(2\theta))\Bigg\} \\ \coloneqq \min_{0\le\theta\le\frac{\pi}{2}} \max\Big\{& g_1(\theta),\ g_2(\theta)\Big\}. \end{aligned} \tag{C.37}$$

We argue that the intersection point $\theta^* \in [0, \frac{\pi}{2}]$ of the trigonometric curves $g_1(\theta),\ g_2(\theta)$ solves the min-max problem (C.37).

(i) To establish the existence of such $\theta^* \in [0, \frac{\pi}{2}]$, we use the following facts: (i) $k_1, k_2, k_3 > 0$, (ii) $g_1(0) = 0$, $g_1(\frac{\pi}{2}) = 2k_1 > 0$, $\frac{dg_1(\theta)}{d\theta} = -2\sqrt{k_1^2+k_2^2}\cos\Big(2\theta+\tan^{-1}(\frac{k_1}{k_2})\Big)$, (ii) $g_2(0) = 2k_3 > 0$, $g_2(\frac{\pi}{2}) = 0$, $\frac{dg_2(\theta)}{d\theta} = -2k_3\sin(2\theta)$. Since g_1 and g_2 are continuous functions, the previous conditions imply the existence of a point θ^* such that $g_1(\theta^*) = g_2(\theta^*)$. Moreover, since g_2 is decreasing on $[0, \frac{\pi}{2}]$ with $g_2(0) > 0$, $g_2(\frac{\pi}{2}) = 0$, $g_1(0) = 0$, $g_1(\frac{\pi}{2}) > 0$, and $\frac{dg_1(\theta)}{d\theta} = -2\sqrt{k_1^2+k_2^2}\cos\Big(2\theta+\tan^{-1}(\frac{k_1}{k_2})\Big)$, it follows that the intersection point θ^* is in fact the minimum of the point-wise maximum of $g_1(\theta)$ and $g_2(\theta)$. See Figure C.1 for an illustration of this step.

(j) By computing the intersection point θ^*, we obtain

$$\theta^* = \frac{1}{2}\left[\cos^{-1}\left(\frac{k_1-k_3}{\sqrt{(k_1+k_3)^2+k_2^2}}\right)+\tan^{-1}\left(\frac{k_2}{k_1+k_3}\right)\right].$$

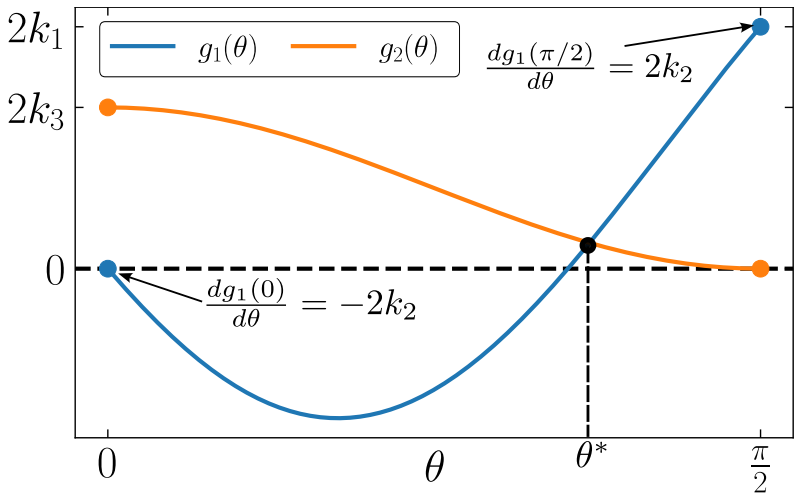


Figure C.1. Illustration of step (i) in the proof of Proposition C.8.

Substituting the values of k_1, k_2 and k_3, establishes the existence of a positive lower bound on the constant $\underline{\sigma}_{\mathbf{\Sigma}}$ that satisfies $\mathbf{\Sigma} \succeq \underline{\sigma}_{\mathbf{\Sigma}} I_{Nn}$, given by

$$\underline{\sigma}_{\mathbf{\Sigma}} \geq \frac{k_c \lambda_2}{4}\big[1 + \cos(\theta^*)\big], \tag{C.38}$$

where $\theta^* = \theta_1^* + \theta_2^*$, with

$$\theta_1^* = \cos^{-1}\left(\frac{2k_r \underline{\sigma}_{\mathbf{Q}}\alpha - k_c\lambda_2}{\sqrt{(2k_r\underline{\sigma}_{\mathbf{Q}}\alpha + k_c\lambda_2)^2 + 16k_r^2\overline{\sigma}_{\mathbf{\Delta}}^2\overline{\sigma}_{\mathbf{Q}}^2}}\right)$$

and $\theta_2^* = \tan^{-1}\left(\frac{4\overline{\sigma}_{\mathbf{\Delta}}k_r\overline{\sigma}_{\mathbf{Q}}}{2\underline{\sigma}_{\mathbf{Q}}k_r\alpha + k_c\lambda_2}\right)$. Note that $\cos(\theta^*) \in [0,1]$ since $\theta^* \in [0, \pi/2]$, which implies that $\underline{\sigma}_{\mathbf{\Sigma}} > 0$. ■

Lemma C.9. Let λ_N be the largest eigenvalue of $\mathcal{L}^\top\mathcal{Q} + \mathcal{Q}\mathcal{L}$. Then, under Assumption 5.6, the matrix $\mathbf{\Sigma}$ satisfies

$$\mathbf{\Sigma} \preceq \left(k_r\overline{\sigma}_{\mathbf{Q}}\overline{\sigma}_{\mathbf{\Delta}} + \frac{k_c}{2}\lambda_N\right) I_{Nn}. \tag{C.39}$$

Proof. By the definition of $\mathbf{\Delta}$ and $\overline{\sigma}_{\mathbf{\Delta}}$, the term $\mathbf{Q\Delta}$ satisfies: $\mathbf{Q\Delta} \preceq \overline{\sigma}_{\mathbf{Q}}\overline{\sigma}_{\mathbf{\Delta}} I_{Nn}$. By the definition of λ_N and the fact that $\mathbf{QL} = \mathcal{QL} \otimes I_n$ by the properties of the Kronecker product, it follows that $\mathbf{L}^\top\mathbf{Q} + \mathbf{QL} \preceq \lambda_N I_{Nn}$. Note that $\lambda_N > 0$ since, as stated in the proof of Lemma 5.2, $\mathcal{QL} + \mathcal{L}^\top\mathcal{Q}$ is a nonzero and symmetric M-matrix. Combining these arguments we obtain (C.39). ■

Proof of Theorem 5.5

We follow a (hybrid) Lyapunov-based approach to study the HDS $\mathcal{H}_c$ with input u, in the error coordinates

$$\tilde{y}_c = (\tilde{x}_c, s) \coloneqq ((\tilde{\theta}, \tilde{p}, \tau_c), s),$$

where $\tilde{\theta} = \theta - \mathbf{1}_N \otimes \theta^\star$, $\tilde{x}_c = (\tilde{\theta}, \tilde{p}, \tau_c)$, and $\tilde{p} = p - \mathbf{1}_N \otimes \theta^\star$. In these new coordinates, the HDS with input u becomes

$$\tilde{\mathcal{H}}_c = (\mathbf{C}_c \times \mathbb{R}_{\geq 0}, \tilde{\mathbf{F}}_c, \mathbf{D}_c \times \mathbb{R}_{\geq 0}, \mathbf{G}_c),$$

where $\tilde{\mathbf{F}}_c(\tilde{y}_c, u) \coloneqq \hat{\mathbf{F}}_c(\tilde{y}_c, u) \times [0, \omega] \times \{1\}$, with $\hat{\mathbf{F}}_c$ given by (5.38). For this system, we will study stability properties with respec to the set $\tilde{\mathcal{A}}_c \times \mathbb{R}_{\geq 0}$, where

$$\tilde{\mathcal{A}}_c \coloneqq \{0\} \times \{0\} \times [T_0, T]. \tag{C.40}$$

Proof of Theorem 5.5-(a)

We establish item (a) of Theorem 5.5 via a sequence of lemmas. The following lemma follows directly from the uniform boundedness assumption on the regressors ϕ and the definition of $\mathbf{U}$ in (5.39).

Lemma C.10. There exist $\overline{\phi} > 0$ such that $|\mathbf{U}(s)| \leq \overline{\phi}|u|$ for all $s \geq 0$.

Next, we consider the Lyapunov function

$$V(\tilde{y}_c) \coloneqq \frac{|\tilde{p}-\tilde{\theta}|^2_{\mathbf{Q}}}{4} + \frac{|\tilde{p}|^2_{\mathbf{Q}}}{4} + \tau_c^2 \frac{|\tilde{\theta}|^2_{\mathbf{\Sigma}}}{2}. \tag{C.41}$$

and we study its behavior during flows and jumps of $\tilde{\mathcal{H}}_c$. and present a lemma and two auxiliary propositions.

Lemma C.11. There exist constants $\overline{c} > \underline{c} > 0$ such that

$$\underline{c}|\tilde{y}_c|^2_{\tilde{\mathcal{A}}_c \times \mathbb{R}_{\geq 0}} \leq V(\tilde{y}_c) \leq \overline{c}|\tilde{y}_c|^2_{\tilde{\mathcal{A}}_c \times \mathbb{R}_{\geq 0}},$$

for all $\tilde{y}_c \in (\mathbf{C}_c \cup \mathbf{D}_c) \times \mathbb{R}_{\geq 0}$.

Proof. Since, by the definition of $\tilde{\mathcal{H}}_c$, we always have $s \in \mathbb{R}_{\geq 0}$, we just need to study $|\tilde{x}_c|_{\tilde{\mathcal{A}}_c}$. To establish the lower bound, and using the definition of the norm $|\cdot|_{\mathbf{P}}$, and since $\tau_c \geq T_0$ for all $\tilde{x}_c \in \mathbf{C}_c \cup \mathbf{D}_c$, we directly obtain that $|\tilde{p}|^2_{\mathbf{Q}} \geq \underline{\sigma}_{\mathbf{Q}}|\tilde{p}|^2$ and $\tau^2|\tilde{\theta}|^2_{\mathbf{\Sigma}} \geq \underline{\sigma}_{\mathbf{\Sigma}} T_0^2 |\tilde{\theta}|^2$. Therefore, $V(\tilde{y}_c) \geq \underline{c}|\tilde{x}_c|^2_{\tilde{\mathcal{A}}_c}$, where $\underline{c} \coloneqq \frac{1}{4}\min\left\{\underline{\sigma}_{\mathbf{Q}}, 2\underline{\sigma}_{\mathbf{\Sigma}} T_0^2\right\}$. To establish the upper bound, we use (5.50) together with the fact that $\tau \leq T$ to obtain that $V(\tilde{y}_c) \leq \frac{1}{4}(2\overline{\sigma}_{\mathbf{Q}}|\tilde{\theta}|^2 + 3\overline{\sigma}_{\mathbf{Q}}|\tilde{p}|^2 + 2T^2|\tilde{\theta}|^2_{\mathbf{\Sigma}})$, where we also used the fact that $|\tilde{p}-\tilde{\theta}|^2 \leq 2(|\tilde{\theta}|^2 + |\tilde{p}|^2)$. Using Lemma C.9, we obtain $V(\tilde{y}_c) \leq \overline{c}|\tilde{x}_c|^2_{\tilde{\mathcal{A}}}$, with $\overline{c} \coloneqq \frac{1}{4}\max\left\{3\overline{\sigma}_{\mathbf{Q}}, T^2(2k_r\overline{\sigma}_{\mathbf{Q}}\overline{\sigma}_{\mathbf{\Delta}} + k_c\lambda_N) + 2\overline{\sigma}_{\mathbf{Q}}\right\}$. ■

Lemma C.12. Suppose that $T < \overline{\mathbf{T}}$; then, there exists $\varrho > 0$ and $\gamma > 0$ such that V satisfies $\dot{V}(\tilde{y}_c) \leq -\varrho V(\tilde{y}_c) + \gamma|u|^2$, for all $\tilde{y}_c \in \mathbf{C}_c \times \mathbb{R}_{\geq 0}$.

Proof. By direct computation, we have:

$$\dot{V}(\tilde{y}_c) = -\tau_c \begin{pmatrix} (\tilde{p}-\tilde{\theta})^\top & \tilde{\theta}^\top \end{pmatrix} \mathbf{V}_w(\tau_c, s) \begin{pmatrix} \tilde{p}-\tilde{\theta} \\ \tilde{\theta} \end{pmatrix}$$

$$+ \tau_c(2\tilde{p} - \tilde{\theta})^\top \mathbf{Q}\mathbf{U}(s), \tag{C.42}$$

where

$$\mathbf{V}_w(\tau_c, s) \coloneqq \begin{pmatrix} \frac{\mathbf{Q}}{\tau_c^2} & \hat{\mathbf{\Omega}}(s) \\ \hat{\mathbf{\Omega}}(s)^\top & (1-w)\mathbf{\Sigma} + k_t \mathbf{Q}\mathbf{A}(s) \end{pmatrix},$$

for all $w \in [0, \omega]$, where

$$\hat{\mathbf{\Omega}}(s) \coloneqq \mathbf{\Omega} + k_t \mathbf{Q}\mathbf{A}(s),$$

and where we used the fact that $x_1^\top \mathbf{\Omega} x_1 = 0$ and Proposition 5.2. Using the definitions of $\underline{\sigma}_{\mathbf{Q}}$, $\underline{\sigma}_{\mathbf{\Sigma}}$, and $\overline{\sigma}_{\mathbf{\Omega}}$ provided in Propositions 5.2-5.3, and Lemma C.15 in the Appendix, it follows that $\mathbf{V}_w(\tau_c, s) \succeq \underline{\mathbf{v}} I_{Nn}$, for all $\tau_c \in [T_0, T]$, all $w \in [0, \omega]$, and all $s \in \mathbb{R}_{\geq 0}$, with

$$\underline{\mathbf{v}} \coloneqq \frac{(1-\omega)\underline{\sigma}_{\mathbf{\Sigma}}\underline{\sigma}_{\mathbf{Q}} - T^2(\overline{\sigma}_{\mathbf{\Omega}}^2 + \chi(k_t)^2)}{T^2(1-\omega)\underline{\sigma}_{\mathbf{\Sigma}} + \underline{\sigma}_{\mathbf{Q}}} > 0, \tag{C.43}$$

and $\chi \in \mathcal{K}_\infty$. Using the Cauchy-Schwartz inequality to upper-bound the last term in (C.42), and since $T_0 \leq \tau_c \leq T$ and $|\tilde{x}_c|^2 \leq 3(|\tilde{p} - \tilde{\theta}|^2 + |\tilde{p}|^2)$ for all $\tilde{x}_c \in \mathbf{C}_c \cup \mathbf{D}_c$, we obtain:

$$\begin{aligned} \dot{V}(\tilde{y}_c) &\leq -T_0 \underline{\mathbf{v}}(|\tilde{p} - \tilde{\theta}|^2 + |\tilde{\theta}|^2) + 2T(|\tilde{p}| + |\tilde{\theta}|)\|\mathbf{Q}\||\mathbf{U}(s)| \\ &\leq -\frac{\mathbf{v}}{3} T_0 |\tilde{x}_c|^2 + 2\sqrt{2}\overline{\sigma}_{\mathbf{Q}} \overline{\phi} T |\tilde{x}_c||u| \\ &\leq -\left(\frac{\mathbf{v}}{3} T_0 - \frac{1}{\epsilon}\right)|\tilde{x}_c|^2 + 2\epsilon \left(\overline{\sigma}_{\mathbf{Q}} \overline{\phi} T\right)^2 |u|^2, \end{aligned} \tag{C.44}$$

for all $\epsilon > 0$ and all $w \in [0, \omega]$, where the last inequality follows from the fact that $ab \leq \frac{1}{4\epsilon} a^2 + \epsilon b^2$ for all $\epsilon > 0$. Setting $\epsilon \coloneqq \frac{3(1+\varepsilon)}{T_0 \nu}$ for $\varepsilon > 0$ and using the lower bound of Lemma C.11, the expression in (C.44) yields:

$$\dot{V}(\tilde{y}_c) \leq -\frac{\varepsilon}{1+\varepsilon} \frac{\nu T_0}{3\overline{c}} V(\tilde{y}_c) + (1+\varepsilon) \frac{6}{\nu T_0} \left(\overline{\sigma}_{\mathbf{Q}} \overline{\phi} T\right)^2 |u|^2. \tag{C.45}$$

The result follows by setting $\varrho \coloneqq \frac{\nu T_0}{3\overline{c}} \frac{\varepsilon}{1+\varepsilon}$ and letting $\gamma \in \mathcal{K}_\infty$ be defined as $\gamma(r) \coloneqq (1+\varepsilon)\frac{6}{\nu T_0}\left(\overline{\sigma}_{\mathbf{Q}}\overline{\phi}T\right)^2 r$. ■

Lemma C.13. Suppose that $T > \underline{\mathbf{T}}$; then,

$$V(\tilde{y}_c^+) \leq (\mu_T)^{\underline{\eta}} V(\tilde{y}_c), \quad \forall\ \tilde{y}_c \in (\mathbf{C}_c \cup \mathbf{D}_c) \times \mathbb{R}_{\geq 0},$$

where $\underline{\eta} \coloneqq \min_{i\in\mathcal{V}} \eta_i$, and $\mu_T \coloneqq (\underline{\mathbf{T}}/T)^2$.

Proof. Using the definition of the jump map $\mathbf{G}_c$, for all $\tilde{y}_c \in \mathbf{D}_c \times \mathbb{R}_{\geq 0}$ we have:

$$\begin{aligned} 4V(\tilde{y}_c^+) = {} & |\mathbf{R}_\eta\tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\tilde{p} - \tilde{\theta}|^2_{\mathbf{Q}} \\ & + |\mathbf{R}_\eta\tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\tilde{p}|^2_{\mathbf{Q}} + 2T_0|\tilde{\theta}|^2_{\mathbf{\Sigma}}, \end{aligned} \tag{C.46}$$

where $\mathbf{R}_\eta \coloneqq \operatorname{diag}(\eta) \otimes I_n$. By Lemma C.16 in the Appendix, the change of V during jumps, given by $\Delta V(\tilde{y}_c) \coloneqq V(\tilde{y}_c^+) - V(\tilde{y}_c)$, satisfies:

$$\begin{aligned} 4\Delta V(\tilde{y}_c) &= |\tilde{\theta}|^2_{\mathbf{R}_\eta\mathbf{Q}} - |\tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} - |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} + 2T_0^2|\tilde{\theta}_c|^2_{\mathbf{\Sigma}} - 2T^2|\tilde{\theta}|^2_{\mathbf{\Sigma}} \\ &\leq -\left(|\tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} + |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}}\right) + \frac{\overline{\sigma}_{\mathbf{Q}}}{\underline{\sigma}_{\mathbf{\Sigma}}}|\tilde{\theta}|^2_{\mathbf{\Sigma}} + 2T_0^2|\tilde{\theta}|^2_{\mathbf{\Sigma}} - 2T^2|\tilde{\theta}|^2_{\mathbf{\Sigma}} \\ &= -\left(|\tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} + |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}}\right) - (1-\mu_T)2T^2|\tilde{\theta}|^2_{\mathbf{\Sigma}} \\ &\leq -(1-\mu_T)\left(|\tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} + |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta\mathbf{Q}} + 2T^2|\tilde{\theta}|^2_{\mathbf{\Sigma}}\right), \end{aligned}$$

where we also used the fact that $\mu_T \in (0,1)$ whenever $T > \underline{\mathbf{T}}$. It then follows that $\Delta V(\tilde{y}_c) \leq 0$ for all $\tilde{y}_c \in \mathbf{D}_c \times \mathbb{R}_{\geq 0}$. When $\underline{\eta} = 1$, the previous inequality yields $\Delta V(\tilde{y}_c) \leq -(1-\mu_T)V(\tilde{y}_c)$, wich in turn implies that $V(\tilde{y}_c) \leq \mu_T V(\tilde{y}_c)$. ■

By the construction of the dynamics of τ_c, every solution to $\mathcal{H}_c$ is guaranteed to

have intervals of flow with a duration of at least $(T - T_0)/\omega$ between any two consecutive jumps. Combining this fact with Lemmas C.11, C.12, and C.13, it follows that $\tilde{\mathcal{H}}_c$ renders the set $\tilde{\mathcal{A}}_c$ ISS with respect to the input u. The ISS property of the HDS $\mathcal{H}_c$ with respect to $|\cdot|_{\mathcal{A}_c} \times \mathbb{R}_{\geq 0}$ follows directly by employing the change of coordinates $\tilde{y}_c \to y_c$.

Proof of Theorem 5.5-(b)

Let the initial condition satisfy $\tilde{y}_0 := ((\tilde{\theta}(0,0), \tilde{p}(0,0), \tau_c(0,0)), s(0,0)) \in (\mathbf{C}_c \times \mathbf{D}_c) \times \mathbb{R}_{\geq 0}$, and let $(\tilde{y}_c, u)$ be a maximal solution pair to $\tilde{\mathcal{H}}_c$ from the initial condition $\tilde{y}_0$, satisfying during flows $\dot{\tau}_c(t,j) = \omega \in (0,1)$ for all $(t,j) \in \mathrm{dom}(\tilde{y}_c)$. By Lemma C.12, we have that $\dot{V}(\tilde{y}_c) \leq -\frac{\varrho}{2} V(\tilde{y}_c)$ for all $\tilde{y}_c \in (\mathbf{C}_c \cup \mathbf{D}_c) \times \mathbb{R}_{\geq 0}$ such that $V(\tilde{y}_c) \geq \frac{2\gamma}{\varrho}|u|^2$. Let

$$\mathcal{R} := \left\{ \tilde{y}_c \in \mathbb{R}^{2nN+1} \times \mathbb{R}_{\geq 0} \ : \ V(\tilde{y}_c) \leq \frac{2\gamma}{\varrho} |u|_\infty^2 \right\}, \tag{C.47}$$

and let $\mathbb{T} := \sup\{\sigma \in \mathbb{R}_{\geq 0} \ : \ \tilde{y}_c(\tilde{t}, \tilde{j}) \notin \mathcal{R}, \ (\tilde{t}, j) \in \mathrm{dom}(\tilde{y}_c), \ 0 \leq \tilde{t} + \tilde{j} \leq \sigma\}$. Then, letting $t_j := \min\{t \in \mathbb{R}_{\geq 0} \ : \ (t,j) \in \mathrm{dom}(\tilde{y}_c)\}$ for every $j \in \mathbb{Z}_{\geq 0}$, and via the comparison lemma, it follows that $V(\tilde{y}_c(t,j)) \leq e^{-\rho(t-t_j)/2} V(\tilde{y}_c(t_j, j))$, for all $(t,j) \in \mathrm{dom}(\tilde{y}_c)$ such that $t_j + j \leq t + j \leq \mathbb{T}$. On the other hand, from Lemma C.13, it follows that $V(\tilde{y}_c(t_{j+1}, j+1)) \leq \mu_T V(\tilde{y}_c(t_{j+1}, j))$, which iterating over j yields:

$$V(\tilde{y}_c(t,j)) \leq e^{-\varrho t/2} \mu_T^j V(\tilde{y}_0), \tag{C.48}$$

for all $(t,j) \in \mathrm{dom}(\tilde{y}_c)$ such that $t + j \leq \mathcal{T}$ and where we have used that $t_0 = 0$. Since $\dot{V}(\tilde{y}_c(t,j)) \leq 0$ if $\tilde{y}_c(t,j) \in \mathcal{R}$ and Proposition C.13 holds for all $(t_j, j) \in \mathrm{dom}(\tilde{y})$, it follows that $\tilde{y}_c(t,j) \in \mathcal{R}$ for all $t + j \geq \mathcal{T}$, meaning that

$$V(\tilde{y}_c(t,j)) \leq \frac{2\gamma}{\varrho} |u|_\infty, \tag{C.49}$$

for all $t+j \geq \mathcal{T}$. The bounds (C.48) and (C.49), together with Lemma C.11 and the time-invariance of $\mathcal{H}_c$, imply that $|\tilde{y}_c(t,j))|^2_{\tilde{\mathcal{A}}_c} \leq \frac{\overline{c}}{\underline{c}}\mu_T^j |\tilde{y}_0|^2_{\tilde{\mathcal{A}}_c} + \frac{2\gamma}{\varrho}|u|_{(t,j)}$ for all $(t,j) \in \mathrm{dom}(\tilde{y}_c)$, where we also used the fact that $e^{-\varrho t/2} \leq 1$ for all $t \in \mathbb{R}_{\geq 0}$. The bound (5.57), is obtained by evaluating the above bound at the hybrid times (t_j, j), noting that $|\tilde{\theta}| \leq |\tilde{y}_c|_{\mathcal{A}_c}$, and via the change of coordinates $\tilde{y}_c \mapsto y_c$. ■

Proof of Theorem 5.6

The proof uses the reduction principle for hybrid systems [16, Corollary 7.24]. First, note that, by construction, $\mathcal{H}$ satisfies the hybrid basic conditions [16, Assump. 6.5]. Since the flow map $\mathbf{F}$ is globally Lipschitz in $\mathbf{C}$, the HDS does not exhibit finite escape times. To study the stability properties of the system, we first intersect the flow set $\mathbf{C}$, the jump set $\mathbf{D}$, and the values of the jump map $\mathbf{G}$ with a compact set $K \subset \mathbb{R}^{(2n+1)N}$. Since τ already evolves in a compact set, we take K only to restrict the states (θ, p, s). The new restricted system is denoted as $\mathcal{H}_K = (\mathbf{C} \cap K, \mathbf{F}, \mathbf{D} \cap K, \mathbf{G} \cap K)$. Since the dynamics of the state τ are independent of (θ, p), we can directly use [148, Prop. 1-(a)] to conclude that, under condition (b) of Theorem 5.6: 1) the set $K \times \mathcal{A}_{\mathrm{sync}}$ is UGAS for the HDS $\mathcal{H}_K$, and 2) τ converges to $\mathcal{A}_{\mathrm{sync}}$ before the hybrid time $(2t^*, 2N)$. It follows that, for all solutions (y, s), and all times $(t, j) \in \mathrm{dom}((y, s))$ such that $t \geq 2t^*$ and $j \geq 2N$, the restricted synchronized HDS behaves as having the centralized master timer τ_c of Section 5.2. Next, we intersect the data of the HDS $\mathcal{H}_K$ with the set $K \times \mathcal{A}_{\mathrm{sync}}$. For this restricted HDS, denoted $\mathcal{H}_{K,\mathcal{A}_{\mathrm{sync}}}$, Theorem 5.5 guarantees UGES of the set $\mathcal{A}$ when $u = 0$. By invoking the reduction principle of [16, Corollary 7.24], we conclude UGES of the set $\mathcal{A}$ for the HDS $\mathcal{H}_K$. Since this system has bounded solutions, and K was arbitrary large, for each compact set of initial conditions K_0 of system $\mathcal{H}$, we can select K sufficiently large such that the restriction in $\mathcal{H}_K$ does not affect the solutions from K_0, obtaining UGES of $\mathcal{A}$ for the original hybrid system $\mathcal{H}$. Now, since the convergence of $\tau \in \mathbb{R}^N$ to $\mathcal{A}_{\mathrm{sync}}$ occurs in finite time after which the stability properties are characterized by Theorem 5.5, we

obtain that $\mathcal{A}$ is UGES for $\mathcal{H}$. ■

Proof of Corollary 5.1

First, note that $j \geq \frac{t'}{(T-T_0)/\omega}$ for any $(t', j) \in \mathrm{dom}(y)$. Therefore, since $\mu(T) \in (0,1)$, the bound (5.57) implies the following slightly looser bound when $u \equiv 0$:

$$|y_j|_{\mathcal{A}}^2 \leq \frac{\overline{c}}{\underline{c}} \left(\mu_T^{\frac{1}{T-T_0}} \right)^{\omega t'} |y_0|_{\mathcal{A}}^2. \tag{C.50}$$

where $\overline{c}$ and $\underline{c}$ come from Lemma C.11. Following similar ideas to [93, 25], and using the definition of $\mu(T)$, we solve the following optimization problem to maximize the rate of contraction over any window of time t':

$$\min_{T \in \mathbb{R}_{>0}} \phi(T) \coloneqq \mu_T^{\frac{1}{T-T_0}}.$$

Computing the derivative of ϕ with respect to T, and equating to zero, we obtain: $T^* = e\sqrt{\frac{\overline{\sigma}_{\mathbf{Q}}}{2\underline{\sigma}_{\mathbf{\Sigma}}} + T_0^2}$, which is the unique minimizer of ϕ. By substituting $T = T^*$ in (C.50), we obtain

$$|y_j|_{\mathcal{A}}^2 \leq \frac{\overline{c}}{\underline{c}} e^{-\frac{2\omega t'}{T^*-T_0}} |y_0|_{\mathcal{A}}^2. \tag{C.51}$$

Thus, to have $|y_j|_{\mathcal{A}}^2 \leq \varepsilon$ for a given $\varepsilon > 0$, it suffices to have that

$$t' \geq \frac{1}{2\omega} (T^* - T_0) \log \left(\frac{1}{\varepsilon} \frac{\overline{c}}{\underline{c}} |y_0|_{\mathcal{A}}^2 \right). \tag{C.52}$$

Moreover, note that the right hand side of (C.51) is of order $\mathcal{O}\left(e^{-\sqrt{\underline{\sigma}_{\mathbf{\Sigma}}/\overline{\sigma}_{\mathbf{Q}}}t'}\right)$. ■

Proof of Corollary 5.2

The arguments are similar to those used in the proof of Theorem 5.5 by using the fact that in Lemma C.13 the expression in (C.46) yields $\Delta V(\tilde{y}_c) \leq 0$ whenever $\eta = 0$.

Auxiliary Lemmas

Lemma C.14. Consider the following block triangular matrix:

$$M \coloneqq \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

Suppose that M is non-singular. Then, the minimum singular value of M, $\sigma_{\min}(M)$, satisfies

$$\sigma_{\min}(M) \geq \frac{1}{\sqrt{\|A^{-1}\|^2(1+\|BD^{-1}\|^2)+\|D^{-1}\|^2}}.$$

Proof. First, since the inverse of the block triangular matrix M is given by

$$M^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix},$$

we can upper-bound the 2-norm matrix of M^{-1}:

$$\begin{aligned} \|M^{-1}\|^2 &= \max_{|u|^2+|v|^2=1} \left| \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \right|^2 \\ &= \max_{|u|^2+|v|^2=1} \left| \begin{bmatrix} A^{-1}u - A^{-1}BD^{-1}v \\ D^{-1}v \end{bmatrix} \right|^2 \\ &= \max_{|u|^2+|v|^2=1} \left|A^{-1}u - A^{-1}BD^{-1}v\right|^2 + \left|D^{-1}v\right|^2 \\ &\leq \|A^{-1}\|^2(1+\|BD^{-1}\|^2)+\|D^{-1}\|^2. \end{aligned} \tag{C.53}$$

Then, since the minimum singular value of a matrix is the inverse of the 2-norm of the inverse matrix, i.e., $\sigma_{\min}(M) = \frac{1}{\|M^{-1}\|}$, we can use (C.53) to obtain the result. ■

Lemma C.15. For each $\tau_c \in [T_0, T]$ and $s \in \mathbb{R}_{\geq 0}$, consider the following block matrix

$$\mathbf{V}_w(\tau_c, s) \coloneqq \begin{pmatrix} \frac{1}{\tau^2}\mathbf{Q} & \hat{\mathbf{\Omega}}(s) \\ \hat{\mathbf{\Omega}}(s) & \hat{\mathbf{\Sigma}}(s) \end{pmatrix},$$

where

$$\hat{\mathbf{\Sigma}}(s) \coloneqq (1-\omega)\mathbf{\Sigma} + k_{\mathrm{t}}\mathbf{Q}\mathbf{A}(s) \tag{C.54}$$

$$\hat{\mathbf{\Omega}}(s) \coloneqq \mathbf{\Omega} + k_t\mathbf{Q}\mathbf{A}(s) \tag{C.55}$$

where $w \in [0, \omega]$, $\omega \in (0, 1)$, and the matrices $\mathbf{Q}$, $\mathbf{\Omega}$, and $\mathbf{\Sigma}$ are defined as in Proposition 5.2. Then, under Assumption 5.6, we have that:

$$\mathbf{V}_w(\tau_c, s) \succeq \nu I_{Nn}, \tag{C.56}$$

for all $\tau_c \in [T_0, T]$, all $w \in [0, \omega]$, and all $s \in \mathbb{R}_{\geq 0}$, where

$$\nu := \frac{(1-\omega)\underline{\sigma}_{\mathbf{\Sigma}}\underline{\sigma}_{\mathbf{Q}} - T^2(\overline{\sigma}^2_{\mathbf{\Omega}} + k_t\chi^2)}{T^2((1-\omega)\underline{\sigma}_{\mathbf{\Sigma}}) + \underline{\sigma}_{\mathbf{Q}}} > 0, \tag{C.57}$$

with $\underline{\sigma}_{\mathbf{Q}}$, $\underline{\sigma}_{\mathbf{\Sigma}}$, and $\overline{\sigma}^2_{\mathbf{\Omega}}$ as defined in Proposition 5.2.

Proof. First, we show that the matrix-valued function $\mathbf{V}_w(\cdot, \cdot)$ is positive-definite uniformly over $\tau_c \in [T_0, T]$, $s \in \mathbb{R}_{\geq}$, and $w \in [0, \omega]$. To do this, we decompose $\mathbf{V}_w$ as follows:

$$\mathbf{V}_w(\tau_c, s) = \mathbf{U}(\tau_c, s)^\top \mathbf{D}(\tau_c, s)\mathbf{U}(\tau_c, s) \tag{C.58}$$

where

$$\mathbf{D}(\tau_c, s) \coloneqq \begin{pmatrix} \frac{\mathbf{Q}}{\tau_c^2} & 0 \\ 0 & \hat{\mathbf{\Sigma}}(s) - \tau_c^2 \hat{\mathbf{\Omega}}(s)^\top \mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s) \end{pmatrix},$$

and

$$\mathbf{U}(\tau_c, s) \coloneqq \begin{pmatrix} I & \tau^2 \mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s) \\ 0 & I \end{pmatrix},$$

Using the definition of $\mathbf{Q}$, and the fact that $\tau_c \leq T$ for all $\tilde{y}_c \in \mathbf{C}_c \cup \mathbf{D}_c$, we obtain

$$\frac{\mathbf{Q}}{\tau^2} \succeq \left(\frac{\underline{\sigma}_{\mathbf{Q}}}{T^2} \right) I_{Nn}. \tag{C.59}$$

Also, it follows that

$$\hat{\mathbf{\Sigma}}(s) - \tau_c^2 \hat{\mathbf{\Omega}}(s)^\top \mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s) \succeq \zeta I_{Nn}, \tag{C.60}$$

for all $s \in \mathbb{R}_{\geq 0}$, where

$$\zeta \coloneqq (1 - \omega) \underline{\sigma}_{\mathbf{\Sigma}} - \frac{T^2}{\underline{\sigma}_{\mathbf{Q}}} (\overline{\sigma}_{\mathbf{\Omega}}^2 + \chi^2(k_t)). \tag{C.61}$$

Note that $\zeta > 0$ since condition (5.56) holds by assumption. Therefore, since

$$\frac{\mathbf{Q}}{\tau_c^2} \succ 0 \quad \text{and} \quad \hat{\mathbf{\Sigma}}(s) - \tau_c^2 \hat{\mathbf{\Omega}}^\top(s) \mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s) \succ 0.$$

it follows that the matrix $\mathbf{V}_w(\tau_c, s)$ is positive definite uniformly over $\tau_c \in [T_0, T]$, $s \in \mathbb{R}_{\geq 0}$, and $w \in [0, \omega]$ [278, Theorem 7.7.7].

Now, we establish the matrix inequality (C.56). To do so, we use the bounds (C.59)

and (C.60) for (C.58) to obtain that

$$\begin{aligned}\mathbf{V}_w(\tau_c, s) &\succeq \mathbf{U}^\top(\tau_c, s) \begin{bmatrix} \frac{\sigma_{\mathbf{Q}}}{T^2} I_{Nn} & 0 \\ 0 & \zeta I_{Nn} \end{bmatrix} \mathbf{U}(\tau_c, s) \\ &= \mathbf{V}(\tau_c, s)^\top \mathbf{V}(\tau_c, s), \end{aligned} \tag{C.62}$$

where $\mathbf{V}(\tau_c, s)$ is the upper block triangular matrix

$$\mathbf{V}(\tau_c, s) \coloneqq \begin{pmatrix} \sqrt{\frac{\sigma_{\mathbf{Q}}}{T^2}} I_{Nn} & \sqrt{\frac{\tau^4 \sigma_{\mathbf{Q}}}{T^2}} \mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s) \\ 0 & \sqrt{\zeta} I_{Nn} \end{pmatrix}.$$

By applying Lemma C.14 on the matrix $\mathbf{V}(\tau_c, s)$, and using (C.62) together with the fact that $\mathbf{V}$ has full column rank and thus that $\sigma_{\min}(\mathbf{V}^\top \mathbf{V}) \geq \sigma_{\min}(\mathbf{V}^\top)\sigma_{\min}(\mathbf{V}) = \sigma^2_{\min}(\mathbf{V})$, we obtain

$$\begin{aligned}\mathbf{V}_w(\tau_c, s) &\succeq \frac{1}{\frac{T^2}{\underline{\sigma}_{\mathbf{Q}}} \left(1 + \frac{\tau^4 \sigma_{\mathbf{Q}}}{\zeta T^2} \|\mathbf{Q}^{-1} \hat{\mathbf{\Omega}}(s)\|^2\right) + \frac{1}{\zeta}} I_{2Nn} \\ &= \frac{\zeta \underline{\sigma}^2_{\mathbf{Q}}}{T^2(\zeta \underline{\sigma}_{\mathbf{Q}} + T^2(\overline{\sigma}^2_{\mathbf{\Omega}} + \chi^2(k_t))) + \underline{\sigma}^2_{\mathbf{Q}}} I_{2Nn} \\ &= \frac{(1-\omega)\underline{\sigma}_{\mathbf{\Sigma}} \underline{\sigma}_{\mathbf{Q}} - T^2(\overline{\sigma}^2_{\mathbf{\Omega}} + \chi^2(k_t))}{T^2((1-\omega)\underline{\sigma}_{\mathbf{\Sigma}}) + \underline{\sigma}_{\mathbf{Q}}} I_{2Nn} \end{aligned}$$

where we have used the fact that the induced 2-norm is sub-multiplicative and that $\|\mathbf{Q}^{-1}\| \leq 1/\underline{\sigma}_{\mathbf{Q}}$ and $\|\mathbf{\Omega}\|^2 \leq \overline{\sigma}^2_{\mathbf{\Omega}}$. This completes the proof. ■

Lemma C.16. Let $\eta := (\eta_1, \eta_2, \ldots, \eta_N)$ with $\eta_i \in \{0, 1\}$ for all $i \in \mathcal{V} = \{1, 2, \ldots, N\}$ and $\mathbf{R}_\eta = \operatorname{diag}(\eta) \otimes I_n$. Then, for all $\tilde{\theta}, \tilde{p} \in \mathbb{R}^{Nn}$ we have:

$$|\mathbf{R}_\eta \tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\, \tilde{p} - \tilde{\theta}|^2_{\mathbf{Q}} + |\mathbf{R}_\eta \tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\, \tilde{p}|^2_{\mathbf{Q}}$$

$$- |\tilde{p}|^2_{\mathbf{Q}} - |\tilde{p} - \tilde{\theta}|^2_{\mathbf{Q}} = |\tilde{\theta}|^2_{\mathbf{R}_\eta \mathbf{Q}} - |\tilde{p}|^2_{\mathbf{R}_\eta \mathbf{Q}} - |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta \mathbf{Q}}$$

where $\mathbf{Q}$ is defined in (5.50).

Proof. By direct computation, we have:

$$\begin{aligned}
\left|\mathbf{R}_\eta \tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\, \tilde{p} - \tilde{\theta}\right|^2_{\mathbf{Q}} &= \left|\mathbf{R}_\eta(\tilde{\theta} - \tilde{p}) - (\tilde{\theta} - \tilde{p})\right|^2_{\mathbf{Q}} \\
&= |\,(I_{Nn} - \mathbf{R}_\eta)\,(\tilde{\theta} - \tilde{p})|^2_{\mathbf{Q}} \\
&= |\mathbf{R}^c_\eta(\tilde{\theta} - \tilde{p})|^2_{\mathbf{Q}} \\
&= |z|^2_{\mathbf{Q}},
\end{aligned}$$

where $z \coloneqq \mathbf{R}^c_\eta(\tilde{\theta} - \tilde{p})$, and $\mathbf{R}^c_\eta \coloneqq I_{Nn} - \mathbf{R}_\eta$. . Writing $z = (z_1, \ldots, z_N)$, with $z_i = (\eta_i - 1)\left(\tilde{\theta}_i - \tilde{p}_i\right) \in \mathbb{R}^n,\ \forall i \in \mathcal{V}$, it follows that

$$\begin{aligned}
|z|^2_{\mathbf{Q}} &= \sum_{i=1}^{N} q_i |\tilde{\theta}_i - \tilde{p}_i|^2 (\eta_i - 1)^2 \\
&= \sum_{i=1}^{N} q_i |\tilde{\theta}_i - \tilde{p}_i|^2 \,(1 - \eta_i)\,.
\end{aligned} \tag{C.63}$$

Similarly,

$$|\mathbf{R}_\eta \tilde{\theta} + (I_{Nn} - \mathbf{R}_\eta)\, \tilde{p}|^2_{\mathbf{Q}} = |\mathbf{R}_\eta(\tilde{\theta} - \tilde{p}) + \tilde{p}|^2_{\mathbf{Q}} = |\tilde{z}|^2_{\mathbf{Q}},$$

where $\tilde{z} = \mathbf{R}_\eta(\tilde{\theta} - \tilde{p}) + \tilde{p}$. Writing $\tilde{z} \coloneqq (\tilde{z}_1, \ldots, \tilde{z}_N)$, with $\tilde{z}_i = \eta_i \left(\tilde{\theta}_i - \tilde{p}_i\right) + \tilde{p}_i \in \mathbb{R}^n,\ \forall i \in \mathcal{V}$, we get:

$$|\tilde{z}|^2_{\mathbf{Q}} = \sum_{i=1}^{N} q_i |\tilde{z}_i|^2$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} q_i |\eta_i(\tilde{\theta}_i - \tilde{p}_i) + \tilde{p}_i|^2 \\
&= \sum_{i=1}^{N} q_i \left(\eta_i^2 |\tilde{\theta}_i - \tilde{p}_i|^2 + 2\eta_i (\tilde{\theta}_i - \tilde{p}_i)^\top (\tilde{p}_i) + |\tilde{p}_i|^2 \right) \\
&= \sum_{i=1}^{N} q_i \left(\eta_i (\tilde{\theta}_i - \tilde{p}_i)^\top (\tilde{\theta}_i + \tilde{p}_i) + |\tilde{p}_i|^2 \right) \\
&= \sum_{i=1}^{N} q_i \eta_i |\tilde{\theta}_i|^2 + \sum_{i=1}^{N} q_i |\tilde{p}_i|^2 (1 - \eta_i). \qquad \text{(C.64)}
\end{aligned}
$$

Together (C.63) and (C.64) yield:

$$
\begin{aligned}
|\mathbf{R}_\eta \tilde{\theta} + \left(I_{Nn} - \mathbf{R}_\eta\right) \tilde{p} - \tilde{\theta}|^2_{\mathbf{Q}} + |\mathbf{R}_\eta \tilde{\theta} + \left(I_{Nn} - \mathbf{R}_\eta\right) \tilde{p}|^2_{\mathbf{Q}} \\
-|\tilde{p}|^2_{\mathbf{Q}} - |\tilde{p} - \tilde{\theta}|^2_{\mathbf{Q}} = \sum_{i=1}^{N} q_i \eta_i |\tilde{\theta}_i|^2 \\
+ \sum_{i=1}^{N} q_i (1 - \eta_i) \left(|\tilde{p}_i|^2 + |\tilde{\theta}_i - \tilde{p}_i|^2 \right) \\
- \sum_{i=1}^{N} q_i |\tilde{p}_i|^2 - \sum_{i=1}^{N} q_i |\tilde{\theta}_i - \tilde{p}_i|^2 \\
= \sum_{i=1}^{N} q_i \eta_i |\tilde{\theta}_i|^2 \\
- \sum_{i=1}^{N} q_i \eta_i \left(|\tilde{p}_i|^2 + |\tilde{\theta}_i - \tilde{p}_i|^2 \right) \\
= |\tilde{\theta}|^2_{\mathbf{R}_\eta \mathbf{Q}} - |\tilde{p}|^2_{\mathbf{R}_\eta \mathbf{Q}} - |\tilde{\theta} - \tilde{p}|^2_{\mathbf{R}_\eta \mathbf{Q}}.
\end{aligned}
$$

■

Appendix D

Proofs of Chapter 6

Proof of Lemma 6.2

Let $x \in \mathbb{R}^n$ be arbitrary. Then, by using that the data is SR with level of richness γ we find that

$$\begin{aligned} x^\top \Omega(s) x &= x^\top \sigma \frac{\phi(s)\phi(s)^\top}{(1+\phi(s)^\top\phi(s))^2} x + x^\top \rho P x \\ &= \sigma \frac{|\phi(s)x|^2}{(1+\phi(s)^\top\phi(s))^2} + \rho x^\top P x \\ &\geq \rho x^\top P x \\ &\geq \gamma\rho |x|^2 \end{aligned}$$

$$\implies \Omega(s) \succeq \kappa I_n \qquad \forall s \in R_{\geq 0}, \tag{D.1}$$

where $\kappa \coloneqq \gamma\rho$. On the other hand, using the fact that $\left|aa^\top\right| = |a|^2,\ \forall a \in \mathbb{R}^n$ we obtain that

$$\left|\Psi(s)\Psi(s)^\top\right| = |\Psi(s)|^2 \leq 1,\ \forall x \in \mathbb{R}^n,$$

and thus, using Hölder's inequality, it follows that:

$$x^\top \Omega(s) x = x^\top \sigma \Psi(s)\Psi(s)^\top x + x^\top \rho P x$$

$$
\begin{aligned}
&\leq \sigma \left| x^\top \Psi(s) \Psi(s)^\top x \right| + \rho \lambda_{\max}(P) |x|^2 \\
&\leq \sigma |x| \left| \Psi(s) \Psi(s)^\top x \right| + \rho \lambda_{\max}(P) |x|^2 \\
&\leq \sigma |x| \left| \Psi(s) \Psi(s)^\top \right| |x| + \rho \lambda_{\max}(P) |x|^2 \\
&\leq \left(\sigma + \rho \lambda_{\max}(P) \right) |x|^2 \\
&\implies \Omega(s) \preceq K I_n,
\end{aligned}
$$

where $K \coloneqq \sigma + \rho \lambda_{\max}(P)$. ■

Proof of Theorem 6.9

We will establish the stability results by using hybrid Lyapunov functions. In particular, we consider the function

$$
V(x, s) = \frac{(p - \hat{\theta})^2}{4} + \frac{(p - \theta^*)^2}{4} + k \rho \tau^2 \frac{\tilde{\theta}^\top P \tilde{\theta}}{2}.
$$

Therefore, for all $x \in C \cup D$, using the fact that $\tau(t, j) \geq T_0$ together with Lemma 6.2, for all $(t, j) \in \operatorname{dom}(x)$, we obtain:

$$
\begin{aligned}
V(x, s) &\geq \frac{\left| p - \hat{\theta} \right|^2}{4} + \frac{|p - \theta^*|^2}{4} + \frac{k \rho T_0^2}{2} \tilde{\theta}^\top P \tilde{\theta} \\
&\geq \frac{\left| p - \hat{\theta} \right|^2}{4} + \frac{|p - \theta^*|^2}{4} + k \rho \gamma T_0^2 \frac{\left| \tilde{\theta} \right|^2}{2} \\
&\geq \frac{\left| p - \hat{\theta} \right|^2}{4} + \min \left\{ \frac{1}{4}, \ \frac{k \rho \gamma T_0^2}{2} \right\} \left(|p - \theta^*|^2 + \left| \tilde{\theta} \right|^2 \right) \\
&\geq \underline{c} |\tilde{x}|_{\mathcal{A}}^2 \qquad \forall s \in \mathbb{R}_{\geq 0},
\end{aligned} \tag{D.2}
$$

where $\underline{c} \coloneqq \min\left\{\frac{1}{4},\ \frac{k\rho\gamma T_0^2}{2}\right\}$. On the other hand, since $\tau(t,j) \leq T$ for all $(t,j) \in \mathrm{dom}(x)$, we obtain that V satisfies:

$$\begin{aligned}
V(x,s) &\leq \frac{|p-\theta|^2}{4} + \frac{|p-\theta^*|^2}{4} + k\rho T^2 \frac{\tilde{\theta}^\top P \tilde{\theta}}{2} \\
&\leq \frac{|p-\theta|^2}{4} + \frac{|p-\theta^*|^2}{4} + k\rho T^2 \lambda_{\max}(P) \frac{\left|\tilde{\theta}\right|^2}{2} \\
&\leq \frac{|p-\theta|^2}{4} + \frac{|p-\theta|^2 + \left|\hat{\theta}-\theta^*\right|^2}{2} + k\rho T^2 \lambda_{\max}(P) \frac{\left|\tilde{\theta}\right|^2}{2} \\
&= \frac{3\left|p-\hat{\theta}\right|^2}{4} + \frac{1}{2}\left(1 + k\rho T^2 \lambda_{\max}(P)\right)\left|\tilde{\theta}\right|^2 \\
&\leq \overline{c}|\tilde{x}|_{\mathcal{A}}^2, \qquad \text{(D.3)}
\end{aligned}$$

where $\overline{c} \coloneqq \max\left\{\frac{3}{4}, \frac{1}{2}\left(1 + k\rho\lambda_{\max}(P)\, T^2\right)\right\}$. Similarly, using

$$|p-\theta^*|^2 \leq 2\left(\left|p-\hat{\theta}\right|^2 + \left|\tilde{\theta}\right|^2\right),$$

we obtain:

$$\begin{aligned}
V(x,s) &\leq \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{1}{2}\left(\left|p-\hat{\theta}\right|^2 + \left|\tilde{\theta}\right|^2\right) + k\rho\lambda_{\max}(P)\, T^2 \frac{\left|\hat{\theta}-\theta^*\right|^2}{2} \\
&= \frac{3}{4}\left|p-\hat{\theta}\right|^2 + \frac{1}{2}\left(1 + k\rho\lambda_{\max}(P)\, T^2\right)\left|\hat{\theta}-\theta^*\right|^2 \\
&\leq \overline{c}\left(\left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2\right) \\
&\implies \left|p-\hat{\theta}\right|^2 + \left|\hat{\theta}-\theta^*\right|^2 \geq \frac{V(x,s)}{\overline{c}}. \qquad \text{(D.4)}
\end{aligned}$$

Additionally, let $\eta \coloneqq 1 - \frac{T_0}{T^2} - \frac{1}{2k\gamma T^2}$, and $\lambda \coloneqq \min\left\{\frac{kT_0(\rho\gamma - \sigma/2)}{2\overline{c}}, -\log(1-\eta)\right\}$. Then, by Assumption 6.2, we have that $\lambda \in \mathbb{R}_{>0}$, since $\frac{kT_0(\rho\gamma-\sigma/2)}{2\overline{c}} > 0$ and $\eta \in (0,1)$.

Hence, by letting $\tilde{x} \coloneqq (x,s)$, the time-derivative of V along the trajectories generated by

the flows of Accelerated Hybrid Dynamics with Periodic Restarting satisfies:

$$
\begin{aligned}
\dot{V}(x,s) &= \frac{\partial V}{\partial \tilde{x}} F(x,s) \\
&= \left(-\frac{p-\theta}{2} + k\rho\tau^2 P\tilde{\theta}\right)^\top \frac{2}{\tau}(p-\theta) \\
&\qquad + \left(\frac{p-\hat{\theta}}{2} + \frac{p-\theta^*}{2}\right)^\top \left(-2k\tau\Omega(s)\tilde{\theta}\right) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} + 2k\tau\tilde{\theta}^\top\rho P(p-\hat{\theta}) \\
&\qquad - k\tau(p-\hat{\theta})^\top\Omega(s)\tilde{\theta} - k\tau(p-\theta^*)^\top\Omega(s)\tilde{\theta} + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} + k\tau\tilde{\theta}^\top\left(2\rho P - \Omega(s)\right)(p-\hat{\theta}) - k\tau\tilde{\theta}^\top\Omega(s)(p-\theta^*) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} + k\tau\tilde{\theta}^\top\left(\rho P - \sigma\Psi(s)\Psi(s)^\top\right)(p-\hat{\theta}) \\
&\qquad - k\tau\tilde{\theta}^\top\left(\sigma\Psi(s)\Psi(s)^\top + \rho P\right)(p-\theta^*) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} + k\tau\tilde{\theta}^\top\left(\rho P - \sigma\Psi(s)\Psi(s)^\top\right)(p-\hat{\theta}+\theta^*-\theta^*) \\
&\qquad - k\tau\tilde{\theta}^\top\left(\sigma\Psi(s)\Psi(s)^\top + \rho P\right)(p-\theta^*) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - k\tau\tilde{\theta}^\top\left(\rho P - \sigma\Psi(s)\Psi(s)^\top\right)\tilde{\theta} + k\tau\tilde{\theta}^\top\left(\rho P - \sigma\Psi(s)\Psi(s)^\top\right)(p-\theta^*) \\
&\qquad - k\tau\tilde{\theta}^\top\left(\sigma\Psi(s)\Psi(s)^\top + \rho P\right)(p-\theta^*) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - k\tau\tilde{\theta}^\top\left(\rho P - \sigma\Psi(s)\Psi(s)^\top\right)\tilde{\theta} - k\tau\tilde{\theta}^\top\sigma\Psi(s)\Psi(s)^\top(p-\theta^*) \\
&\qquad - k\tau\tilde{\theta}^\top\sigma\Psi(s)\Psi(s)^\top(p-\theta^*) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - k\rho\tau\tilde{\theta}^\top P\tilde{\theta} - k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(2p-2\theta^*-\tilde{\theta}) + \frac{k\rho}{2}\tau\tilde{\theta}^\top P\tilde{\theta} \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{k}{2}\rho\tau\tilde{\theta}^\top P\tilde{\theta} - k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(2p-\hat{\theta}-\theta^*)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{k}{2}\rho\tau\tilde{\theta}^\top P\tilde{\theta} - k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(2p+\tilde{\theta}-2\hat{\theta}) \\
&= -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{k}{2}\rho\tau\tilde{\theta}^\top P\tilde{\theta} - k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top\tilde{\theta} - 2k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(p-\hat{\theta}) \\
&\leq -\frac{\left|p-\hat{\theta}\right|^2}{\tau} - \frac{k}{2}\tau\tilde{\theta}^\top\Omega(s)\tilde{\theta} - 2k\sigma\tau\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(p-\hat{\theta}) \\
&= -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\tilde{\theta}^\top\Omega(s)\tilde{\theta} + 2k\sigma\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(p-\hat{\theta})\right) \qquad \text{(D.5)} \\
&\leq -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\tilde{\theta}^\top\Omega(s)\tilde{\theta} - 2k\sigma\left|\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(p-\hat{\theta})\right|\right) \\
&\leq -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\kappa\left|\tilde{\theta}\right|^2 - 2k\sigma\left|\tilde{\theta}^\top\Psi(s)\Psi(s)^\top(p-\hat{\theta})\right|\right) \\
&\leq -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\kappa\left|\tilde{\theta}\right|^2 - 2k\sigma\left|\tilde{\theta}\Psi(s)\right|\left|\Psi(s)^\top(p-\hat{\theta})\right|\right) \\
&\leq -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\kappa\left|\tilde{\theta}\right|^2 - 2k\sigma\left|\tilde{\theta}\right||\Psi(s)|^2\left|p-\hat{\theta}\right|\right) \\
&\leq -\tau\left(\frac{\left|p-\hat{\theta}\right|^2}{\tau^2} + \frac{k}{2}\kappa\left|\tilde{\theta}\right|^2 - \frac{k}{2}\sigma\left|\tilde{\theta}\right|\left|p-\hat{\theta}\right|\right) \\
&= -\frac{k\tau}{2}\left(\frac{2\left|p-\hat{\theta}\right|^2}{k\tau^2} + \kappa\left|\tilde{\theta}\right|^2 - \sigma\left|\tilde{\theta}\right|\left|p-\hat{\theta}\right|\right) \\
&= -\frac{k\tau}{2}\begin{pmatrix}\left|p-\hat{\theta}\right| & \left|\tilde{\theta}\right|\end{pmatrix}\hat{M}(\tau)\begin{pmatrix}\left|p-\hat{\theta}\right| \\ \left|\tilde{\theta}\right|\end{pmatrix} \qquad \text{(D.6)}
\end{aligned}
$$

where

$$
\hat{M} \coloneqq \begin{pmatrix}\frac{2}{k\tau^2} & -\frac{\sigma}{2} \\ -\frac{\sigma}{2} & \kappa\end{pmatrix}. \qquad \text{(D.7)}
$$

Next, we compute the eigenvalues of $\hat{M}$:

$$\begin{aligned}
0 &= \left(r - \frac{2}{k\tau^2}\right)(r-\kappa) - \frac{\sigma^2}{4} \\
&= r^2 - r\left(\frac{2}{k\tau^2} + \kappa\right) - \left(\frac{\sigma^2}{4} - \frac{2\kappa}{k\tau^2}\right) \\
\implies r_{\pm}(\tau) &= \frac{1}{2}\left(\frac{2}{k\tau^2} + \kappa \pm \sqrt{\left(\frac{2}{k\tau^2} + \kappa\right)^2 + \left(\sigma^2 - \frac{8\kappa}{k\tau^2}\right)}\right) \\
&= \frac{1}{2}\left(\frac{2}{k\tau^2} + \kappa \pm \sqrt{\left(\frac{2}{k\tau^2} - \kappa\right)^2 + \sigma^2}\right).
\end{aligned}$$

In order to have that $r_-(\tau) > 0$ we have the following necessary condition

$$\begin{gathered}
\left(\frac{2}{k\tau^2} + \kappa\right)^2 > \left(\frac{2}{k\tau^2} - \kappa\right)^2 + \sigma^2 \\
\iff \\
\frac{8}{k\tau^2}\kappa > \sigma^2 \quad \Longleftarrow \quad T^2 < \frac{8\kappa}{k\sigma^2} = \frac{8\rho\gamma}{k\sigma^2}.
\end{gathered}$$

Now, using the fact that $\sqrt{a^2+b^2} \leq a + b$, it follows that

$$r_-(\tau) \geq \frac{1}{2}\left(\frac{2}{k\tau^2} + \kappa - \left|\frac{2}{k\tau^2} - \kappa\right| - \sigma\right). \tag{D.8}$$

If

$$T^2 < \frac{2}{k}\max\left\{\frac{4\rho\gamma}{\sigma^2}, \frac{1}{\rho\gamma}\right\}, \tag{D.9}$$

from (D.8) we obtain that:

$$r_-(\tau) \geq \kappa - \frac{\sigma}{2}.$$

Moreover, if $2\rho\gamma > \sigma$ holds (D.9), reduces to:

$$T^2 < \frac{2}{k}\max\left\{\frac{4\rho\gamma}{\sigma^2}, \frac{1}{\rho\gamma}\right\} = \frac{2}{k\rho\gamma}\max\left\{\frac{4\rho^2\gamma^2}{\sigma^2}, 1\right\} = \frac{8\rho\gamma}{k\sigma^2},$$

and hence to

$$T^2 < \frac{8\rho\gamma}{k\sigma^2}. \tag{D.10}$$

Hence, provided that (D.10) holds, and defining $\underline{r} \coloneqq \frac{kT_0}{2}\left(\rho\gamma - \frac{\sigma}{2}\right)$ we can further upper bound (D.6) by

$$\begin{aligned}\dot{V} &\leq -\underline{r}\left(\left|p-\hat{\theta}\right|^2 + \left|\tilde{\theta}\right|^2\right) \\ &\leq -\frac{\underline{r}}{\overline{c}}V(x,s)\end{aligned} \tag{D.11}$$

Since by definition we have $\lambda \leq \frac{\underline{r}}{\overline{c}}$, it follows that

$$\dot{V}(x,s) \leq -\lambda V(x,s), \qquad \forall x \in C. \tag{D.12}$$

Note that this bound implies that V does not increase during flows, and that it satisfies $V(x(t,j)) \leq V(x(s,j))$ for all $t > s$ and each fixed j such that $(t,j) \in \mathrm{dom}(x)$. This generates the bound (6.8) (acceleration bound in the theorem).

On the other hand, during jumps the restarting policy $q = 0$ generates changes in the Lyapunov function given by

$$\begin{aligned}V(x^+,s^+) - V(x,s) &= \frac{\left|p-\hat{\theta}\right|^2}{4} + \frac{|p-\theta^*|^2}{4} + \frac{kT_0^2}{2}\tilde{\theta}^\top P\tilde{\theta} \\ &\quad - \frac{\left|p-\hat{\theta}\right|^2}{4} - \frac{|p-\theta^*|^2}{4} - \frac{kT^2}{2}\tilde{\theta}^\top P\tilde{\theta} \\ &= -k\frac{T^2 - T_0^2}{2}\tilde{\theta}^\top P\tilde{\theta}\end{aligned}$$

$$
\begin{aligned}
&\leq -k\gamma \frac{T^2 - T_0^2}{2} \left|\tilde{\theta}\right|^2 \\
&\leq 0, \qquad \text{(D.13)}
\end{aligned}
$$

where in the last step we have used Lemma 6.2. Additionally, when $q = 1$, the change of the Lyapunov function during jumps is given by

$$
\begin{aligned}
V(x^+, s^+) - V(x, s) &= \frac{\left|\tilde{\theta}\right|^2}{4} + \frac{kT_0^2}{2}\tilde{\theta}^\top P \tilde{\theta} - \frac{\left|p - \hat{\theta}\right|^2}{4} - \frac{|p - \theta^*|^2}{4} - \frac{k\tau^2}{2}\tilde{\theta}^\top P \tilde{\theta} \\
&\leq \frac{1}{4\gamma}\tilde{\theta}^\top P \tilde{\theta} + \frac{kT_0^2}{2}\tilde{\theta}^\top P \tilde{\theta} - \frac{\left|p - \hat{\theta}\right|^2}{4} - \frac{|p - \theta^*|^2}{4} - \frac{k\tau^2}{2}\tilde{\theta}^\top P \tilde{\theta} \\
&= -\frac{\left|p - \hat{\theta}\right|^2}{4} - \frac{|p - \theta^*|^2}{4} - \left(1 - \frac{T_0^2}{\tau^2} - \frac{1}{2k\gamma\tau^2}\right)\frac{k\tau^2}{2}\tilde{\theta}^\top P \tilde{\theta} \\
&\leq -\eta V(x, s) \\
&\implies V(x^+, s^+) \leq e^{-\lambda} V(x, s), \qquad \text{(D.14)}
\end{aligned}
$$

for all $x \in D,\ x^+ \in G(x), s \in \mathbb{R}_{\geq 0}$, and where we used $(1 - \eta) \leq e^{-\lambda}$ which follows from the definition of λ. Inequality (D.14) and the non-increment of V during flows implies the bound (6.9) (exponential bound in Theorem). Inequalities (D.11) and (D.14), together with the quadratic bounds on the Lyapunov function (D.2) and (D.3), imply UGES of the set $\mathcal{A}$ ([223, Thm. 1]). Similarly, inequalities (D.11) and (D.13) imply UGAS via the hybrid invariance principle [33, Thm. 8.8], which in turn implies that $V \to 0^+$, i.e., the sequence $\{\beta_j\}_{j=0}^{\infty}$ with $\beta_j \coloneqq V(x(t_j, j), s(t_j, j))$ and $t_j \coloneqq \min\{t \in \mathbb{R}_{\geq 0} : (t, j) \in \mathrm{dom}(x)\}$, is monotonically decreasing and converges to zero. ■

Proof of Theorem 6.2

First, notice that the Hamiltonian function H satisfies:

$$\dot{H}(\hat{\theta}, p) = \nabla H^\top \begin{pmatrix} \dot{\hat{\theta}} \\ \dot{p} \end{pmatrix} = k\rho \left(\frac{\partial H}{\partial \hat{\theta}}\right)^\top \left(\frac{\partial H}{\partial p}\right) - k\rho \left(\frac{\partial H}{\partial p}\right)^\top \left(\frac{\partial H}{\partial \hat{\theta}}\right) = 0, \tag{D.15}$$

where we used $\frac{\partial H}{\partial \hat{\theta}} = B(\hat{\theta})$. To analyze the behaviour of H during jumps, we follow similar arguments as in [167], and we first note that for any pair of vectors $u, v \in \mathbb{R}^n$ we have that:

$$\frac{|u|_P^2}{2} = \frac{|v|_P^2}{2} + v^\top P(u - v) + \frac{1}{2}(u - v)^\top P(u - v).$$

Therefore, since the data is SR with level of richness $\gamma > 0$, we obtain:

$$\frac{|u|_P^2}{2} \geq \frac{|v|_P^2}{2} + v^\top P(u - v) + \frac{\gamma}{2}|u - v|^2. \tag{D.16}$$

After minimizing with respect to u at both sides of (D.16), we find that

$$0 = \min_{u \in \mathbb{R}^n} |u|_P^2 \geq \frac{|v|_P^2}{2} - \frac{1}{2\gamma}|Pv|^2 \implies \frac{|Pv|^2}{2\gamma} \geq \frac{|v|_P^2}{2}. \tag{D.17}$$

Now, let $\varepsilon \coloneqq \frac{1}{\frac{\bar{\lambda}}{\gamma}+1}$, and note that $\varepsilon \in (0, 1)$ since $\bar{\lambda} \geq \gamma$. Hence, using (D.17), we can write an upper bound for the value of the Hamiltonian after a jump as follows:

$$\begin{aligned} H(\hat{\theta}^+, p^+) &= H(\hat{\theta}, 0) \\ &= \frac{|\tilde{\theta}|_P^2}{2} \\ &= (1 - \varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{|\tilde{\theta}|_P^2}{2} \end{aligned}$$

$$\leq (1-\varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{\left|P\tilde{\theta}\right|^2}{2\gamma}$$
$$= (1-\varepsilon)\frac{|\tilde{\theta}|_P^2}{2} + \varepsilon\frac{|B(\hat{\theta})|^2}{2\gamma}. \tag{D.18}$$

Using the fact that

$$1-\varepsilon = 1 - \frac{1}{\frac{\bar{\lambda}}{\gamma}+1} = \frac{\bar{\lambda}}{\gamma}\frac{1}{\frac{1}{\frac{\bar{\lambda}}{\gamma}+1}} = \varepsilon\frac{\bar{\lambda}}{\gamma}, \tag{D.19}$$

from (D.18), we obtain that

$$H(\hat{\theta}^+, p^+) \leq \frac{\varepsilon}{2\gamma}\left(\bar{\lambda}\left|\tilde{\theta}\right|_P^2 + \left|B(\hat{\theta})\right|^2\right),$$

which, using the definition of H, further reduces to

$$H(\hat{\theta}^+, p^+) \leq \frac{\varepsilon}{2\gamma}\left(\bar{\lambda}\left|\tilde{\theta}\right|_P^2 + \bar{\lambda}|p|^2\right)$$
$$= \frac{\varepsilon\bar{\lambda}}{2\gamma}\left(\left|\tilde{\theta}\right|_P^2 + |p|^2\right)$$
$$= \frac{\varepsilon\bar{\lambda}}{\gamma}H(\hat{\theta}, p). \tag{D.20}$$

Therefore, given any arbitrary initial condition $(\hat{\theta}_0, p_0, \tau_0)$, and using (D.19) together with (D.20), after j jumps we obtain

$$H(\hat{\theta}(t,j), p(t,j)) \leq \left(\varepsilon\frac{\bar{\lambda}}{\gamma}\right)^j H(\hat{\theta}_0, p_0) = (1-\varepsilon)^j H(\hat{\theta}_0, p_0) = e^{-jT\alpha}H(\hat{\theta}_0, p_0), \tag{D.21}$$

where $\alpha \coloneqq \frac{1}{T}\ln\left(\frac{1}{1-\varepsilon}\right) = \frac{1}{T}\ln\left(1+\frac{\gamma}{\bar{\lambda}}\right)$. As described in Remark 6.2 we have that $\tau(t,j) < T$ for all $t + j > 0$, and thus $t < (j+1)T \implies e^{-jT\alpha} < e^{-\alpha(t-T)}$. Using this inequality in

(D.21), we obtain

$$H(\hat{\theta}(t,j),p(t,j)) \leq \min\left\{1, e^{-\alpha(t-T)}\right\} H(\hat{\theta}_0, p_0). \tag{D.22}$$

Additionally, by the definition of the Hamiltonian, there exist $\kappa \coloneqq \min\{\gamma/2, 1/2\}$, and $K \coloneqq \max\left\{\lambda_{\max}(P)/2, \frac{1}{2}\right\}$, such that

$$\kappa |x|^2_{\mathcal{A}_H} \leq H(\hat{\theta}, p) \leq K |x|^2_{\mathcal{A}_H}$$

Hence, using (D.22) and (D.15), by [223, Thm. 1] the set $\mathcal{A}_H$ is UGES.

Now, we note that under the Hybrid Concurrent Learning dynamics defined in (6.10), by the definition of the jump map and set, for each initial condition there exists a time $t_1 \leq T$ such that $x(t_1, 1) = (\theta(t_1, 1), 0, 0)$ for $\theta(t_1, 1) \in \mathbb{R}^n$. Thus, given the state value $(\theta(t_1, 1), 0, 0)$ at $(t_1, 1) \in \text{dom}(x)$, and using (D.19) together with (D.20), after $j - 1$ additional jumps we have that

$$H(\hat{\theta}(t,j),p(t,j)), \leq e^{-(j-1)T\alpha} H(\theta(t_1,1),0). \tag{D.23}$$

Moreover, by the definition of the Hamiltonian, we obtain that

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{2}{\gamma} H(\hat{\theta}, p) \qquad (t,j) \in \text{dom}(x), \tag{D.24a}$$

$$H(\hat{\theta}(t_1,1),0) \leq \lambda_{\max}(P) \frac{\left|\tilde{\theta}(t_1,1)\right|^2}{2}. \tag{D.24b}$$

Therefore, using (D.23) together with (D.24), we obtain

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{\lambda_{\max}(P)}{\gamma} \min\left\{1, e^{-\alpha(t-2T)}\right\} \left|\tilde{\theta}(t_1,1)\right|^2, \tag{D.25}$$

where we have used the fact that $e^{-(j-1)T\alpha} < e^{-\alpha(t-2T)}$. Since the flow map in (6.10) is globally Lipschitz, and $t_1 < \infty$, there exists $c_1 \in \mathbb{R}_{>0}$ such that

$$\left|\tilde{\theta}(t,0)\right| \leq c_1 \left|\tilde{\theta}(0,0)\right| \qquad \forall t \leq t_1. \tag{D.26}$$

Since $\tilde{\theta}(t,1) = \tilde{\theta}(t,0)$, from (D.25) and (D.26), we obtain that

$$\left|\tilde{\theta}(t,j)\right|^2 \leq \frac{\lambda_{\max}(P)c_1}{\gamma} \min\left\{1, e^{-\alpha(t-\tilde{T})}\right\} \left|\tilde{\theta}(0,0)\right|^2,$$

which is equivalent to (6.11) with $c_0 \coloneqq \frac{\lambda_{\max}(P)c_1}{2}$ and $\tilde{T} = 2T$. This concludes the proof.■

Proof of Lemma 6.3

a) By the definition of $B(\hat{\theta})$ we have that $B(\theta^*) = 0$. Now, we prove that $B(\hat{\theta}) = 0 \implies \hat{\theta} = \theta^*$. Moreover, note that we can write

$$B(\hat{\theta}) = P\tilde{\theta}. \tag{D.27}$$

Since the data is SR with level of richness $\gamma > 0$, meaning that $P \succeq \gamma I_n$, we obtain that $\ker(P) = \{0_n\}$. Consequently, from (D.27), we have that

$$B(\hat{\theta}) = 0 \implies \tilde{\theta} = 0_n \implies \hat{\theta} = \theta^*.$$

This concludes the proof.

b) The fact that the vector field vanishes at θ^* follows directly by construction. On the other hand, from the item above $B(\hat{\theta}) = 0 \iff \hat{\theta} = \theta^*$ whenever the data is SR, and

thus, that the the following implication holds:

$$0 = -k\left(\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})\right)\frac{1}{|B(\hat{\theta})|^{\frac{1}{2}}} \implies \sigma A(s,\hat{\theta}) + \rho B(\hat{\theta}) = 0. \tag{D.28}$$

Rewriting (D.28) by using the fact that $\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta}) = \Omega(s)\tilde{\theta}$

$$0 = \Omega(s)\tilde{\theta}. \tag{D.29}$$

Therefore, using the fact that $\Omega(s)$ is positive definite via Lemma 6.2, it follows from (D.29) that $\tilde{\theta} = 0 \implies \hat{\theta} = \theta^*$. This concludes the proof.

■

Proof of Lemma 6.4

a) We follow similar arguments as in [185]. First, by continuity of the ℓ_2-norm we note that

$$\begin{aligned}
\left|\lim_{\hat{\theta}\to\theta^*}\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right| &= \lim_{\hat{\theta}\to\theta^*}\left|\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right|, \\
&= \lim_{\hat{\theta}\to\theta^*}\frac{1}{\left|B(\hat{\theta})\right|^a}\left|B(\hat{\theta})\right|, \\
&= \lim_{\hat{\theta}\to\theta^*}\left|B(\hat{\theta})\right|^{1-a}, \\
&= \left|\lim_{\hat{\theta}\to\theta^*}B(\hat{\theta})\right|^{1-a} = 0, \qquad \forall a \in (0,1).
\end{aligned}$$

Hence, by positive-definiteness of the norm, we obtain that

$$\lim_{\hat{\theta}\to\theta^*}\frac{B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0, \qquad \forall a \in (0,1). \tag{D.30}$$

On the other hand, by assumption, we have that ϕ is uniformly bounded, and thus, there exists $M \in \mathbb{R}_{\geq 0}$ such that $\left|\Psi(s)\phi(s)^\top\right| \leq M$. With this in mind, using (D.27) and rewriting A in terms of its definition we find

$$\left|B(\hat{\theta})\right|^a = \left|P\tilde{\theta}\right|^a \geq \gamma^a \left|\tilde{\theta}\right|^a, \tag{D.31}$$

$$A(s,\hat{\theta}) = \frac{\phi(s)}{\left(1+\phi(s)^\top\phi(s)\right)^2}\left(\phi(s)^\top\hat{\theta} - y(s)\right) = \Psi(s)\Psi(s)^\top\tilde{\theta} \implies \left|A(s,\hat{\theta})\right| \leq \left|\tilde{\theta}\right|, \tag{D.32}$$

where we have used the fact that $\left|aa^\top\right| = |a|^2$ for all $a \in \mathbb{R}^n$ together with $|\Psi(s)| \leq 1$. Hence, using (D.32) and (D.31), we have that

$$0 \leq \frac{\left|A(s,\hat{\theta})\right|}{\left|B(\hat{\theta})\right|^a} \leq \frac{1}{\gamma^a}\left|\hat{\theta} - \theta^*\right|^{1-a} \qquad \forall \hat{\theta} \in \mathbb{R}^n,\ a \in (0,1). \tag{D.33}$$

Moreover, using continuity of the ℓ_2-norm together with (D.33), we obtain that

$$\left|\lim_{\hat{\theta}\to\theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a}\right| = \lim_{\hat{\theta}\to\theta^*} \frac{\left|A(s,\hat{\theta})\right|}{\left|B(\hat{\theta})\right|^a} \leq \lim_{\hat{\theta}\to\theta^*} \frac{1}{\gamma^a}\left|\hat{\theta} - \theta^*\right|^{1-a} = \frac{1}{\gamma^a}\lim_{\hat{\theta}-\theta^*}\left|\hat{\theta} - \theta^*\right|^{1-a} = 0,$$

for all $a \in (0,1)$. Therefore, by positive-definitness of the norm, we obtain

$$\lim_{\hat{\theta}\to\theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \qquad \forall a \in (0,1). \tag{D.34}$$

This concludes the proof.

b) Continuity for every $\hat{\theta} \in \mathbb{R}^n$ such that $B(\hat{\theta}) \neq 0$, follows directly from the continuity of A and B. Hence, in order to show continuity we only need to analyze the cases where B vanishes. As shown by Lemma 6.3, B vanishes only when $\hat{\theta} = \theta^*$. Hence, using Lemma

6.4, with $a = \frac{1}{2}$, we have that

$$\lim_{\hat{\theta}\to\theta^*} -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}} = -k\sigma \lim_{\hat{\theta}\to\theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}} - k\rho \lim_{\hat{\theta}\to\theta^*} \frac{B(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}} = 0, \tag{D.35}$$

where in the second to last step we have used the continuity of A and B, together with the fact that A and B vanish at $\hat{\theta} = \theta^*$. Equation (D.35), together with the fact that (6.12) is defined to be 0 when $\hat{\theta} = \theta^*$, shows that the vector field describing the Finite-Time Concurrent Learning dynamics is continuous at θ^*, and hence, by the aforementioned arguments, that it is continuous everywhere. ■

Proof of Theorem 6.3

We consider the Lyapunov function $V(\hat{\theta}) = \frac{1}{2}\left|\tilde{\theta}\right|^2$, whose time-derivative along the trajectories of the Finite-Time Concurrent Learning dynamics satisfies:

$$\begin{aligned}
\dot{V}(\hat{\theta}) &= \nabla V(\hat{\theta})^\top \dot{\hat{\theta}} \\
&= -k\tilde{\theta}^\top \frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{|B(\hat{\theta})|^{\frac{1}{2}}} \\
&= -k\tilde{\theta}^\top \frac{\Omega(s)}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}}\tilde{\theta} \\
&\leq -k\kappa \frac{\left|\tilde{\theta}\right|^2}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}},
\end{aligned} \tag{D.36}$$

where in the last step we have used (D.1). Now, note that

$$\begin{aligned}
\left|B(\hat{\theta})\right| = \left(\tilde{\theta}^\top P^\top P \tilde{\theta}\right)^{\frac{1}{2}} \leq \left(\lambda_{\max}^2(P)\left|\tilde{\theta}\right|^2\right)^{\frac{1}{2}} = \lambda_{\max}(P)\left|\tilde{\theta}\right| \\
\implies \frac{1}{\left|B(\hat{\theta})\right|^{\frac{1}{2}}} \geq \frac{1}{\lambda_{\max}^{\frac{1}{2}}(P)\left|\tilde{\theta}\right|^{\frac{1}{2}}},
\end{aligned} \tag{D.37}$$

and thus, from (D.36) we obtain that

$$\begin{aligned}\dot{V}(\hat{\theta}) &\leq -\frac{k\kappa}{\lambda_{\max}^{\frac{1}{2}}(P)}\frac{\left|\tilde{\theta}\right|^2}{\left|\tilde{\theta}\right|^{\frac{1}{2}}} \\ &= -\frac{k\kappa}{\lambda_{\max}^{\frac{1}{2}}(P)}\left|\tilde{\theta}\right|^{\frac{3}{2}} \\ &= -\frac{k\kappa}{\lambda_{\max}^{\frac{1}{2}}(P)}\left(2V(\hat{\theta})\right)^{\frac{3}{4}} \\ &= -\frac{2^{3/4}k\kappa}{\lambda_{\max}^{\frac{1}{2}}(P)}V(\hat{\theta})^{\frac{3}{4}} \\ &= -\frac{2^{3/4}k\gamma\rho}{\lambda_{\max}^{\frac{1}{2}}(P)}V(\hat{\theta})^{\frac{3}{4}}\end{aligned}$$

Hence, it follows by [279, Thm. 4.2] that the set $\mathcal{A}_0$ is UGFTS with settling time function $T : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ bounded as

$$T(\hat{\theta}(0)) \leq \frac{2}{k\gamma\rho}\lambda_{\max}^{1/2}(P)\sqrt{|\hat{\theta}(0) - \theta^*|}. \tag{D.38}$$

This concludes the proof. ■

Proof of Lemma 6.5

Continuity for every $\hat{\theta} \in \mathbb{R}^n$ such that $B(\hat{\theta}) \neq 0$, follows directly from the continuity of A and B. Hence, in order to show continuity we only need to analyze the cases where B vanishes. As shown by Lemma 6.3, B vanishes only when $\hat{\theta} = \theta^*$. Hence, using Lemma 6.4, for all $a \in (0, 1)$ we have that

$$\lim_{\hat{\theta}\to\theta^*} -k\frac{\sigma A(s,\hat{\theta}) + \rho B(\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = -k\sigma \lim_{\hat{\theta}\to\theta^*} \frac{A(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} - k\rho \lim_{\hat{\theta}\to\theta^*} \frac{B(s,\hat{\theta})}{\left|B(\hat{\theta})\right|^a} = 0 \tag{D.39}$$

$$\lim_{\hat{\theta}\to\theta^*} -k\frac{\sigma A(s,\hat{\theta})+\rho B(\hat{\theta})}{\left|B(\hat{\theta})\right|^{-a}} = \lim_{\hat{\theta}\to\theta^*} -k\left(\sigma A(s,\hat{\theta})+\rho B(\hat{\theta})\right)\left|B(\hat{\theta})\right|^{a}$$
$$= -k\left(\sigma A(s,\theta^*)+\rho B(\theta^*)\right)|B(\theta^*)|^{-a}$$
$$= 0, \tag{D.40}$$

where in the second to last step we have used the continuity of A and B, together with the fact that A and B vanish at $\hat{\theta}=\theta^*$. Equations (D.39) and (D.40), and the fact that (6.13) is defined to be 0 when $\hat{\theta}=\theta^*$, show that the vector field describing the Fixed-Time Concurrent Learning dynamics is continuous at θ^*, and hence, by the previous arguments, that it is continuous everywhere. ■

Proof of Theorem 6.4

We consider the Lyapunov function $V(\hat{\theta})=\frac{\tilde{V}(\hat{\theta})^2}{2}$, where $\tilde{V}$ is an auxiliary function defined as $\tilde{V}(\hat{\theta}) \coloneqq \frac{1}{2}\left|\tilde{\theta}\right|^2$. Hence, the time-derivative of V along the trajectories of the Fixed-Time Concurrent Learning dynamics satisfies

$$\begin{aligned}
\dot{V}(\hat{\theta}) &= \tilde{V}(\hat{\theta})\dot{\tilde{V}}(\hat{\theta}) \\
&= -k\tilde{V}(\hat{\theta})\tilde{\theta}^\top\left(\frac{\sigma A(s,\hat{\theta})+\rho B(\hat{\theta})}{|B(\hat{\theta})|^a}+\frac{\sigma A(s,\hat{\theta})+\rho B(\hat{\theta})}{|B(\hat{\theta})|^{-a}}\right) \\
&= -k\tilde{V}(\hat{\theta})\tilde{\theta}^\top\left(\frac{\Omega(s)\tilde{\theta}}{|B(\hat{\theta})|^a}+\frac{\Omega(s)\tilde{\theta}}{|B(\hat{\theta})|^{-a}}\right) \\
&= -k\tilde{V}(\hat{\theta})\tilde{\theta}^\top\Omega(s)\tilde{\theta}\left(\frac{1}{|B(\hat{\theta})|^a}+\frac{1}{|B(\hat{\theta})|^{-a}}\right) \\
&\leq -k\kappa\tilde{V}(\hat{\theta})\left|\tilde{\theta}\right|^2\left(\frac{1}{|B(\hat{\theta})|^a}+\frac{1}{|B(\hat{\theta})|^{-a}}\right),
\end{aligned} \tag{D.41}$$

where in the last step we have used (D.1). Now, note that

$$\left|B(\hat{\theta})\right| \leq \lambda_{\max}(P)\left|\tilde{\theta}\right| \implies \frac{1}{\left|B(\hat{\theta})\right|^a} \geq \frac{1}{\lambda_{\max}^a(P)\left|\tilde{\theta}\right|^a},$$
$$\left|B(\hat{\theta})\right|^a = \left|\tilde{\theta}^\top P^\top P\tilde{\theta}\right|^{\frac{\alpha}{2}} \geq \gamma^a\left|\tilde{\theta}\right|^a,$$

and thus, from (D.41), we obtain that

$$\begin{aligned}
\dot{V}(\hat{\theta}) &\leq -k\kappa\tilde{V}(\hat{\theta})\left|\tilde{\theta}\right|^2\left(\frac{1}{\lambda_{\max}^a(P)}\frac{1}{\left|\tilde{\theta}\right|^a} + \frac{\gamma^a}{\left|\tilde{\theta}\right|^{-a}}\right) \\
&= -k\kappa\tilde{V}(\hat{\theta})\left(\frac{1}{\lambda_{\max}^a(P)}\left|\tilde{\theta}\right|^{2-a} + \gamma^a\left|\tilde{\theta}\right|^{2+a}\right) \\
&= -k\kappa\tilde{V}(\hat{\theta})\left(\frac{1}{\lambda_{\max}^a(P)}\left(2\tilde{V}(\hat{\theta})\right)^{1-\frac{a}{2}} + \gamma^a\left(2\tilde{V}(\hat{\theta})\right)^{1+\frac{a}{2}}\right) \\
&= -\frac{2^{1-\frac{a}{2}}k\kappa}{\lambda_{\max}^a(P)}\tilde{V}(\hat{\theta})^{2-\frac{a}{2}} - 2^{1+\frac{a}{2}}k\kappa\gamma^a\tilde{V}(\hat{\theta})^{2+\frac{a}{2}} \\
&= -\frac{2^{2-3\frac{a}{4}}k\kappa}{\lambda_{\max}^a(P)}V(\hat{\theta})^{1-\frac{a}{4}} - 2^{2+3\frac{a}{4}}k\kappa\gamma^aV(\hat{\theta})^{1+\frac{a}{4}} \\
&= -\frac{2^{2-3\frac{a}{4}}k\gamma\rho}{\lambda_{\max}^a(P)}V(\hat{\theta})^{1-\frac{a}{4}} - 2^{2+3\frac{a}{4}}k\rho\gamma^{a+1}V(\hat{\theta})^{1+\frac{a}{4}}.
\end{aligned}$$

The last inequality implies UGFXS via [203, Lemma 1]. Moreover, by [280, Lemma 2], a sharp bound T^* on the settling time function T can be computed as

$$T^* = \frac{\pi}{2a\gamma\rho k}\sqrt{\frac{\lambda_{\max}^a(P)}{\gamma^a}}. \tag{D.42}$$

This concludes the proof. ■

Appendix E

Proofs of Chapter 7

Proof of Theorem 7.1

The result of Theorem 7.1 can be established by showing that all the assumptions needed to apply [111, Thm.1] are satisfied in a neighborhood of the optimal incentive. First, note that, by Assumption 7.1, the plant has a well-defined steady state input-to-output map $\tilde{\varphi}$. Also, under Assumption 7.2, this response map is strictly convex, and thus has a unique minimizer. Since under Assumption 7.4 the average dynamics of (7.11) can be computed to be $\dot{\hat{u}} = -k\nabla\tilde{\varphi}(u) + \mathcal{O}(\varepsilon_a)$ (see, e.g., [25, Sec. 7]), it follows that for ε_a sufficiently small, in a neighborhood of the optimal incentive u^* the average dynamics converge to a neighborhood of u^*. By averaging theory and the results of [111, Thm.1], the original system retains the stability properties in a practical sense. The result follows by using a (local) singular perturbation argument to interconnect the dynamics (7.11) with the dynamics (7.5). ■

Proof of Theorem 7.2

We prove Theorem 7.2 following a similar approach as in Theorem 7.1. In particular, first note that the hybrid dynamics (7.12) are well-posed in the sense of [33, Sec. 6] because the sets C_2 and D_2 are closed, and the maps F_2 and G_2 are continuous on these sets. Moreover, neglecting the socio-technical dynamics, and using Assumption 7.4, the average

dynamics of (7.12) correspond to an $\mathcal{O}(\varepsilon_a)$-perturbed version of the hybrid Nesterov gradient dynamics studied in [25] for the model-free optimization of static maps. Under Assumptions 7.1, 7.2 or 7.3, and 7.4, these average hybrid dynamics render the set $\{(\hat{u}, p, \tau) : \ \hat{u} = p = \text{argmin } \tilde{\varphi}(u), \tau \in [T_0, T]\}$ (locally) practically asymptotically stable. By using, sequentially, averaging and singular perturbation theory for perturbed hybrid systems [25, Thm. 7], we obtain the desired result for the interconnection between the controller and the socio-technical dynamics, which are stable under Assumption 7.1. ■

Proof of Theorem 7.3

Neglecting the socio-technical dynamics, the average dynamics of (7.13) are given by a perturbed version of the fixed-time gradient flows studied in [185]. Under Assumption 7.3, these dynamics render the optimal incentive fixed-time stable. A direct application of averaging theory for non-smooth systems [180] allows us to conclude practical (with respect to $\tilde{\Lambda}_u$) fixed-time stability for the ISC interconnected with the socio-technical dynamics (7.1). ■

Appendix F

Proofs of Chapter 8

Before presenting the proofs of our main results we start by establishing the following auxiliary Lemma.

Lemma F.1. Under the conditons of Lemma 8.1, for every $z_0 = (x_0, \mu_0) \in (\overline{C_x} \cup D_x) \times \mathcal{X}_\mu$, it follows that

a) $\hat{z} \in \mathcal{S}_{\hat{\mathcal{H}}}(z_0) \implies \hat{z} \circ \mathbb{D}_{\mu_0} \in \mathcal{S}_{\mathcal{H}}(z_0)$ with $\mathrm{dom}(\hat{z} \circ \mathbb{D}_{\mu_0}) = \mathbb{D}_{\mu_0}^{-1}(\mathrm{dom}(\hat{z}))$, and

b) $z \in \mathcal{S}_{\mathcal{H}}(z_0) \implies z \circ \mathbb{D}_{\mu_0}^{-1} \in \mathcal{S}_{\hat{\mathcal{H}}}(z_0)$ with $\mathrm{dom}(z \circ \mathbb{D}_{\mu_0}^{-1}) = \mathbb{D}_{\mu_0}(\mathrm{dom}(z))$,

where $\mathbb{D}_{\mu_0}^{-1} \coloneqq \mathcal{D}_{\mu_0}^{-1} \times \mathrm{id}_{\mathbb{Z}_{\geq 0}}$ for all $\mu_0 \in \mathcal{X}_\mu$. □

Proof: (a) Let $z_0 \coloneqq (x_0, \mu_0) \in (\overline{C_x} \cup D_x) \times \mathcal{X}_\mu$, and $\hat{z}$ be a maximal solution to the HDS $\hat{\mathcal{H}}$ from z_0. Then, by the definition of solutions to hybrid dynamical systems, for each $j \in \mathbb{Z}_{\geq 0}$ such that the interior of $\hat{I}_j := \{s \geq 0 : (s, j) \in \mathrm{dom}(\hat{z})\}$ is nonempty, $\hat{z}$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s,j) \in F_x(\hat{x}(s,j)) \times \left\{\frac{1}{\hat{\mu}(s,j)} F_\mu(\hat{\mu}(s,j))\right\}, \tag{F.1}$$

for almost all $s \in \hat{I}_j$. Now, consider the hybrid signal $z \coloneqq \hat{z} \circ \mathbb{D}_{\mu_0}$. Then, using the chain rule, the signal z satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} z(t,j) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\hat{z} \circ \mathbb{D}_{\mu_0}\right)(t,j)$$

$$= \frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s,j)\Big|_{s=\mathcal{D}_{\mu_0}(t)} \frac{\mathrm{d}\mathcal{D}_{\mu_0}}{\mathrm{d}t}(t), \tag{F.2}$$

for almost all $t \in \mathcal{D}_{\mu_0}^{-1}\left(\hat{I}_j\right) =: I_j$. Since $\frac{\mathrm{d}\mathcal{D}_{\mu_0}}{\mathrm{d}t}(t) = (\hat{u} \circ \mathcal{D}_{\mu_0})(t)$ for all $t \in \mathcal{T}_{\mu_0}$ by assumption. Noting that $\hat{\mu}$ does not change during the jumps of (8.7), and using the fact that $\sup_t \mathrm{dom}(\hat{z}) = \infty$, it follows that $\mu(t,j) = (\hat{\mu} \circ \mathbb{D}_{\mu_0})(t,j) = (\hat{u} \circ \mathcal{D}_{\mu_0})(t)$ for all $t \in I_j$. Therefore, using (F.1) and (F.2) we obtain that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}z(t,j) &= \mu(t,j)\, F_x\left((\hat{x} \circ \mathbb{D}_{\mu_0})(t,j)\right) \times F_\mu\left((\hat{\mu} \circ \mathbb{D}_{\mu_0})(t,j)\right) \\ &= \mu(t,j)\, F_x\left(x(t,j)\right) \times F_\mu\left(\mu(t,j)\right), \end{aligned} \tag{F.3}$$

for almost all $t \in I_j$. Equation (F.3) implies that z satisfies the continuous-time dynamics of the HDS $\mathcal{H}$ for almost all $t \in I_j$. Moreover, note that by the definition of I_j, we directly get that $\mathcal{D}_{\mu_0}(\underline{t}_j) = \underline{s}_j$ and $\mathcal{D}_{\mu_0}(\overline{t}_j) = \overline{s}_j$ where $\underline{t}_j := \min I_j$, $\overline{t}_j = \sup I_j$, $\underline{s}_j := \min \hat{I}_j$, $\overline{s}_j = \sup \hat{I}_j$. Similarly, we have that $\hat{x}(s, j+1) \in G(\hat{x}(s,j))$ for every $(s,j) \in \mathrm{dom}(\hat{x})$ such that $(s, j+1) \in \mathrm{dom}(\hat{x})$, and therefore $x(t, j+1) \in G(x(t,j))$ since the HDSs $\hat{\mathcal{H}}$ and the HDS $\mathcal{H}$ have the same discrete-time dynamics. Thus z is a maximal solution to $\mathcal{H}$. Using these arguments, we also obtain that

$$\begin{aligned}\mathrm{dom}\left(\hat{z} \circ \mathbb{D}_{\mu_0}\right) &= \bigcup_{j=0}^{\sup_j \mathrm{dom}(z)} I_j \times \{j\} \\ &= \bigcup_{j=0}^{\sup_j \mathrm{dom}(\hat{z})} \mathbb{D}_{\mu_0}^{-1}\left(\hat{I}_j \times \{j\}\right) = \mathbb{D}_{\mu_0}^{-1}\left(\mathrm{dom}(\hat{z})\right). \end{aligned}$$

(b) Follows by the same ideas of the proof of item (a), via the use of the inverse mapping $\mathbb{D}_c^{-1}$. ■

Proof of Lemma 8.1

A first consequence of Lemma F.1 is that it allows us to certify that the map $\mathcal{W}$ introduced in (8.9) is well-defined. Indeed, let $\hat{z} \in \mathcal{S}_{\hat{\mathcal{H}}}$ be arbitrary. Then, by the definition of solutions to hybrid dynamical systems, there exists $z_0 = (x_0, \mu_0) \in (\overline{C}_x \cup D_x) \times \mathcal{X}_\mu$ such that $\hat{z}(0,0) = z_0$. Thus, using Lemma F.1, we obtain that $\mathcal{W}(\hat{z}) = \hat{z} \circ \mathbb{D}_{\mu_0} \in \mathcal{S}_{\mathcal{H}}(z_0) \subset \mathcal{S}_{\mathcal{H}}$, meaning that $\mathcal{W}$ indeed maps $\mathcal{S}_{\hat{\mathcal{H}}}$ into $\mathcal{S}_{\mathcal{H}}$. Using these facts, we are now prepared to present the proofs of our main results.

Now, to prove that $\mathcal{W}$ is a bijection it suffices to establish the existence of both a left and a right inverse. For this purpose, consider the following map:

$$\hat{\mathcal{W}} : \quad \mathcal{S}_{\mathcal{H}} \quad \longrightarrow \quad \mathcal{S}_{\hat{\mathcal{H}}} \tag{F.4a}$$

$$z = (x, \mu) \longmapsto z \circ \mathbb{D}^{-1}_{\mu(0,0)}, \tag{F.4b}$$

which is well-defined by the results of Lemma F.1. Then, for every $z \in \mathcal{S}_{\mathcal{H}}$, we have that

$$\begin{aligned}
\left(\mathcal{W} \circ \hat{\mathcal{W}}\right)(z) &= \mathcal{W}\left(\hat{\mathcal{W}}(z)\right) \\
&= \mathcal{W}\left(z \circ \mathbb{D}^{-1}_{\mu(0,0)}\right) \\
&= z \circ \mathbb{D}^{-1}_{\mu(0,0)} \circ \mathbb{D}_{\mu(0,0)} \\
&= z \circ \mathrm{id}_{\mathbb{R}_{\geq 0} \times \mathbb{Z}_{\geq 0}} \\
&= z \implies \mathcal{W} \circ \hat{\mathcal{W}} = \mathrm{id}_{\mathcal{S}_{\mathcal{H}}}.
\end{aligned}$$

Similarly, it follows that $\hat{\mathcal{W}} \circ \mathcal{W} = \mathrm{id}_{\mathcal{S}_{\hat{\mathcal{H}}}}$. Hence, we have that $\mathcal{W}^{-1} = \hat{\mathcal{W}}$, and thus that $\mathcal{W}$ is a bijection between $\mathcal{S}_{\mathcal{H}}$ and $\mathcal{S}_{\hat{\mathcal{H}}}$. The fact that $\mathrm{dom}\left(\mathcal{W}(\hat{z})\right) = \mathbb{D}^{-1}_{\hat{\mu}(0,0)}(\mathrm{dom}(\hat{z}))$ follows directly from the results of Lemma F.1. ■

Proof of Theorem 8.1

Let $z_0 = (x_0, \mu_0) \in (\overline{C}_x \cup D_x) \times \mathcal{X}_\mu$ and $z \in \mathcal{S}_{\mathcal{H}}(z_0)$. Then, by Lemma 8.1 it follows that there exists $\hat{z} \in \mathcal{S}_{\hat{\mathcal{H}}}$ such that $z = \mathcal{W}(\hat{z})$, where $\hat{\mathcal{W}}$ is the inverse of $\mathcal{W}$ as defined in (F.4). Therefore, we obtain that

$$\begin{aligned} |z(t,j)|_{\mathcal{A}_0} &= |\mathcal{W}(\hat{z})(t,j)|_{\mathcal{A}_0} \\ &= \left|\left(\hat{z} \circ \mathbb{D}_{\hat{\mu}(0,0)}\right)(t,j)\right|_{\mathcal{A}_0} \\ &= \left|\hat{z}\left(\mathcal{D}_{\hat{\mu}(0,0)}(t),\ j\right)\right|_{\mathcal{A}_0} \\ &\leq \beta\left(|\hat{z}(0,0)|_{\mathcal{A}_0},\ \mathcal{D}_{\hat{\mu}(0,0)}(t),\ j\right), \end{aligned} \tag{F.5}$$

for all $(t,j) \in \mathrm{dom}(z)$, where in the last step we used that $\mathcal{A}_0$ is β-UGAS for $\hat{\mathcal{H}}$. Additionally, it follows that

$$\begin{aligned} z_0 = z(0,0) = \mathcal{W}(\hat{z})(0,0) &= (\hat{z} \circ \mathbb{D}_{\hat{\mu}(0,0)})(0,0) \\ &= \hat{z}(\mathcal{D}_{\hat{\mu}(0,0)}(0), 0) \\ &= \hat{z}(0,0), \end{aligned} \tag{F.6}$$

where we used that $\mathcal{D}_{\hat{\mu}(0,0)}(0) = 0$ by Assumption 8.1. Now, by the definition of solutions to hybrid dynamical systems we necessarily have that $\hat{z}(0,0) = (\hat{x}(0,0), \hat{\mu}(0,0)) \in (\overline{C}_x \cup D_x) \times \mathcal{X}_\mu$. In particular, since $\mathcal{A}_0 = \mathcal{A} \times \mathcal{X}_\mu$, this implies that $|\hat{z}(0,0)|_{\mathcal{A}_0} = |\hat{x}(0,0)|_{\mathcal{A}} = |x(0,0)|_{\mathcal{A}} = |x_0|_{\mathcal{A}}$, where the last equality follows by (F.6). By replacing this expression in (F.5), and using the fact that $\hat{\mu}(0,0) = \mu(0,0) = \mu_0$ by (F.6), we obtain the following bound:

$$|z(t,j)|_{\mathcal{A}_0} \leq \beta(|x_0|_{\mathcal{A}}, \mathcal{D}_{\mu_0}(t), j), \tag{F.7}$$

for all $(t, j) \in \mathrm{dom}(z)$. Since $z = (x, \mu) \in \mathcal{S}_{\mathcal{H}}(z_0)$, by the definition of solutions to hybrid dynamical systems, and the construction of the jump and flow sets of the HDS $\mathcal{H}$, it follows that $\mu(t, j) \in \mathcal{X}_\mu$ for all $(t, j) \in \mathrm{dom}(z)$. Then, using this fact in (F.7) finally yields

$$|x(t, j)|_{\mathcal{A}} \leq \beta(|x_0|_{\mathcal{A}}, \mathcal{D}_{\mu_0}(t), j) \tag{F.8}$$

for all $(t, j) \in \mathrm{dom}(z)$. ■

Proof of Proposition 8.1

The bound in (8.14) follows directly by the results of Theorem 8.1, and leveraging the fact that $\mathcal{A}_0$ is β_g-UGAS for $\hat{\mathcal{H}}_g$, as discussed above the statement of Proposition 8.1.

To obtain Inequality (8.15), let $z_0 = (x_0, \mu_0) \in \mathbb{R}^n \times \mathcal{X}_\mu$, and $z = (x, \mu) \in \mathcal{S}_{\mathcal{H}_g}(z_0)$ be arbitrary. By Lemma 8.1, it follows that there exists $\hat{z} = (\hat{x}, \hat{\mu}) \in \mathcal{S}_{\hat{\mathcal{H}}_g}$ such that $z = \mathcal{W}(\hat{z})$. Using this fact, and letting Π_x be the natural projection that selects the $\tilde{x}$-coordinates from a point $(\tilde{x}, \tilde{\mu}) \in \mathbb{R}^n \times \mathcal{X}_\mu$, we obtain the following:

$$\begin{aligned} f(x(t)) - f^\star &\leq f(\Pi_x(\mathcal{W}(\hat{z}))) - f^\star \\ &= f(\hat{x} \circ \mathcal{D}_{\mu_0}(t)) - f^\star \\ &\leq \frac{V_0}{\mathcal{D}_{\mu_0}(t)}, \end{aligned}$$

for all $t \in \mathrm{dom}(z)$, where in the last step we have employed bound (8.13). This obtains the result. ■

Proof of Proposition 8.2

Using [24, Theorem 3.1.a)], the fact that for any solution $\hat{z} = (\hat{x}, \hat{\mu}) \in \mathcal{S}_{\hat{\mathcal{H}}_m}$ we have $|\hat{\mu}(s, j)|_{\mathcal{X}_\mu} = 0$ for all $(s, j) \in \mathrm{dom}(\hat{z})$, and since the $\hat{x}$ dynamics are uncoupled from the $\hat{\mu}$ dynamics, it follows that there exists $\beta_m \in \mathcal{KLL}$ such that $\mathcal{A}_0$ is β_m-UGAS for $\hat{\mathcal{H}}_m$.

Also by using the completeness of solutions result of [24, Theorem 3.1.a)], together with the fact that F_μ satisfies Assumption 8.1, it follows that Assumption 8.2 is also satisfied. These results imply Inequality (8.17) via Theorem 8.1.

Now, by the same arguments of [24, Sec. IV.A], it follows that for every solution $\hat{z} = ((\hat{x}_1, \hat{x}_2, \hat{x}_3), \hat{\mu}) \in \mathcal{S}_{\hat{\mathcal{H}}_m}$ there exists a monotonously decreasing sequence of positive numbers $\{v_j\}_{j=1}^{\infty}$ such that the following bound holds

$$f(\hat{x}_1(s,j)) - f^\star \leq \frac{v_j}{\hat{x}_3(s,j)^2}$$

for every $(s,j) \in \mathrm{dom}(\hat{z})$. Now, let $\underline{s}_j \coloneqq \min\{s \ : \ (s,j) \in \mathrm{dom}(\hat{z})\}$. Noting that $\frac{\mathrm{d}}{\mathrm{d}s}\hat{x}_3(s,j) = \frac{1}{2}$, and since $\hat{x}_3(\underline{s}_j, j) = \underline{T}$, by the definition of the jump map G_m, it follows that $\hat{x}_3(s,j)^2 = \left(\frac{1}{2}(s - \underline{s}_j) + \underline{T}\right)^2 > \frac{1}{4}(s - \underline{s}_j)^2$ for all $(s,j) \in \mathrm{dom}(\hat{z})$. This result, together with the above inequality yields the following bound

$$f(\hat{x}_1(s,j)) - f^\star \leq \frac{4v_j}{(s - \underline{s}_j)^2}.$$

Following the same arguments as in the proof of Proposition 8.1, and the relations between $\underline{s}_j$ and $\underline{t}_j \coloneqq \min\{t \ : \ (t,j) \in \mathrm{dom}(\mathcal{W}(\hat{z}))\}$ obtains Inequality (8.18). ■

Appendix G

Proofs of Chapter 9

G.1 Proofs of Section 9.1

The results below follow directly by computations and/or straightforward extensions or specializations of existing results in the literature.

Proof of Lemma 9.1

By direct integration, we have that:

$$\int_{\mu_0}^{\mu_k(t)} \frac{d\mu_k}{\mu_k^{1+\frac{1}{k}}} = \int_0^t \frac{k}{T} dt \implies -k\mu_k^{\frac{-1}{k}}\Big|_{\mu_0}^{\mu_k(t)} = \frac{k}{T}t.$$

Thus, it follows that $k\left(-\mu_k(t)^{\frac{-1}{k}} + \mu_k(0)^{\frac{-1}{k}}\right) = \frac{k}{T}t$, and:

$$\frac{1}{\mu_k(t)^{\frac{1}{k}}} = \frac{1}{\mu_k(0)^{\frac{1}{k}}} - \frac{t}{T} = \frac{T - t\mu_k(0)^{\frac{1}{k}}}{T\mu_k(0)^{\frac{1}{k}}},$$

from which we obtain the result. ■

Proof of Lemma 9.3

By direct integration, we have that:

$$\int_{\mu_0}^{\hat{\mu}_k(t)} \frac{d\hat{\mu}_k}{\hat{\mu}_k^{\frac{1}{k}}} = \int_0^t \frac{k}{T} dt \implies \frac{1}{1-\frac{1}{k}}\hat{\mu}_k^{1-\frac{1}{k}k}\Big|_{\mu_0}^{\hat{\mu}_k(t)} = \frac{k}{T}t.$$

Therefore, we obtain $\frac{k}{k-1}\left(\hat{\mu}_k^{1-\frac{1}{k}}(t) - \mu_0^{1-\frac{1}{k}}\right) = \frac{k}{T}t$, and:

$$\hat{\mu}_k(t) = \left(\frac{k-1}{T}t + \mu_0^{\frac{k-1}{k}}\right)^{\frac{k}{k-1}}.$$

This obtains the result. ■

Proof of Proposition 9.1

(P1) Follows by the monotonicity of $\omega_k(\cdot,\cdot)$ in its first argument, combined with the limit $\lim_{t\to\Upsilon_{T,k}} \mu_k(t) = \infty$.

(P2) For $k > 1$, the result follows by direct computation. For $k = 1$, the result is obtained by the properties of the logarithm.

(P3) By definition, the equality $\mathcal{T}_k(0) = 0$ holds for all $k \in \mathbb{R}_{\geq 1}$. For $k = 1$, by direct computation, we have: $\frac{d\mathcal{T}_1(t)}{dt} = \frac{T}{\mu_1(t)}\dot{\mu}_1(t) = \mu_1(t)$. For $k > 1$, by the chain rule, we obtain:

$$\frac{d\mathcal{T}_k(t)}{dt} = \frac{\partial \omega_k(b, \mu_k(0))}{\partial b}\bigg|_{b=\mu_k(t)} \dot{\mu}_k = \mu_k(t).$$

(P4) For $k = 1$, we have that $\mu_1(t) = \frac{\mu_0 T}{T - \mu_0 t}$. It then follows that $s = \left(\mathcal{T}_1 \circ \mathcal{T}_1^{-1}\right)(s) = T \ln\left(\frac{\mu_1\left(\mathcal{T}^{-1}(s)\right)}{\mu_0}\right)$. Solving for $\mathcal{T}_1^{-1}(s)$ leads to $\mathcal{T}_1^{-1}(s) = \Upsilon_{T,1}\left(1 - e^{-\frac{s}{T}}\right)$. For $k > 1$, let $y_k \coloneqq \mathcal{T}_k^{-1}$. By using (9.4), and the inverse function theorem, we obtain that $\frac{dy_k}{ds} = \frac{\left(\Upsilon_{T,k} - y\right)^k}{T^k}$. Then, by direct integration and using the fact that $y_k(0) = 0$, we obtain the following equality $\Upsilon_{T,k} - y_k(s) = \left(\frac{(k-1)s}{T^k} + \Upsilon_{T,k}^{1-k}\right)^{\frac{1}{1-k}}$. Solving for $\mathcal{T}_k^{-1}(s)$, we obtain that $\mathcal{T}_k^{-1}(s) = \Upsilon_{T,k} - \Upsilon_{T,k}\left(1 + \frac{(k-1)s}{\Upsilon_{T,k}\mu_0}\right)^{\frac{1}{1-k}}$.

(P5) Follows directly by the inverse function theorem.

(P6) For $k = 1$, using the equality $\ln(1-x) = \sum_{l=1}^{\infty} \frac{-1}{l}x^l$, $|x| < 1$, we obtain that $\mathcal{T}_1(t) = \mu_0 t + \sum_{l=2}^{\infty} \frac{1}{l}\mu_0^l t^l T^{1-l}$, for all $t\mu_0 < T$. Letting $T \to \infty$, the second term in this expression vanishes, and we obtain that the equality $\lim_{T\to\infty} \mathcal{T}_1(t) = \mu_0 t$ holds for all

$(t, \mu_0) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 1}$. For $k > 1$, from Remark 9.5 it follows that

$$\mathcal{T}_k(t) = \frac{T\mu_0^{\frac{k-1}{k}}}{k-1}\left(\left(1 - \frac{t\mu_0^{\frac{1}{k}}}{T}\right)^{1-k} - 1\right). \tag{G.1}$$

Now, using the binomial theorem we have that

$$\left(1 - \frac{t\mu_0^{\frac{1}{k}}}{T}\right)^{1-k} - 1 = \frac{(k-1)t\mu_0^{\frac{1}{k}}}{T} + \sum_{l=2}^{\infty} g_{k,l}\left(\frac{t\mu_0^{\frac{1}{k}}}{T}\right)^l,$$

for all $t\mu_0^{\frac{1}{k}} < T$, and where $g_{k,l} = \frac{(k-1)k(k+1)\cdots(k+l-2)}{l!}$. Thus, for all $t\mu_0^{\frac{1}{k}} < T$, equality (G.1) can be written as $\mathcal{T}_k(t) = \mu_0^{\frac{k-1}{k^2}} t + \sum_{l=2}^{\infty} \frac{g_{k,l}}{k-1} t^l \mu_0^{\frac{(k-1)l}{k^2}} T^{1-l}$. Letting $T \to \infty$, the second term in this expression vanishes. Thus, it follows that the limit $\lim_{T\to\infty} \mathcal{T}_k(t) = \mu_0^{\frac{k-1}{k^2}} t$ holds for all $(t, \mu_0) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 1}$. ■

Proof of Proposition 9.2

We prove each item separately:

(a) Let $(\hat{z}, \hat{u})$ be a maximal solution pair of $\widehat{\mathcal{H}}$ from z_0. Then, for each $j \in \mathbb{Z}_{\geq 0}$ such that the interior of $\hat{I}_j \coloneqq \{s \geq 0 : (s, j) \in \mathrm{dom}(\hat{z})\}$ is nonempty, $\hat{z}$ satisfies:

$$\frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s, j) \in \frac{1}{\hat{\mu}_k(s, j)} F(\hat{z}(s, j), \hat{u}(s, j)), \tag{G.2}$$

for almost all $s \in I_j$. Using the chain rule, z satisfies:

$$\frac{\mathrm{d}}{\mathrm{d}t} z(t, j) = \frac{\mathrm{d}}{\mathrm{d}t}\hat{z}(\mathcal{T}_k(t), j) = \frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s, j) \cdot \dot{\mathcal{T}}_k(t),$$

and since $\dot{\mathcal{T}}_k(t) = \mu_k(t, j)$ for all $t \in [0, \Upsilon_{T,k})$ due to (9.7), and μ_k does not change during

the jumps (9.3d), using (G.2) we obtain:

$$\frac{\mathrm{d}}{\mathrm{d}t}z(t,j) = \mu_k(t,j)\frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s,j) \in \frac{\mu_k(t,j)}{\hat{\mu}_k(s,j)}F(\hat{z}(s,j),\hat{u}(s,j)).$$

By construction, $\mu_k(t,j) = \hat{\mu}_k(s,j)$, $u(t,j) = \hat{u}(s,j)$ and $z(t,j) = \hat{z}(s,j)$ via the time dilation $s = \mathcal{T}_k(t)$. Thereforc, substituting in the above inclusion we obtain that $\dot{z}(t,j)$ satisfies (9.3b) for almost all $t \in I_j \coloneqq \{t \geq 0 : (t,j) \in \mathrm{dom}(z)\}$. Moreover, note that $\mathcal{T}_k(\underline{t}_j) = \underline{s}_j$ and $\mathcal{T}_k(\overline{t}_j) = \overline{s}_j$ where $\underline{t}_j \coloneqq \min I_j$, $\overline{t}_j = \sup I_j$, $\underline{s}_j \coloneqq \min \hat{I}_j$, $\overline{s}_j = \sup \hat{I}_j$. Similarly, for every $(s,j) \in \mathrm{dom}(\hat{z})$ such that $(s,j+1) \in \mathrm{dom}(\hat{z})$, we have that $\hat{z}(s,j+1) \in G(\hat{z}(s,j))$, and therefore $z(t,j+1) \in G(z(t,j))$. Thus (z,u) is a maximal solution to $\mathcal{H}$.

(b) Let (z,u) be a maximal solution pair of $\mathcal{H}$ from z_0. Using again the chain rule, and the definition of $\hat{z}$, we obtain that for each j for which the interior of $I_j \coloneqq \{t \geq 0 : (t,j) \in \mathrm{dom}(z)\}$ is nonempty, the signal $\hat{z}$ satisfies:

$$\frac{\mathrm{d}}{\mathrm{d}s}\hat{z}(s,j) = \frac{\mathrm{d}z}{\mathrm{d}\mathcal{T}_k^{-1}}\frac{\mathrm{d}\mathcal{T}_k^{-1}}{\mathrm{d}s} = \frac{\dot{z}(t,j)}{\mu_k(t,j)} \in \frac{F(z(t,j),u(t,j))}{\mu_k(t,j)},$$

where we used (9.3b) and (9.9). Note that by construction $\hat{z}(s,j) = z(t,j)$, $\hat{\mu}_k(s,j) = \mu_k(t,j)$, and $\hat{u}(s,j) = u(t,j)$ via the time contraction $t = \mathcal{T}_k^{-1}(s)$. Then, by substituting in the above expression we obtain that $\hat{z}$ satisfies $\dot{\hat{z}}_s \in \frac{1}{\hat{\mu}_k}F(\hat{z},\hat{u}_k)$ for almost all $s \in \hat{I}_j = \{s \geq 0 : (s,j) \in \mathrm{dom}(\hat{z})\}$. Moreover, note that $\mathcal{T}_k^{-1}(\underline{s}_j) = \underline{t}_j$ and $\mathcal{T}_k^{-1}(\overline{s}_j) = \overline{t}_j$ where $\underline{t}_j \coloneqq \min I_j$, $\overline{t}_j = \sup I_j$, $\underline{s}_j \coloneqq \min \hat{I}_j$, $\overline{s}_j = \sup \hat{I}_j$. Since for every $(t,j) \in \mathrm{dom}(z)$ such that $(t,j+1) \in \mathrm{dom}(z)$, we have that $z(t,j+1) \in G(z(t,j))$, and therefore $\hat{z}(s,j+1) \in G(\hat{z}(s,j))$, it follows that $(\hat{z},\hat{u})$ is a maximal solution pair to $\widehat{\mathcal{H}}$. ■

G.2 Proofs of Section 9.2

Proof of Lemma 9.4

Let $\tau_d > 0$, $N_0 \geq 1$, and $\sigma(t) \in \Sigma_{\text{ADT}}(\tau_d, N_0)$. Then, it follows that

$$N(t_2, t_1) \leq \frac{1}{\tau_d}(t_2 - t_1) + N_0, \tag{G.3}$$

for all $t_1 \leq t_2$. We prove that expression (G.3) can be upper bounded by the right-hand side of (9.20).

Case $k = 1$: Assume that $t_1, t_2 \in [0, \Upsilon_{T,k})$ and define $X = \frac{\Upsilon_{T,k} - t_1}{\Upsilon_{T,k} - t_2}$, where $\Upsilon_{T,1} = T\mu_0^{-1}$ and $\mu_0 \geq 1$ fixed. Then, $X \geq 1$ and:

$$t_2 - t_1 = (\Upsilon_{T,1} - t_1)\left(1 - \frac{1}{X}\right).$$

Now, fix t_1 and define $f(X) = \ln(X) - T^{-1}(\Upsilon_{T,1} - t_1)\left(1 - \frac{1}{X}\right)$. Since t_1 satisfies $t_1 \leq \Upsilon_{T,1} \leq T$ by assumption, it follows that there exists $\delta_{t_1} \in [0, 1]$ such that:

$$f(X) = \ln(X) - \delta_{t_1}\left(1 - \frac{1}{X}\right).$$

By noting that $f(1) = 0$, and since $X \geq 1$, it follows that the derivative of f satisfies:

$$f'(X) = \frac{1}{X}\left(1 - \frac{\delta_{t_1}}{X}\right) \geq 0,$$

for all $\delta_{t_1} \in [0, 1]$. Thus, $f(X) \geq 0$ for all $X \geq 1$ and $t_1 \leq \Upsilon_{T,1}$. Equivalently, by using the definition of X, it follows that:

$$T \ln\left(\frac{\Upsilon_{T,1} - t_1}{\Upsilon_{T,1} - t_2}\right) - (t_2 - t_1) \geq 0,$$

for all $0 \leq t_1 \leq t_2 < \Upsilon_{T,1}$, where we have used the definition of X. Using this bound in (G.3) yields:

$$N(t_2, t_1) \leq \frac{T}{\tau_d} \ln \left(\frac{\Upsilon_{T,1} - t_1}{\Upsilon_{T,1} - t_2} \right) + N_0,$$

for all $0 \leq t_1 \leq t_2 < \Upsilon_{T,1}$, which implies that $\sigma(t)$, when restricted to $[0, \Upsilon_{T,1})$, satisfies the bound (9.20) for $k = 1$.

Case $k > 1$: Assume that $t_1, t_2 \in [0, \Upsilon_{T,k})$, with $\Upsilon_{T,k} = T\mu_0^{-\frac{1}{k}}$, $T > 0$ and $\mu_0 \geq 1$. Let $\Delta = t_2 - t_1$, and define

$$f(\Delta) = \mathcal{T}_k(t_1 + \Delta) - \mathcal{T}_k(t_1) - \Delta, \quad \Delta \in [0, \Upsilon_{T,k}).$$

Then, by using the result of Proposition 9.1-(P3) the derivative of f satisfies:

$$f'(\Delta) = \mu_k(t_1 + \Delta) - 1,$$

for all $t_1, \Delta \in [0, \Upsilon_{T,k})$. Since $\mu_k(t) \geq 1$ for all $t \in \mathbb{R}_{\geq 0}$, the previous equality implies that $f'(\Delta) \geq 0$. This result, together with the fact that $f(0) = 0$, implies that $f(\Delta) \geq 0$ for all $t_1, \Delta \in [0, \Upsilon_{T,k})$. Equivalently, by using the definition of Δ we obtain:

$$\begin{aligned} 0 \leq \mathcal{T}_k(t_2) - \mathcal{T}_k(t_1) - (t_1 - t_2) \\ \implies (t_1 - t_2) \leq \omega_k(\mu_k(t_2), \mu_k(t_1)), \end{aligned}$$

where the implication follows from the result of Proposition 9.1-(P2). Using this bound in (G.3) yields:

$$N(t_2, t_1) \leq \frac{1}{\tau_d} \omega_k(\mu_k(t_2), \mu_k(t_1)) + N_0,$$

for all $0 \leq t_1 \leq t_2 < \Upsilon_{T,1}$, which implies that $\sigma(t)$, when restricted to $[0, \Upsilon_{T,1})$, satisfies the bound (9.20) for $k \in \mathbb{Z}_{\geq 1}$. ■

Proof of Lemma 9.5

The case $k = 1$ follows directly by the definition of $\mathcal{T}_1$ and Remark 9.3. For $k > 1$, consider expanding the right-hand side of (9.20):

$$\begin{aligned} N(t_2, t_1) &\leq \frac{T}{\tau_d} \left(\frac{\mu_k(t_2)^{\frac{k-1}{k}}}{k-1} - \frac{\mu_k(t_1)^{\frac{k-1}{k}}}{k-1} \right) + N_0 \\ &= \frac{T^k}{\tau_d(k-1)} \left(\frac{(\Upsilon_{T,k} - t_1)^{k-1} - (\Upsilon_{T,k} - t_2)^{k-1}}{((\Upsilon_{T,k} - t_2)(\Upsilon_{T,k} - t_1))^{k-1}} \right) + N_0. \end{aligned}$$

Taking the limit as $k \to 1$, one obtains (9.21), see also Remark 9.3. On the other hand, when $k \in \mathbb{Z}_{>1}$, the Binomial theorem can be used to write $(\Upsilon_{T,k} - t_i)^{k-1} = \sum_{\ell=0}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} (-t_i)^\ell$, for $i \in \{1, 2\}$, where $b_{k,l} \coloneqq \frac{(k-1)!}{\ell!(k-\ell-1)!}$ are the so-called Binomial coefficients. Let

$$\begin{aligned} S &\coloneqq \sum_{\ell=0}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} (-t_1)^\ell - \sum_{\ell=0}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} (-t_2)^\ell \\ &= \sum_{\ell=1}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} (-t_1)^\ell - \sum_{\ell=1}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} (-t_2)^\ell \\ &= b_{k,1} \Upsilon_{T,k}^{k-2} (t_2 - t_1) + \sum_{\ell=2}^{k-1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} \left((-t_1)^\ell - (-t_2)^\ell \right) \\ &= b_{k,1} \Upsilon_{T,k}^{k-2} (t_2 - t_1) + \sum_{\ell=2}^{k-1} (-1)^{\ell+1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} \left(t_2^\ell - t_1^\ell \right). \end{aligned}$$

Therefore, the BU_k-ADT bound can be written as

$$\begin{aligned} N(t_2, t_1) &\leq \frac{T^k}{\tau_d(k-1)} \left(\frac{S}{((\Upsilon_{T,k} - t_2)(\Upsilon_{T,k} - t_1))^{k-1}} \right) + N_0 \\ &= \frac{\gamma_k(t_1, t_2)}{\tau_d} \left[(t_2 - t_1) + \sum_{\ell=2}^{k-1} \tilde{c}_{\ell,k} \left(t_2^\ell - t_1^\ell \right) \right] + N_0, \end{aligned}$$

where

$$\tilde{c}_{\ell,k} = (-1)^{\ell+1} b_{k,l} \Upsilon_{T,k}^{k-1-\ell} \left(b_{k,1} \Upsilon_{T,k}^{k-2} \right)^{-1} = (-1)^{\ell+1} \frac{b_{k,l}}{b_{k,1}} \Upsilon_{T,k}^{1-\ell},$$

and

$$\begin{aligned}
\gamma_k(t_1, t_2) &= \frac{b_{k,1} T^k \Upsilon_{T,k}^{k-2}}{(k-1)} \left(\frac{1}{(\Upsilon_{T,k} - t_2)(\Upsilon_{T,k} - t_1)} \right)^{k-1} \\
&= \frac{T^k}{\Upsilon_{T,k}} \left[\frac{\Upsilon_{T,k}}{(\Upsilon_{T,k} - t_2)\,(\Upsilon_{T,k} - t_1)} \right]^{k-1} \\
&= \mu_0 \left[\frac{\Upsilon_{T,k}^2}{(\Upsilon_{T,k} - t_2)\,(\Upsilon_{T,k} - t_1)} \right]^{k-1}
\end{aligned}$$

where we have used the fact that $b_{k,1} = k - 1$. ■

Proof of Lemma 9.6

Proof: The HDS $\mathcal{H}$ given by (9.3) has state $z = (\psi, \mu_k) \in \mathbb{R}^{n+3}$ with $\psi = (x, \tau, q) \in \mathbb{R}^{n+2}$, and dynamics

$$z \in C \coloneqq \mathbb{R}^n \times [0, N_0] \times \mathcal{Q}_s \times \mathbb{R}_{\geq 1} \tag{G.4a}$$

$$\dot{z} = \begin{pmatrix} \dot{x} \\ \dot{\tau} \\ \dot{q} \\ \dot{\mu}_k \end{pmatrix} \in F(z, u) \coloneqq \begin{pmatrix} \mu_k \cdot f_q(x, \mu_k, u, \tau) \\ \left[0, \dfrac{\mu_k}{\tau_d}\right] \\ 0 \\ \dfrac{k}{T} \mu_k^{1+\frac{1}{k}} \end{pmatrix}, \tag{G.4b}$$

$$z \in D \coloneqq \mathbb{R}^n \times [1, N_0] \times \mathcal{Q}_s \times \mathbb{R}_{\geq 1}, \tag{G.4c}$$

$$z^+ = \begin{pmatrix} x^+ \\ \tau^+ \\ q^+ \\ \mu_k^+ \end{pmatrix} \in G(z) \coloneqq \begin{pmatrix} R_q(x) \\ \tau - 1 \\ \mathcal{Q}_s \backslash \{q\} \\ \mu_k \end{pmatrix}. \tag{G.4d}$$

Since the function μ_k generated by (G.4) is precisely (9.4), any solution $z : \mathrm{dom}(z) \to \mathbb{R}^{n+3}$ to (G.4) will necessarily satisfy $\mathrm{length}_t(\mathrm{dom}(z)) \leq \Upsilon_{T,k}$. By Proposition 9.2, the corresponding HDS (9.11) in the (s,j)-time scale is given by:

$$\hat{z} \in C, \ \dot{\hat{z}}_s = \begin{pmatrix} \dot{\hat{x}}_s \\ \dot{\hat{\tau}}_s \\ \dot{\hat{q}}_s \\ \dot{\hat{\mu}}_{k_s} \end{pmatrix} \in \hat{F}_T(\hat{z}, \hat{u}) \coloneqq \begin{pmatrix} f_{\hat{q}}(\hat{x}, \hat{\mu}_k, \hat{u}, \hat{\tau}) \\ \left[0, \frac{1}{\tau_d}\right] \\ 0 \\ \frac{k}{T} \hat{\mu}_k^{\frac{1}{k}} \end{pmatrix} \tag{G.5a}$$

$$\hat{z} \in D, \qquad \hat{z}^+ \in G(\hat{z}), \tag{G.5b}$$

where C, D, and G were defined in (G.4). Since Assumption 9.3 ensures that the state $\hat{x}$ does not exhibit finite escape times, by noting that the dynamics of $(\hat{\tau}, \hat{q})$ are decoupled from $\hat{\mu}_k$, and since $\hat{\mu}_k$ remains constant during jumps, we can directly obtain $\hat{\mu}_k(s,j)$ for any $(s,j) \in \mathrm{dom}(\hat{z})$ using Proposition 9.3: $\hat{\mu}_k(s,j) = \left(\frac{(k-1)}{T} t + \hat{\mu}(\underline{s}_j, j)^{\frac{k-1}{k}}\right)^{\frac{k}{k-1}}$, for $k > 1$, and $\hat{\mu}_k(s,j) = \hat{\mu}_k(\underline{s}_j, j) e^{\frac{s}{T}}$, for $k = 1$, where $\underline{s}_j \coloneqq \min\{s \geq 0 : (s,j) \in \mathrm{dom}(\hat{z})\}$. By [33, Ex. 2.15] it follows that every solution $\hat{z}$ of (G.5) has a HTD that satisfies the ADT bound in the (s,j)-time scale:

$$j_2 - j_1 \leq \frac{1}{\tau_d}(s_2 - s_1) + N_0, \tag{G.6}$$

for all $(s_1, j_1), (s_2, j_2) \in \mathrm{dom}(\hat{z})$, with $s_2 > s_1 \geq 0$. Additionally, by [33, Ex. 2.15], for every hybrid time domain satisfying (G.6), there exists a solution to the HDS (G.5) having

said hybrid time domain. Thus, it remains to show that (G.6) is equivalent to (9.20) in the original (t, j)-time scale. Using the time scaling function $\mathcal{T}_k$ given by (9.6), for any solution z of (G.4) and all $(t_1, j_1), (t_2, j_2) \in \mathrm{dom}(z)$ with $0 \leq t_1 < t_2$, we have that $(s_1, j_1), (s_2, j_2) \in \mathrm{dom}(\hat{z})$, where $s_1 = \mathcal{T}_k(t_1)$, $s_2 = \mathcal{T}_k(t_2)$, and $0 \leq s_1 < s_2$. Substituting in (G.6):

$$j_2 - j_1 \leq \frac{1}{\tau_d}(\mathcal{T}_k(t_2) - \mathcal{T}_k(t_1)) + N_0.$$

The result follows now by using (P2) in Proposition 9.1. ■

Proof of Theorem 9.1

The proof has three main steps.

Step 1: Stability of the "target" HDS $\widehat{\mathcal{H}}$ in the (s, j)-Hybrid Time Scale: The overall HDS is given by (G.4), which in the (s, j)-time scale is given by (G.5). To study the stability properties of system (G.5), we consider the Lyapunov function $W(\hat{z}) \coloneqq V_{\hat{q}}(\hat{x}, \hat{\tau})e^{\ln(r)\hat{\tau}}$. By Assumption 9.3, this function satisfies $\underline{c}|\hat{z}|^p_{\mathcal{A}} \leq W(\hat{z}) \leq \overline{c}|\hat{z}|^p_{\mathcal{A}}, \quad \forall\ \hat{z} \in C \cup D$, with $\underline{c} \coloneqq \min_{p\in\mathcal{Q}} c_{1,p}$, $\overline{c} \coloneqq e^{\ln(r)N_0}\overline{c}_2$, and $\overline{c}_2 \coloneqq \max_{p\in\mathcal{Q}} c_{2,p}$. When $\hat{z} \in C$, for all $\eta \in [0, 1/\tau_d]$, we have:

$$\begin{aligned}\langle \nabla W(\hat{z}), \hat{F}_T(\hat{z}, \hat{u})\rangle &= \left\langle \nabla V_{\hat{q}}(\hat{x}, \hat{\tau}), \begin{pmatrix} f_{\hat{q}}(\hat{x}, \hat{u}, \hat{\tau}) \\ \eta \end{pmatrix} \right\rangle e^{\ln(r)\hat{\tau}} \\ &\quad + \langle \ln(r) V_{\hat{q}}(\hat{x}, \hat{\tau})e^{\ln(r)\hat{\tau}}, \dot{\hat{\tau}}_s\rangle \\ &\leq -\underline{c}_3\left(1 - \frac{\ln(r)}{\underline{c}_3\tau_d}\right)W(\hat{z}) + \overline{c}_4 e^{\ln(r)N_0}\Delta(\hat{\mu}_k)|\hat{u}|^p,\end{aligned}$$

where $\underline{c}_3 \coloneqq \min_{p\in\mathcal{Q}} c_{p,3}$, $\overline{c}_4 \coloneqq \max_{p\in\mathcal{Q}} c_{p,4}$, and where we used item (b) in Assumption 9.3.

On the other hand, when $\hat{z} \in D$ we can use Assumption 9.3-(d) to obtain

$$\begin{aligned} W(\hat{z}^+) &= V_{\hat{q}^+}(\hat{x}^+, \hat{\tau}^+) e^{\ln(r)\hat{\tau}^+} = V_{\hat{q}^+}\left(R_{\hat{q}}(\hat{x}), \hat{\tau} - 1\right) e^{\ln(r)(\hat{\tau}-1)} \\ &\leq \chi V_{\hat{q}}(\hat{x}, \hat{\tau}) e^{\ln(r)(\hat{\tau}-1)} = \frac{\chi}{r} W(\hat{z}). \end{aligned}$$

Thus, using the definition of r, during jumps we obtain

$$W(\hat{z}^+) - W(\hat{z}) \leq -\left(1 - \frac{\chi}{\max\{1, \chi\}}\right) W(\hat{z}) \leq 0$$

Using Lemma G.3, we conclude that every solution $\hat{z}$ of system (G.5) satisfies:

$$|\hat{z}(s,j)|_{\mathcal{A}} \leq \kappa_1 e^{-\kappa_2(s+j)} |\hat{z}(0,0)|_{\mathcal{A}} + \kappa_3 \cdot \sup_{0 \leq \zeta \leq s} |\hat{\Delta}(\zeta)|, \tag{G.7}$$

for all $(s,j) \in \mathrm{dom}(\hat{z})$, where $\kappa_1 = (\overline{c}/\underline{c})^{1/p} e^{\frac{\lambda}{2p}\frac{\tau_d}{1+\tau_d} N_0}$ $\kappa_2 = \lambda\tau_d/(2p(1+\tau_d))$, $\kappa_3 = \left(2\overline{c}_4 r^{N_0}/[\lambda \underline{c}]\right)^{1/p}$, $\lambda = \underline{c}_3 - \ln(r)/\tau_d$, and $\hat{\Delta}(s) \coloneqq \Delta(\hat{\mu}_k(s))\hat{u}(s)$. Moreover, when $\Delta(\hat{\mu}_k) = \hat{\mu}_k^{-\ell}$, via Lemma G.4, there exists $\beta_k \in \mathcal{KL}$ such that every solution $\hat{z}$ of system (G.5) satisfies:

$$|\hat{z}(s,j)|_{\mathcal{A}} \leq \beta_k\Big(\overline{\kappa}_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\overline{\kappa}_2(s+j)} + \overline{\kappa}_3 |\hat{u}|_{(s,j)}, s\Big), \tag{G.8}$$

for all $(s,j) \in \mathrm{dom}(\hat{z})$, with $\overline{\kappa}_1 \coloneqq \kappa_1$, $\kappa_2 \coloneqq \frac{\kappa_2}{2}$, $\overline{\kappa}_3 \coloneqq 2\kappa_3$.

Step 2: PT-ISS$_F$ of the HDS in the (t,j) - Time Scale: We now use the properties of the solutions $\hat{z}$ of system (G.5) to establish properties for the solutions z of system (G.4). First, we use Proposition 9.2 and let $s = \mathcal{T}_k(t)$, which yields $e^{-\kappa_2(s+j)} = e^{-\kappa_2(\mathcal{T}_k(t)+j)}$, and $|\hat{z}(\mathcal{T}_k(t), j)|_{\mathcal{A}} = |z\left(\mathcal{T}_k^{-1}(\mathcal{T}_k(t)), j\right)|_{\mathcal{A}} = |z(t,j)|_{\mathcal{A}}$. Then, by substituting in (G.7) and noting that $\mathcal{T}_k(0) = \mathcal{T}_k^{-1}(0) = 0$, it follows that when $\Delta = 0$ or $\Delta = 1$, every solution $z = (\psi, \mu_k)$ of the HDS (G.4) with $\mu_k(0,0) = \mu_0 \geq 1$ satisfies the bound:

$$|z(t,j)|_{\mathcal{A}} \leq \kappa_1 e^{-\kappa_2(\mathcal{T}_k(t)+j)} |z(0,0)|_{\mathcal{A}} + \kappa_3 \Delta |u|_{(t,j)}, \tag{G.9}$$

for all $(t,j) \in \mathrm{dom}(\psi)$, which implies that $\mathcal{A}$ is PT-ISS$_F$. Similarly, when $\Delta(\hat{\mu}_k) = \hat{\mu}_k^{-\ell}$, (G.8) leads to:

$$|z(t,j)|_{\mathcal{A}} \leq \beta_k\Big(\overline{\kappa}_1 |z(0,0)|_{\mathcal{A}} e^{-\overline{\kappa}_2(\mathcal{T}_k(t)+j)} + \overline{\kappa}_3 |u|_{(t,j)}, \mathcal{T}_k(t)\Big), \tag{G.10}$$

for all $(t,j) \in \mathrm{dom}(z)$. Inequality (G.10) implies that $\mathcal{A}$ is PT-ISS-C$_F$.

Step 3: Length of solutions in the (t,j) - Time Scale: Finally, we show that

$$\sup_t(\mathrm{dom}(z)) = \Upsilon_{T,k}$$

for all solutions z of (G.4). First, note that by the definition of $\mathcal{T}_k$ and Proposition 9.2, we have $\sup_t(\mathrm{dom}(z)) = \Upsilon_{T,k}$ if and only if $\sup_s(\mathrm{dom}(\hat{z})) = \infty$. Furthermore, based on the bound (G.6), we obtain hat $j \leq \frac{1}{\tau_d} s + N_0$ for any $(s,j) \in \mathrm{dom}(\hat{z})$. Since every complete solution $\hat{z}$ of (G.5) satisfies $\mathrm{length}(\mathrm{dom}(\hat{z})) = \infty$, and noting that $\mathrm{length}(\mathrm{dom}(\hat{z})) = \sup_s(\mathrm{dom}(\hat{z})) + \sup_j(\mathrm{dom}(\hat{z}))$, we can infer that if $j \to \infty$, then $s \to \infty$. Consequently, every complete solution of (G.5) must satisfy $\sup_s(\mathrm{dom}(\hat{z})) = \infty$, which in turn implies that $\sup_t(\mathrm{dom}(z)) = \Upsilon_{T,k}$ for such solutions. ■

Proof of Corollary 9.1

Using (9.10b) and the bounds obtained in Step 2 of the proof of Theorem 9.1, it follows that $e^{-\kappa_2(\mathcal{T}_k(t)+j)} = e^{-\alpha \ln\left(\frac{\mu_1(t)}{\mu_0}\right)} e^{-\kappa_2 j} = \left(\frac{\mu_0}{\mu_1(t)}\right)^{\alpha} e^{-\frac{\alpha}{T} j}$ where $\alpha = \kappa_2 T$. Since $|z|_{\mathcal{A}} = |x|$ for every solution, inequality (G.9) becomes (9.25). Similarly, inequality (G.10) becomes (9.26) with $\alpha_1 \coloneqq \overline{\kappa}_1$, $\alpha_2 \coloneqq (\overline{\kappa}_3 + \overline{\kappa}_2)T$, $\alpha_3 \coloneqq \overline{\kappa}_3 T$, $\alpha_4 \coloneqq \overline{\kappa}_2$, $\alpha_5 \coloneqq \overline{\kappa}_2 T$, and $\alpha_6 \coloneqq \overline{\kappa}_4$. ■

Proof of Lemma 9.7

The overall HDS has state $z = (\psi, \mu_k) \in \mathbb{R}^{n+4}$ with $\psi = (x, \tau, \rho, q,)$, and the following dynamics:

$$z \in C \coloneqq \mathbb{R}^n \times [0, N_0] \times [0, T_0] \times \mathcal{Q} \times \mathbb{R}_{\geq 1}, \tag{G.11a}$$

$$\dot{z} = \begin{pmatrix} \dot{x} \\ \dot{\tau} \\ \dot{\rho} \\ \dot{q} \\ \dot{\mu}_k \end{pmatrix} \in F(z, u) \coloneqq \begin{pmatrix} \mu_k \cdot f_q(x, \mu_k, u, \tau) \\ \left[0, \dfrac{\mu_k}{\tau_d}\right] \\ \left[0, \dfrac{\mu_k}{\tau_a}\right] - \mu_k \mathbb{I}_{\mathcal{Q}_u}(q) \\ 0 \\ \dfrac{k}{T} \mu_k^{1+\frac{1}{k}} \end{pmatrix}, \tag{G.11b}$$

$$z \in D \coloneqq \mathbb{R}^n \times [1, N_0] \times [0, T_0] \times \mathcal{Q} \times \mathbb{R}_{\geq 1}, \tag{G.11c}$$

$$z^+ = \begin{pmatrix} x^+ \\ \tau^+ \\ \rho^+ \\ q^+ \\ \mu_k^+ \end{pmatrix} \in G(z, u) \coloneqq \begin{pmatrix} R_q(x) \\ \tau - 1 \\ \rho \\ \mathcal{Q} \setminus \{q\} \\ \mu_k \end{pmatrix}. \tag{G.11d}$$

This system has a finite escape time at $t = \Upsilon_{T,k}$, induced by μ_k. Note that, by construction, the states (τ, ρ, q) are confined to the compact sets $[0, N_0]$, $[0, T_0]$, and $\mathcal{Q}$ respectively. Using the time variable $s = \mathcal{T}_k(t)$ defined in (9.6), and Proposition 9.2, we obtain the

following HDS in the (s, j)-time scale:

$$\hat{z} \in C,\ \dot{\hat{z}}_s = \begin{pmatrix} \dot{\hat{x}}_s \\ \dot{\hat{\tau}}_s \\ \dot{\hat{\rho}}_s \\ \dot{\hat{q}}_s \\ \dot{\hat{\mu}}_{k_s} \end{pmatrix} \in \hat{F}(\hat{z}, \hat{u}) \coloneqq \begin{pmatrix} f_q(\hat{x}, \hat{\mu}_k, \hat{u}, \hat{\tau}) \\ \left[0, \frac{1}{\tau_d}\right] \\ \left[0, \frac{1}{\tau_a}\right] - \mathbb{I}_{\mathcal{Q}_u}(\hat{q}) \\ 0 \\ \frac{k}{T}\hat{\mu}_k^{\frac{1}{k}} \end{pmatrix}, \tag{G.12a}$$

$$\hat{z} \in D, \qquad \hat{z}^+ \in G(\hat{z}), \tag{G.12b}$$

where the subscript s in (G.12a) indicates that the time derivative is taken with respect to s. Since (G.12) incorporates an ADT automaton $\hat{\tau}$ and a time-ratio monitor $\hat{\rho}$, by [111, Lemma 7] every solution $\hat{z}$ of (G.12) has a hybrid time domain such that for any pair $(s_1, j_1), (s_2, j_2) \in \mathrm{dom}(\hat{z})$ the bound (G.6) is satisfied, as well as the following bound:

$$\mathbb{T}(s_1, s_2) \coloneqq \int_{s_1}^{s_2} \mathbb{I}_{\mathcal{Q}_u}(\hat{q}(s, \underline{\hat{j}}(s)))ds \leq \frac{1}{\tau_a}(s_2 - s_1) + T_0, \tag{G.13}$$

where $\underline{\hat{j}}(s) \coloneqq \min\{j \in \mathbb{Z}_{\geq 0} \ :\ (s, j) \in \mathrm{dom}(\hat{q})\}$. Moreover, by [111, Lemma 7] every hybrid arc satisfying (G.13) can be generated by the HDS (G.12). Using $s = \mathcal{T}_k(t)$, the left-hand side of (G.13) can be expressed in the t-variable as:

$$\begin{aligned} \mathbb{T}(\mathcal{T}_k(t_2), \mathcal{T}_k(t_1)) &= \int_{t_1}^{t_2} \frac{\partial \mathcal{T}_k(t)}{\partial t} \cdot \mathbb{I}_{\mathcal{Q}_u}\Big(\hat{q}\Big(\mathcal{T}_k(t), \hat{j}\big(\mathcal{T}_k(t)\big)\Big)\Big)dt \\ &= \int_{t_1}^{t_2} \mu_k(t) \cdot \mathbb{I}_{\mathcal{Q}_u}\big(q(t, \underline{j}(t))\big)dt, \end{aligned} \tag{G.14}$$

where we used Proposition 9.1-(P3), together with the equality

$$\hat{q}(\mathcal{T}_k(t), \underline{\hat{j}}(\mathcal{T}_k(t))) = q\left(\mathcal{T}_k^{-1}(\mathcal{T}_k(t)),\ \underline{j}\left(\mathcal{T}_k^{-1}(\mathcal{T}_k(t))\right)\right) = q(t, \underline{j}(t)).$$

Using (G.13)-(G.14), together with Proposition 9.1-(P2), the AAT condition in the (t, j)-time scale becomes

$$\int_{t_1}^{t_2} \mu_k(t) \cdot \mathbb{I}_{\mathcal{Q}_u}(q(t, \underline{j}(t)))dt \leq \frac{1}{\tau_a}\omega_k\left(\mu_k(t_2), \mu_k(t_1)\right) + T_0,$$

which is precisely (9.27). The fact that inequality (9.20) holds follows by Lemma 9.6. ■

Proof of Theorem 9.2

The proof follows the same three steps as in the proof of Theorem 9.1. We start by using the time dilation $\mathcal{T}_k^{-1}$ and Proposition 9.2. Hence, we consider the HDS (G.12) in the (s, j)-time scale, with state $\hat{z} = (\hat{x}, \hat{\tau}, \hat{\rho}, \hat{q}, \hat{\mu}_k)$. To study the stability properties of this system, let $\hat{\xi} \coloneqq \ln(r)\hat{\tau} + (\underline{c}_3 + \overline{c}_5)\hat{\rho}$, and consider the Lyapunov function $W_2(\hat{z}) = V_{\hat{q}}(\hat{x}, \hat{\tau})e^{\hat{\xi}}$, which, by Assumption 9.3-(a), satisfies $\underline{\varphi}|\hat{z}|_{\mathcal{A}}^2 \leq W_2(\hat{z}) \leq \overline{\varphi}|\hat{z}|_{\mathcal{A}}^2$, with $\underline{\varphi} = \min_{p\in\mathcal{Q}} c_{p,1},\ \overline{\varphi} = \max_{p\in\mathcal{Q}} c_{p,2}e^{\ln(r)N_0 + (\underline{c}_3+\overline{c}_5)T_0}$. When $\hat{z} \in C$, the time derivative of $\hat{\xi}$ with respect to s satisfies:

$$\dot{\hat{\xi}}_s = \ln(r)\dot{\hat{\tau}}_s + (\underline{c}_3 + \overline{c}_5)\dot{\hat{\rho}}_s \in [0, \delta] - (\underline{c}_3 + \overline{c}_5)\mathbb{I}_{\mathcal{Q}_u}(\hat{q}),$$

where $\delta \coloneqq \frac{1}{\tau_d}\ln(r) + \frac{1}{\tau_a}(\underline{c}_3 + \overline{c}_5)$. Using the above expression together with Assumption 9.3, we evaluate the change of W_2 during the flows of stable and unstable modes. In particular, when $\hat{z} \in C$ and $\hat{q} \in \mathcal{Q}_s$, we have

$$\begin{aligned}\langle \nabla W_2(\hat{z}), \dot{\hat{z}}_s\rangle &= e^{\hat{\xi}}\left\langle \nabla V_{\hat{q}}(\hat{x}, \hat{\tau}),\ \dot{\hat{x}}_s\right\rangle + e^{\hat{\xi}}V_{\hat{q}}(\hat{x}, \hat{\tau})\dot{\hat{\xi}}_s \\ &\leq -(\underline{c}_3 - \delta)W_2(\hat{z}) + \frac{\overline{c}_4}{\overline{c}_2}\overline{\varphi}\hat{\Delta}(s)|\hat{u}|^p, \end{aligned} \tag{G.15}$$

where $\hat{\Delta}(s) \coloneqq \Delta(\hat{\mu}_k(s))\hat{u}(s)$, $\overline{c}_2 \coloneqq \max_{p\in\mathcal{Q}} c_{2,p}$ and $\overline{c}_4 = \max_{p\in\mathcal{Q}} c_{4,p}$, and where $\underline{c}_3 - \delta > 0$ since (9.31) is satisfied by assumption. On the other hand, when $\hat{z} \in C$ and $\hat{q} \in \mathcal{Q}_u$:

$$\begin{aligned}\langle \nabla W_2(\hat{z}), \dot{\hat{z}}_s \rangle &\leq \left(\overline{c}_5 V_{\hat{q}}(\hat{x},\hat{\tau}) + \overline{c}_4 \hat{\Delta}(s)|\hat{u}|\right) e^{\hat{\xi}} + V_{\hat{q}}(\hat{x},\hat{\tau}) e^{\hat{\xi}} \dot{\hat{\xi}}_s \\ &\leq (\delta - \underline{c}_3)\, W_2(\hat{z}) + \overline{c}_4 \hat{\Delta}(s)|\hat{u}| e^{\hat{\xi}} \\ &\leq -(\underline{c}_3 - \delta)\, W_2(\hat{z}) + \frac{\overline{c}_4}{\overline{c}_2}\overline{\varphi}\hat{\Delta}(s)|\hat{u}|^p,\end{aligned}$$

which is the same bound as (G.15).

During jumps, it follows that $\hat{\xi}^+ = \ln(r)\hat{\tau}^+ + (\underline{c}_3 + \overline{c}_4)\hat{\rho}^+ = \hat{\xi} - \ln(r)$ for all $\hat{z} \in D$. Then, using Assumption 9.3, the Lyapunov function satisfies:

$$\begin{aligned}W_2(\hat{z}^+) = V_{\hat{q}^+}(\hat{x}^+,\hat{\tau}^+) e^{\hat{\xi}^+} &= V_{\hat{q}^+}\left(R_{\hat{q}}(\hat{x}), \hat{\tau} - 1\right) e^{\hat{\xi} - \ln(r)} \\ &\leq \chi V_{\hat{q}}(\hat{x},\hat{\tau}) e^{\hat{\xi}-\ln(r)} = \frac{\chi}{\max\{1,\chi\}} W_2(\hat{z}) \leq W_2(\hat{x}).\end{aligned}$$

It follows that $W_2(\hat{z}^+) - W_2(\hat{z}) \leq 0$ for all $\hat{z} \in D$. Using Lemma G.3, we conclude that every solution $\hat{z}$ satisfies the bound

$$|\hat{z}(s,j)|_{\mathcal{A}} \leq \kappa_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\kappa_2(s+j)} + \kappa_3 \hat{\Delta}(s)|\hat{u}|_{(s,j)},$$

for all $(s,j) \in \mathrm{dom}(\hat{z})$, where $\kappa_1 = \left(\overline{\varphi}/\underline{\varphi}\right)^{1/p} e^{\frac{\lambda}{2p}\frac{\tau_d}{1+\tau_d}N_0}$ $\kappa_2 = \lambda\tau_d/(2p(1+\tau_d))$, $\kappa_3 = \left(2\overline{c}_4\overline{\varphi}/[\overline{c}_2\lambda\underline{\varphi}]\right)^{1/p}$, $\lambda = \underline{c}_3 - \delta$, and $\hat{\Delta}(s) \coloneqq \Delta(\hat{\mu}_k(s))\hat{u}(s)$. From here, the bounds (9.14)-(9.15) are obtained following the exact same arguments used in Steps 2 and 3 of the proof of Theorem 9.1. ■

G.3 Proofs of Section 9.3

Proof of Proposition 9.3

Proof: We show that under Assumption 9.4 a suitable Lyapunov function can be used to show that Assumption 9.3 is satisfied. Let $V_{\hat{q}}(\hat{x}) = \frac{1}{2\underline{\sigma}_{\hat{q}}}|\hat{x}|^2$ for every $\hat{q} \in \mathcal{Q}_s$. By employing Young's inequality, we obtain

$$\langle \nabla V_{\hat{q}}(\hat{x}),\ f_{\hat{q}}(\hat{x}, \hat{\mu}_k)\rangle \leq -2\underline{\sigma}_{\hat{q}}\eta_{\hat{q}} V_{\hat{q}}(\hat{x}) + \frac{1}{\hat{\mu}_k^2}\frac{1}{4\underline{\sigma}_{\hat{q}}^2\delta_{\hat{q}}}, \tag{G.16}$$

for all $\hat{q} \in \mathcal{Q}_s$. Similarly, for all $\hat{q} \in \mathcal{Q}_u$ let $V_{\hat{q}}(\hat{x}) = \frac{|\hat{x}|^2}{2}$. Using this function, we obtain

$$\langle \nabla V_{\hat{q}}(\hat{x}),\ f_{\hat{q}}(\hat{x}, \hat{\mu}_k)\rangle \leq V_{\hat{q}}(\hat{x}) + \frac{1}{\hat{\mu}_k^2}\frac{\bar{d}_{\hat{q}}^2}{2}, \tag{G.17}$$

for all $\hat{q} \in \mathcal{Q}_u$. Using $c_{\hat{q},1} = c_{\hat{q},2} = 1/2\underline{\sigma}_{\hat{q}}$, $c_{\hat{q},3} = 2\underline{\sigma}_{\hat{q}}\eta_{\hat{q}}$, $c_{\hat{q},4} = 1/4\underline{\sigma}_{\hat{q}}^2\delta_{\hat{q}}$, when $\hat{q} \in \mathcal{Q}_s$, and $c_{\hat{q},1} = c_{\hat{q},2} = 1/2$, $c_{\hat{q},5} = 1$, $c_{\hat{q},4} = \bar{d}_{\hat{q}}^2/2$ when $\hat{q} \in \mathcal{Q}_u$, together with the set of smooth functions $\{V_{\hat{q}}\}_{\hat{q}\in\mathcal{Q}}$, Assumption 9.3 is satisfied. Thus, we can always pick $\tau_a > 1$ and $\tau_d > 0$ large enough to satisfy the stability condition (9.31). Additionally, Assumption 9.2 is satisfied by the Lipschitz properties of both $d_q(\cdot)$ and $b_q(\cdot)$. Assumption 9.1 is met by the same Lipschitz property and the construction of the HDS $\mathcal{H}$ with data (9.29). It follow that $\mathcal{A}_\psi \times \mathbb{R}_{\geq 1}$ is PT-ISS-C$_F$ for the closed-loop system via Theorem 2-c).

We now prove the boundedness and convergence to 0 of the switching feedback-law u_q given in (9.33). By applying (G.10) from the proof of Theorem 2-c), for any $(x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}_{\geq 1}$ and any solution $z = (x, \tau, \rho, q, \mu_k)$ to the closed-loop system satisfying $x(0,0) = x_0$ and $\mu_k(0,0) = \mu_0$ we obtain:

$$|x(t,j)| \leq \beta_k(\overline{\kappa}_1 e^{-\overline{\kappa}_2(\mathcal{T}_k(t)+j)}|x(0,0)| + \overline{\kappa}_3\overline{u}, \mathcal{T}_k(t)), \tag{G.18}$$

for all $(t,j) \in \mathrm{dom}(z)$, where $\overline{\kappa}_1, \overline{\kappa}_2, \overline{\kappa}_3 > 0$, $\overline{u} \coloneqq \max\left\{\max_{q\in\mathcal{Q}_s} \frac{1}{4\overline{\sigma}_q^2\delta_q},\ \max_{q\in\mathcal{Q}_u} \frac{\overline{d}_q^2}{2}\right\}$, and $\beta_k(r,s) = r\cdot\max\{\overline{\kappa}_1 e^{-\overline{\kappa}_2 s}, \xi_k^{-2}(s)\} =: r\cdot\alpha_k(s)$, with $\xi_k(s) = \left(\frac{k-1}{T}s+1\right)^{\frac{k}{k-1}}$, is the same $\mathcal{KL}$ function obtained in Lemma G.4. Then, from (G.18) we obtain:

$$|x(t,j)| \leq \left(\overline{\kappa}_1 e^{-\overline{\kappa_2}(\mathcal{T}_k(t)+j)}|x(0,0)| + \overline{\kappa}_3\overline{u}\right)\alpha_k(\mathcal{T}(t)),$$

for all $(t,j) \in \mathrm{dom}(z)$. Hence, u_q satisfies:

$$|u_q(x(t,j),\mu_k(t))| \leq \tilde{r}_k(t,j)\left|\eta_q + \delta_q\overline{d}_q^2(x(t,j))\right|\mu_k(t)\alpha_k(\mathcal{T}_k(t)),$$

for all $(t,j) \in \mathrm{dom}(z)$ and all $q \in \mathcal{Q}$, where $\tilde{r}_k(t,j) = \left(\overline{\kappa}_1 e^{-\overline{\kappa}_2(\mathcal{T}_k(t)+j)}|x(0,0)| + \overline{\kappa}_3\overline{u}\right)$. Since $\overline{d}_q(\cdot)$ is assumed to be continuous for all $x \in \mathbb{R}^n$, it is locally bounded. Then, $\tilde{r}_k(t,j)\left|\eta_q + \delta_q\overline{d}_q^2(x(t,j))\right|$ is bounded as $\tilde{r}(t,j)$ is bounded by definition. Now, note that $\alpha_k(s) = \max\{\overline{\kappa}_1 e^{-\overline{\kappa}_2 s}, \xi_k^{-2}(s)\} = \xi_k^{-2}(s)$ for s sufficiently large since the inverse exponential decays faster than any proper rational function. Additionally, by leveraging the result of Proposition 2 it follows that $\mu_k(t) = \hat{\mu}_k(\mathcal{T}_k(t)) = \left(\frac{k-1}{T}\mathcal{T}_k(t) + \mu_0^{\frac{k-1}{k}}\right)^{\frac{k}{k-1}}$ for $k \geq 2$. Then, as $t \to \Upsilon_{T,k}$ we have that

$$\mu_k(t)\alpha_k(\mathcal{T}_k(t)) = \left[\left(\frac{k-1}{T}\mathcal{T}_k(t) + \mu_0^{\frac{k-1}{k}}\right) / \left(\frac{k-1}{T}\mathcal{T}_k(t)+1\right)^2\right]^{\frac{k}{k-1}}$$

which implies that $\mu_k(t)\alpha_k(\mathcal{T}_k(t)) \to 0$. Using this fact, together with the inequality above and the boundedness of $\tilde{r}_k(t,j)\left|\eta_q + \delta_q\overline{d}_q^2(x(t,j))\right|$, allows us to conclude that $u_q \to 0$ as $t \to \Upsilon_{T,k}$. ■

Proof of Proposition 9.4

The following Lemma is instrumental in studying the stability properties of the HDS with data (9.35).

Lemma G.1. Consider the matrix

$$M_{\zeta_q}(x_1,\tau) \coloneqq \begin{pmatrix} \frac{1}{\eta(\tau)^2} I & I - \partial \mathcal{G}_q(x_1)^\top \\ I - \partial \mathcal{G}_q(x_1) & (\zeta_q - \rho\eta'(\tau))I \end{pmatrix}, \tag{G.19}$$

where $q \in \mathcal{Q}$, $\tau \in [0, N_0]$, $\eta(\tau) \in [\underline{\eta}, \overline{\eta}]$, $\rho \in [0, 1/\tau_d]$, and $\eta'(\tau) \coloneqq \frac{d\eta}{d\tau}(\tau)$, $\mathcal{G}_q(\cdot)$, and ζ_q are as introduced in Section 9.3.2. Suppose that Assumption 9.4 is satisfied. Then,

$$M_{\zeta_q}(x_1,\tau) \succeq \nu_M I, \quad \forall\ \tau \in [0, N_0],\ x_1 \in \mathbb{R}^n \tag{G.20}$$

where $\nu_M \coloneqq \frac{(1-\delta_d-\delta_\eta)\overline{\sigma}^2}{\delta_\eta(1-\delta_d)\underline{\zeta}+\overline{\sigma}^2}$, with $\underline{\zeta} \coloneqq \min_{q\in\mathcal{Q}} \zeta_q$ and $\overline{\sigma} \coloneqq \max_{q\in\mathcal{Q}} \sigma_q$. □

Proof. First we show that matrix-valued function $M_{\zeta_q}(\cdot,\cdot)$ is positive-definite uniformly over $\rho \in [0, \tau_d^{-1}]$, $x_1 \in \mathbb{R}^n$, and $\tau \in [0, N_0]$. To this end, we decompose the matrix $M_{\zeta_q}(x_1,\tau)$ as follows:

$$M_{\zeta_q}(x_1,\tau) = U_q(x_1,\tau) W_q(\tau,x_1) U_q(x_1,\tau)^\top, \tag{G.21a}$$

$$W_q(\tau,x_1) \coloneqq \begin{pmatrix} \frac{I}{\eta(\tau)^2} & 0 \\ 0 & \varrho_q(\tau) I - \eta^2(\tau)\Sigma_q(x_1)\Sigma_q(x_1)^\top \end{pmatrix}, \tag{G.21b}$$

$$\varrho_q(\tau) \coloneqq \zeta_q - \rho\eta'(\tau), \quad \Sigma_q(x_1) \coloneqq I - \partial\mathcal{G}_q(x_1), \tag{G.21c}$$

$$U_q(x_1,\tau) \coloneqq \begin{pmatrix} I & 0 \\ \eta^2(\tau)\Sigma_q(x_1) & I \end{pmatrix}. \tag{G.21d}$$

By the fact that $\eta(\tau) \in [\underline{\eta}, \overline{\eta}]$ for all $\tau \in [0, N_0]$ it follows that

$$\frac{1}{\eta(\tau)^2} I \succeq \frac{1}{\overline{\eta}^2} I. \tag{G.22}$$

Also, by Assumptions 9.4, we have that

$$\begin{aligned}\varrho_q(\tau)I - \eta(\tau)^2\Sigma_q(x_1)\Sigma_q(x_1) &\succeq \left(\varrho_q(\tau) - \overline{\eta}^2\sigma_q^2\right) I \\ &\succeq \left(\zeta_q - \frac{\overline{\eta} - \underline{\eta}}{\tau_d N_0} - \overline{\eta}^2\sigma_q^2\right) I \\ &\succeq \tilde{\delta} I, \end{aligned} \tag{G.23}$$

where $\tilde{\delta} \coloneqq (1-\delta)\underline{\zeta}$, with $\underline{\zeta} \coloneqq \min_{q\in\mathcal{Q}} \zeta_q$. Therefore, via [278, Theorem 7.7.7], the matrix $M_{\zeta_q}(x_1, \tau)$ is positive definite for all $x_1 \in \mathbb{R}^n$ and $\tau \in [0, N_0]$. Now, we establish the matrix inequality (G.20). To do so, we use (G.22) and (G.23) in (G.21a) to obtain that

$$\begin{aligned} M_{\zeta_q}(x_1, \tau) &\succeq U_q(x_1, \tau) \begin{pmatrix} \frac{1}{\overline{\eta}^2} I & 0 \\ 0 & \tilde{\delta} I \end{pmatrix} U_q^\top(x_1, \tau) \\ &\succeq Z_q(x_1, \tau) Z_q(x_1, \tau)^\top, \end{aligned} \tag{G.24}$$

where $Z_q(x_1, \tau)^\top$ is the upper block triangular matrix

$$Z_q(x_1, \tau)^\top \coloneqq \begin{pmatrix} \frac{1}{\overline{\eta}} I & \frac{\eta(\tau)^2}{\overline{\eta}} \Sigma_q(x_1, \tau)^\top \\ 0 & \sqrt{\tilde{\delta}} I \end{pmatrix}.$$

By applying [5, Lemma 9], and using (G.24) together with the fact that $Z_q(x_1, \tau)$ has full column rank for all $x_1 \in \mathbb{R}^n$ and $\tau \in [0, N_0]$ and thus that $\sigma_{\min}(Z_q(x_1, \tau) Z_q(x_1, \tau)^\top) \geq \sigma_{\min}(Z_q(x_1, \tau)) \sigma_{\min}(Z_q(x_1, \tau)^\top) = \sigma_{\min}^2(Z_q(x_1, \tau)^\top)$, we obtain

$$M_{\zeta_q}(x_1, \tau) \succeq \frac{1}{\overline{\eta}^2 \left(1 + \frac{\overline{\eta}^2}{\tilde{\delta}} \|\Sigma_q(x_1)^2\|\right) + \frac{1}{\tilde{\delta}}} I$$

$$\succeq \frac{(1-\delta_d-\delta_\eta)\overline{\sigma}^2}{\delta_\eta(1-\delta_d)\underline{\zeta}+\overline{\sigma}^2} I,$$

where in the last two steps we used Assumption 9.4. This completes the proof. ■

With the previous lemma we are now prepared to present the proof of Proposition 9.4.We show that, under Assumption 9.4, a suitable Lyapunov function for the "target" system $\widehat{\mathcal{H}}$ can be used to show that Assumption 9.3 is satisfied. Indeed, for every $\hat{q} \in \mathcal{Q}$ consider the Lyapunov function

$$V_{\hat{q}}(\hat{x},\hat{\tau}) = \frac{1}{4}|\hat{x}_2 - x^*|^2 + \frac{1}{4}|\hat{x}_2 - \hat{x}_1|^2 + \frac{\eta(\hat{\tau})^2}{2}\left|\mathcal{G}_{\hat{q}}(\hat{x}_1)\right|^2,$$

which in the flow set and jump set satisfies: $v_{\hat{q},1}|\hat{x}|^2_{\mathcal{A}_x} \leq V_{\hat{q}}(\hat{x},\hat{\tau}) \leq v_{\hat{q},2}|\hat{x}|^2_{\mathcal{A}_x}$, with $v_{\hat{q},1} \coloneqq 0.25\min\left\{1, 2\kappa^2_{\hat{q}}\underline{\eta}^2\right\}$, and $v_{\hat{q},2} \coloneqq 0.25\max\left\{3,\ 2+2\ell^2_{\hat{q}}\overline{\eta}^2\right\}$. Let

$$\mathcal{L}_{(f_{\hat{q}},\rho)}V_{\hat{q}}(\hat{x},\hat{\tau}) \coloneqq \left\langle \nabla V_{\hat{q}}(\hat{x},\hat{\tau}), \begin{pmatrix} f_{\hat{q}}(\hat{x},\hat{\tau}) \\ \rho \end{pmatrix} \right\rangle \tag{G.25}$$

Since $\mathcal{G}_{\hat{q}}(\cdot)$ is $\kappa_{\hat{q}}$-strongly-monotone and $\ell_{\hat{q}}-$Lipschitz, we have that $\langle x_1 - \tilde{x},\ \mathcal{G}_{\hat{q}}(\hat{x}_1)\rangle \geq \zeta_{\hat{q}}\left|\mathcal{G}_{\hat{q}}(\hat{x}_1)\right|^2$, where $\zeta_{\hat{q}} = \kappa^2_{\hat{q}}/\ell_{\hat{q}}$. During flows:

$$\begin{aligned}
\mathcal{L}_{(f_{\hat{q}},\rho)}V_{\hat{q}}(\hat{x},\hat{\tau}) &= -\frac{1}{\eta(\hat{\tau})}|\hat{x}_2 - \hat{x}_1|^2 \\
&\quad - 2\eta(\hat{\tau})\left\langle \mathcal{G}_{\hat{q}}(\hat{x}_1),\ \left[I - \partial\mathcal{G}_{\hat{q}}(\hat{x}_1)\right](\hat{x}_2 - \hat{x}_1)\right\rangle \\
&\quad - \eta(\hat{\tau})\left[\langle \hat{x}_1 - x^*,\ \mathcal{G}_{\hat{q}}(\hat{x}_1)\rangle - \rho\eta'(\hat{\tau})|\mathcal{G}_{\hat{q}}(\hat{x}_1)|^2\right] \\
&\leq -\eta(\hat{\tau})\left\langle \chi_{\hat{q}},\ M_{\zeta_{\hat{q}}}(\hat{x}_1,\hat{\tau})\chi_{\hat{q}}\right\rangle,
\end{aligned} \tag{G.26}$$

for all $(\hat{x},\hat{\tau},\rho) \in \mathbb{R}^{2n} \times [0, N_0] \times [0, \tau_d^{-1}]$, where $\chi_{\hat{q}} \coloneqq (\hat{x}_2 - \hat{x}_1, \mathcal{G}_{\hat{q}}(\hat{x}_1)) \in \mathbb{R}^{2n}$, and $M_{\zeta_{\hat{q}}}$ is

given by

$$M_{\zeta_{\hat{q}}}(\hat{x}_1, \hat{\tau}) \coloneqq \begin{pmatrix} \frac{1}{\eta(\hat{\tau})^2} I & I - \partial \mathcal{G}_{\hat{q}}(\hat{x}_1)^\top \\ I - \partial \mathcal{G}_{\hat{q}}(\hat{x}_1) & (\zeta_{\hat{q}} - \rho \eta'(\hat{\tau}))I \end{pmatrix}.$$

Using Lemma G.1, we conclude that $\mathcal{L}_{(f_{\hat{q}},\rho)} V_{\hat{q}}(\hat{x}, \hat{\tau}) \leq -\underline{\eta} \nu_M |\chi_{\hat{q}}|^2$ for all $(\hat{x}_1, \hat{x}_2, \hat{\tau}) \in \mathbb{R}^{2n} \times [0, N_0]$. Hence, by noting that $V_{\hat{q}}(\hat{x}, \hat{\tau}) \leq \frac{1}{4} \max\left\{3, 2\left(\frac{1}{\kappa_{\hat{q}}^2} + \overline{\eta}^2\right)\right\} |\chi_{\hat{q}}|^2$ we obtain:

$$\mathcal{L}_{(f_{\hat{q}},\rho)} V_{\hat{q}}(\hat{x}, \hat{\tau}) \leq -\frac{4\underline{\eta} \nu_M}{\max\left\{3, 2\left(\frac{1}{\kappa_{\hat{q}}^2} + \overline{\eta}^2\right)\right\}} V_{\hat{q}}(\hat{x}, \hat{\tau}). \tag{G.27}$$

Now, for all $\hat{p}, \hat{q} \in \mathcal{Q}$, let

$$\Delta V_{\hat{p}}^{\hat{q}}(\hat{x}, \hat{\tau}) \coloneqq V_{\hat{q}}\left(R_{\hat{p}}(\hat{x}), \hat{\tau} - 1\right) - V_{\hat{p}}(\hat{x}, \hat{\tau}), \quad \hat{\tau} \in [1, N_0].$$

During jumps:

$$\begin{aligned} &\Delta V_{\hat{p}}^{\hat{q}}(\hat{x}, \hat{\tau}) = V_{\hat{q}}\left((\hat{x}_1, \hat{x}_1), \hat{\tau} - 1\right) - V_{\hat{p}}(x, \hat{\tau}) \\ &\leq -\frac{1}{4}|\hat{x}_1 - x^*|^2 - \frac{1}{4}|\hat{x}_1 - \hat{x}_2|^2 + \frac{1}{4\kappa_{\hat{p}}^2}|\mathcal{G}_{\hat{p}}(\hat{x}_1)|^2 \\ &\quad + \frac{1}{2}\left(\eta(N_0 - 1)^2 \frac{\ell_{\hat{q}}^2}{\kappa_{\hat{p}}^2} - \eta(1)^2\right)|\mathcal{G}_{\hat{p}}(\hat{x}_1)|^2 \\ &\leq -\left(1 - \gamma_{\hat{p}}^{\hat{q}}\right) V_{\hat{p}}(\hat{x}, \hat{\tau}), \end{aligned} \tag{G.28}$$

where $\gamma_{\hat{p}}^{\hat{q}} \coloneqq \frac{2\eta(N_0-1)^2 \ell_{\hat{q}}^2 + 1}{2\kappa_{\hat{p}}^2 \eta(1)^2}$. The above inequality implies that

$$V_{\hat{q}}\left(R_{\hat{p}}(\hat{x}),\ \hat{\tau} - 1\right) \leq \gamma_{\hat{p}}^{\hat{q}} V_{\hat{p}}(\hat{x}, \hat{\tau}).$$

where $\underline{\ell} \coloneqq \min_{\hat{q} \in \mathcal{Q}} \ell_{\hat{q}}$, $\overline{\kappa} \coloneqq \max_{\hat{q} \in \mathcal{Q}} \kappa_{\hat{q}}$, and $\underline{\kappa} \coloneqq \min_{\hat{q} \in \mathcal{Q}} \kappa_{\hat{q}}$. Thus, noting that $\gamma_{\hat{p}}^{\hat{q}} \leq$

$\frac{\bar{\ell}^2}{\underline{\kappa}^2}\frac{\eta(N_0-1)^2}{\eta(1)^2}+\frac{1}{2\underline{\kappa}^2\eta(1)^2}=:\overline{\gamma}$, we obtain:

$$V_{\hat{q}}\left(R_{\hat{p}}(\hat{x}),\hat{\tau}-1\right)\leq\overline{\gamma}V_{\hat{p}}(\hat{x},\hat{\tau}), \tag{G.29}$$

for all $\tau\in[1,N_0],\ p,q\in\mathcal{Q}$. By the smoothness properties of $\mathcal{G}_q(\cdot)$ and the differentiability of $\eta(\cdot)$, we obtain that $f_q(x,\tau)$ is locally Lipschitz and, thus, that Assumption 9.2 also holds. On the other hand, note that via a simple change of coordinates, and without loss of generality, the results of Theorem 9.1 hold for $\mathcal{A}$ as defined in (9.23) but with the set $\{0\}$ replaced by the set $\mathcal{A}_x$ in (9.38). Therefore, the quadratic bounds on the Lyapunov function, together with condition (9.39), (G.27), and (G.29), imply PT-S_F of $\mathcal{A}_x\times[0,N_0]\times\mathcal{Q}\times\mathbb{R}_{\geq1}$ via Theorem 9.1-a). ■

G.4 Auxiliary Results

Lyapunov Conditions for Exponential-ISS of Hybrid Dynamical Systems

The following lemma is a specialization of [21, Prop. 2.7] for the case when the system is exponentially ISS. We present the complete proof here only for the purpose of completeness.

Lemma G.2. Consider the HDS (1.3), and a closed set $\mathcal{A}\subset\mathbb{R}^m$. Suppose there exist constants $\underline{\alpha},\overline{\alpha},\rho,p>0$, $\lambda\in(0,1)$, and a smooth function $V:C\cup D\to\mathbb{R}_{\geq0}$, such that the following inequalities hold:

$$\begin{aligned}
\underline{\alpha}|z|_{\mathcal{A}}^p\leq V(z)\leq\overline{\alpha}|z|_{\mathcal{A}}^p,\quad&\forall\ z\in C\cup D\cup G(D),\\
\langle\nabla V(z),F(z,u)\rangle\leq-\lambda V(z)+\rho|u|^p,\ \ &\forall\ (z,u)\in C\times\mathbb{R}^m,\\
V(G(z))-V(z)\leq-\lambda V(z)+\rho|u|^p,\ \ &\forall\ (z,u)\in D\times\mathbb{R}^m.
\end{aligned}$$

Then, every solution of (1.3) satisfies

$$|z(s,j)|_{\mathcal{A}} \leq \kappa_1 e^{-\kappa_2(s+j)}|z(0,0)|_{\mathcal{A}} + \kappa_3 \sup_{0\leq\tau\leq s}|u(\tau)|, \tag{G.30}$$

for all $(s,j) \in \mathrm{dom}(z)$, and where $\kappa_1 = (\overline{\alpha}/\underline{\alpha})^p$, $\kappa_2 = \lambda/2p$, and $\kappa_3 = \left(\frac{2\rho}{\lambda\underline{\alpha}}\right)^{1/p}$. □

Proof. We follow similar ideas as in the proof of [21, Prop. 2.7], but considering set-valued flow and jump maps. The proof has four main steps:

Step 1: First, note that for all $(z,u) \in (C \cup D) \times \mathbb{R}^m$:

$$-\lambda V(z) + \rho|u|^p \leq -\frac{\lambda}{2}V(z), \quad \text{if } V(z) \geq \frac{2\rho}{\lambda}|u|^p. \tag{G.31}$$

Therefore, whenever $V(z) \geq \frac{2\rho}{\lambda}|u|^p$ we have that

$$\langle \nabla V(z), F(z,u)\rangle \leq -\tilde{\lambda}V(z),\ \forall (z,u) \in C \times \mathbb{R}^m,$$
$$V(G(z)) - V(z) \leq -\tilde{\lambda}V(z),\ \forall (z,u) \in D \times \mathbb{R}^m,$$

where $\tilde{\lambda} \coloneqq \lambda/2$.

Step 2: For any $r \geq 0$, define $\gamma_{c_4}(r,s,j) = e^{-\tilde{\lambda}(s+j)}r$. We first show that when $V(z) \geq \frac{2\rho}{\lambda}|u|^p$, the function V evaluated along the solutions of (1.3) satisfies

$$V(z(s,j)) \leq \gamma_\lambda(V(z(0,0)),s,j), \quad \forall\ (s,j) \in \mathrm{dom}(z). \tag{G.32}$$

To establish this property, note that since $V(z(\cdot,\cdot))$ is not increasing during flows and jumps, if there is $(s',j') \in \mathrm{dom}(z)$ with $0 < s' + j' < t + j$ and such that $V(z(s',j')) = 0$, then we necessarily must have $V(z(\tilde{s},\tilde{j})) = 0$ for all $(\tilde{s},\tilde{j}) \in \mathrm{dom}(z)$ such that $s'+j' \leq \tilde{s}+\tilde{j} \leq s+j$, and (G.32) would hold for such times $(\tilde{s},\tilde{j})$. Suppose there is no $(s',j') \in \mathrm{dom}(z)$ with $0 < s' + j' < t + j$ such that $V(z(s',j')) = 0$. For

each $(s,j) \in \mathrm{dom}(z)$, we partition the hybrid time domain of z up to time (s,j) as $\mathrm{dom}(z) = \bigcup_{n=0}^{j}[s_n, s_{n+1}] \times \{n\}$, with $s_0 = 0$ and $s_{j+1} = s$. For any $n \in \{0,1,\ldots,j\}$, V satisfies

$$\int_{s_n}^{s_{n+1}} \frac{\overbrace{V(z(\tau,n))}^{\cdot}}{\tilde{\lambda}V(z(\tau,n))}d\tau \leq -\int_{s_n}^{s_{n+1}} d\tau = -(s_{n+1}-s_n).$$

Using the new variable $\varrho = V(z(\tau,n))$, we obtain $d\varrho = \dot{V}d\tau$ and the above integral can be written as

$$\int_{V(z(s_n,n))}^{V(z(s_{n+1},n))} \frac{d\varrho}{\tilde{\lambda}\varrho} \leq -(s_{n+1}-s_n). \tag{G.33}$$

Similarly, note that

$$\begin{aligned}\int_{V(z(s_{n+1},n))}^{V(z(s_{n+1},n+1))} \frac{d\varrho}{\tilde{\lambda}\varrho} &\leq \int_{V(z(s_{n+1},n))}^{V(z(s_{n+1},n+1))} \frac{d\varrho}{\tilde{\lambda}V(z(s_{n+1},n))} \\ &\leq -1,\end{aligned}$$

where the last inequality follows by the inequality $V(z(s,j+1)) - V(z(s,j)) \leq -\tilde{\lambda}V(z(s,j))$. Combining the above two inequalities, we obtain

$$\begin{aligned}\int_{V(z(0,0))}^{V(z(s,j))} \frac{d\rho}{\tilde{\lambda}\varrho} &= \sum_{n=0}^{j}\int_{V(z(s_n,n))}^{V(z(s_{n+1},n))} \frac{d\varrho}{\tilde{\lambda}\varrho} \\ &\quad + \sum_{n=1}^{j}\int_{V(z(s_{n+1},n))}^{V(z(s_{n+1},n+1))} \frac{d\varrho}{\tilde{\lambda}\varrho} \\ &\leq -\left(\sum_{n=0}^{j}(s_{n+1}-s_n) + \sum_{n=1}^{j}1\right) \\ &= -(s_{j+1}-s_0+j) = -(s+j).\end{aligned} \tag{G.34}$$

Integrating the left-hand side, we obtain $\frac{1}{\tilde{\lambda}}\ln\left(\frac{V(z(s,j))}{V(z(0,0))}\right) \leq -(s+j)$, from which we directly get

$$V(z(s,j)) \leq V(z(0,0))e^{-\frac{\lambda}{2}(s+j)} \tag{G.35}$$

Step 3: Let (z, u) be a maximal solution pair of (1.3). Define the set

$$\Omega \coloneqq \left\{ z \in \mathbb{R}^n : V(z) \leq \frac{2\rho}{\lambda} |u|_\infty^p \right\}. \tag{G.36}$$

For each $z_0 \in \mathbb{R}^n$, let

$$T_{z,u,z_0} \coloneqq \sup \Big\{ \tau \in \mathbb{R}_{\geq 0} : z(s,j) \notin \Omega,\ z(0,0) = z_0, \\ \forall\ (s,j) \in \operatorname{dom}(z),\ 0 \leq s + j \leq \tau \Big\}.$$

It follows that for all solutions of (1.3) with $z(0,0) = z_0$ and $(s,j) \in \operatorname{dom}(z)$ such that $0 \leq s + j < T_{z,u,z_0}$ we have that $V(z) > \frac{2\rho}{\lambda}|u|_\infty^p$, which, by Step 2, implies that V satisfies (G.35). Using the quadratic upper and lower bounds on V, we obtain:

$$|z(s,j)|_{\mathcal{A}} \leq \left(\frac{\overline{\alpha}}{\underline{\alpha}} \right)^{\frac{1}{p}} |z(0,0)|_{\mathcal{A}} e^{-\frac{\lambda}{2p}(s+j)}, \tag{G.37}$$

which holds for all $(s,j) \in \operatorname{dom}(z)$ such that $0 \leq s + j < T_{z,u,z_0}$.

Step 4: The last step is to prove forward invariance of Ω. Suppose there exist $(s', j') \in \operatorname{dom}(z)$ such that $z(s', j') \in \Omega$ and $(s', j' + 1) \in \operatorname{dom}(z)$. Since $\tilde{\lambda} < \lambda$, V satisfies

$$\begin{aligned} V(z(s', j' + 1)) &\leq (1 - \tilde{\lambda}) V(z(s', j')) + \rho |u|_\infty^p, \\ &\leq \left(1 - \frac{\lambda}{2} \right) \frac{2\rho}{\lambda} |u|_\infty^p + \rho |u|_\infty^p = \frac{2\rho}{\lambda} |u|_\infty^p. \end{aligned}$$

Moreover, if $(s', j' + 1) \in \operatorname{dom}(z)$, then z cannot leave Ω via flows because $\dot{V} \leq 0$ if $V(z) \geq \frac{2\rho}{\lambda}|u|_\infty^p$. It follows that for all $(s,j) \in \operatorname{dom}(z)$ such that $s + j \geq T_{z,u,z_0}$ the solution z satisfies:

$$\underline{\alpha} |z(s,j)|_{\mathcal{A}}^p \leq V(z(s,j)) \leq \frac{2\rho}{\lambda} |u|_\infty^p, \tag{G.38}$$

that is, $|z(s,j)|_{\mathcal{A}} \leq \left(\frac{2\rho}{\lambda\underline{\alpha}}\right)^{\frac{1}{p}} |u|_\infty$, for all $s+j \geq T_{z,u,z_0}$. Combining this bound with (G.37) we obtain

$$|z(s,j)|_{\mathcal{A}} \leq \max\left\{\left(\frac{\overline{\alpha}}{\underline{\alpha}}\right)^{\frac{1}{p}} |z(0,0)|e^{-\frac{\lambda}{2p}(s+j)}, \left(\frac{2\rho}{\lambda\underline{\alpha}}\right)^{\frac{1}{p}} |u|_\infty\right\}, \tag{G.39}$$

for all $(s,j) \in \mathrm{dom}(z)$. Since $\max\{a,b\} \leq a+b$, we obtain

$$|z(s,j)|_{\mathcal{A}} \leq \kappa_1 |z(0,0)|e^{-\kappa_2(s+j)} + \kappa_3 |u|_\infty, \tag{G.40}$$

with $\kappa_1 = \left(\frac{\overline{\alpha}}{\underline{\alpha}}\right)^{\frac{1}{p}}$, $\kappa_2 = \frac{\lambda}{2p}$ and $\kappa_3 = \left(\frac{2\rho}{\lambda\underline{\alpha}}\right)^{\frac{1}{p}}$. The result follows from the above inequality by time-invariance and causality. ■

The following result relaxes the third condition in Lemma G.2 under a standard average dwell-time condition on the jumps.

Lemma G.3. Consider the HDS (1.3), and suppose that: (a) every solution satisfies the ADT constraint (9.19); (b) there exist constants $\underline{\alpha}, \overline{\alpha}, \rho, p > 0$, $\lambda \in (0,1)$, and a smooth function $V : C \cup D \to \mathbb{R}_{\geq 0}$, such that the following inequalities hold:

$$\begin{aligned} \underline{\alpha}|z|_{\mathcal{A}}^p \leq V(z) \leq \overline{\alpha}|z|_{\mathcal{A}}^p, &\quad \forall\ z \in C \cup D \cup G(D), \\ \langle \nabla V(z), F(z,u)\rangle \leq -\lambda V(z) + \rho|u|^p, &\quad \forall\ (z,u) \in C \times \mathbb{R}^m, \\ V(G(z)) - V(z) \leq 0, &\quad \forall\ z \in D. \end{aligned}$$

Then, every solution of (1.3) satisfies

$$|z(s,j)|_{\mathcal{A}} \leq \kappa_1 e^{-\kappa_2(s+j)} |z(0,0)|_{\mathcal{A}} + \kappa_3 \sup_{0\leq\tau\leq s} |u(\tau)|, \tag{G.41}$$

for all $(s,j) \in \mathrm{dom}(z)$, where $\kappa_i > 0$, for $i \in \{1,2,3\}$.

Proof. The proof follows similar steps as the proof of Lemma G.2. In particular, inequality (G.33) still holds. On the other hand, during jumps, we now have

$$V(z(s_{n+1}, n+1)) - V(z(s_{n+1}, n)) \leq 0 \tag{G.42}$$

Dividing both sides by $\tilde{\lambda} V(z(s_{n+1}, n))$, we obtain

$$\begin{aligned} 0 &\geq \frac{V(z(s_{n+1}, n+1)) - V(z(s_{n+1}, n))}{\tilde{\lambda} V(z(s_{n+1}, n))} \\ &= \int_{V(z(s_{n+1}, n))}^{V(z(s_{n+1}, n+1))} \frac{d\varrho}{\tilde{\lambda} V(z(s_{n+1}, n))}. \end{aligned}$$

It follows that inequality (G.34) now becomes $\int_{V(z(0,0))}^{V(z(s,j))} \frac{d\rho}{\tilde{\lambda}\varrho} \leq -s$, from which we obtain after integration:

$$V(z(s, j)) \leq V(z(0, 0)) e^{-\frac{\lambda}{2} s} \tag{G.43}$$

Finally, the ADT condition (G.6) guarantees that $j \leq \frac{1}{\tau_d} s + N_0$ for any $(s, j) \in \mathrm{dom}(\hat{z})$, which implies that $s + j \leq (\frac{1}{\tau_d} + 1) s + N_0$. In turn, this inequality can be written as $s \geq \frac{\tau_d}{1+\tau_d}(s + j) - \frac{\tau_d}{1+\tau_d} N_0$, so that (G.43) can be upper-bounded as follows:

$$V(z(s, j)) \leq \kappa_7 e^{-\kappa_8 (s+j)} V(z(0, 0)), \tag{G.44}$$

where $\kappa_7 \coloneqq e^{\frac{\lambda}{2} \frac{\tau_d}{1+\tau_d} N_0}$ and $\kappa_8 \coloneqq \frac{\lambda}{2} \frac{\tau_d}{1+\tau_d}$. From here the proof follows the same Steps 3-4 from the proof of Lemma G.2. In particular, the inequality (G.40) now becomes

$$|z(s, j)|_{\mathcal{A}} \leq \tilde{\kappa}_1 |z(0, 0)| e^{-\tilde{\kappa}_2 (s+j)} + \tilde{\kappa}_3 |u|_\infty,$$

with $\tilde{\kappa}_1 \coloneqq \left(\frac{\overline{\alpha}}{\underline{\alpha}}\right)^{\frac{1}{p}} e^{\frac{\lambda}{2p} \frac{\tau_d}{1+\tau_d} N_0}$, $\tilde{\kappa}_2 \coloneqq \frac{\lambda}{2p} \frac{\tau_d}{1+\tau_d}$, and $\kappa_3 = \left(\frac{2\rho}{\lambda \underline{\alpha}}\right)^{\frac{1}{p}}$. ∎

Corollary G.1. Consider the normalized-by-μ_k BU-ODE of Lemma 9.3, $\frac{d\hat{\mu}_k}{ds} = \frac{k}{T}\hat{\mu}_k^{\frac{1}{k}}$. Then, for any $\ell > 0$ and any solution $\hat{\mu}_k$ to the ODE satisfying $\mu_k(0) = \mu_0 \geq 1$ the following bound holds:

$$\mu_1^{-\ell}(s) \leq e^{-\ell\frac{s}{T}}, \quad \forall s \in \mathbb{R}_{\geq 1},$$

when $k = 1$, and

$$\mu_k^{-\ell}(s) \leq \left(\frac{k-1}{T}s + 1\right)^{-\ell\frac{k}{k-1}}, \quad \forall s \in \mathbb{R}_{\geq 1},$$

when $k \in \mathbb{Z}_{\geq 2}$. □

Proof. We divide the proof into two cases.

Case $k = 1$: From Lemma 9.3, for $k = 1$, the solution to the normalized-by-μ_k BU-ODE is given by:

$$\hat{\mu}_k(s) = \mu_0 e^{\frac{s}{T}}.$$

It follows that $\mu_k^{-\ell}(s) = \mu_0^{-\ell}e^{-\frac{\ell}{T}s} \leq e^{-\frac{\ell}{T}s}$ for all $s \geq 0$, where we have used the fact that $\mu_0^{-\ell} \leq 1$ since $\mu_0 \geq 1$ and $\ell > 0$ by assumption.

Case $k > 1$: From Lemma 9.3, for $k > 1$, the solution to the normalized-by-μ_k BU-ODE is given by:

$$\hat{\mu}_k(s) = \left(\frac{k-1}{T}s + \mu_0^{\frac{k-1}{k}}\right)^{\frac{k}{k-1}}.$$

Using the fact that $\mu_0 \geq 1$ and that $(\cdot)^{\frac{k}{k-1}}$ is monotonically increasing in $\mathbb{R}_{\geq 0}$ for any $k > 1$, and thus preserves the order in $\mathbb{R}_{\geq 0}$, it follows that $\hat{\mu}_k(s) \geq \left(\frac{k-1}{T}s + 1\right)^{\frac{k}{k-1}}$. Therefore, we obtain: $\hat{\mu}_k^{-\ell}(s) \leq \left(\frac{k-1}{T}s + 1\right)^{-\ell\frac{k}{k-1}}$ for all $s \geq 0$. ■

Lemma G.4. Suppose that every solution pair $(\hat{z}, \hat{u})$ of the HDS (G.5) satisfies the bound (G.7) for all $(s, j) \in \mathrm{dom}(\hat{z})$. Assume that $\Delta(\hat{\mu}_k) = \hat{\mu}_k^{-\ell}$, where $\ell > 0$. Then, $(\hat{z}, \hat{u})$ satisfies the inequality

$$|\hat{z}(s,j)|_{\mathcal{A}} \leq \beta_k\Big(\bar{\kappa}_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\bar{\kappa}_2(s+j)} + \bar{\kappa}_3 |\hat{u}|_{(s,j)}, s\Big),$$

for all $(s, j) \in \mathrm{dom}(\hat{z})$, where $\bar{\kappa}_1 \coloneqq \kappa_1$, $\kappa_2 \coloneqq \frac{\kappa_2}{2}$, $\bar{\kappa}_3 \coloneqq 2\kappa_3$. Here $\beta_k(r, s) \in \mathcal{KL}$ is defined as $\beta_k(r, s) = r \cdot \max\{\kappa_1 e^{-\kappa_2 s},\ \xi_k^{-\ell}(s)\}$, $\xi_k(s) = \big(\frac{k-1}{T}s + 1\big)^{\frac{k}{k-1}}$ for all $k > 1$, and $\xi_1(s) = e^{\frac{s}{T}}$. □

Proof. Consider a complete solution pair $(\hat{z}, \hat{u})$ of the HDS (G.5) satisfying the bound (G.7). Then, we have that

$$|\hat{z}(s,j)|_{\mathcal{A}} \leq \kappa_1 e^{-\kappa_2(s+j)} |\hat{z}(0,0)|_{\mathcal{A}} + \kappa_3 \cdot \sup_{0 \leq \zeta \leq s} |\hat{\Delta}(\zeta)|, \tag{G.45}$$

for all $(s, j) \in \mathrm{dom}(\hat{z})$, and where $\hat{\Delta}(s) \coloneqq \Delta(\mu_k^{-\ell}(s))\hat{u}(s)$. Next, pick an arbitrary time $(\bar{s}, \bar{j}) \in \mathrm{dom}(\hat{z})$, and let $\hat{y}(r, k) \coloneqq \hat{z}(r + \bar{s}, k + \bar{j})$, and $v(r, k) \coloneqq \mu_k^{-\ell}(\bar{s} + r)$. Since $\hat{y}$ is also a hybrid arc that is a solution to (G.5), using the above bound and by time-invariance, it satisfies:

$$\begin{aligned}
|\hat{y}(r,k)|_{\mathcal{A}} &\leq \kappa_1 |\hat{y}(0,0)| e^{-\kappa_2(r+k)} + \kappa_3 |\hat{u}|_{(r,k)} |v|_{(r,k)} \\
&= \kappa_1 |\hat{z}(\bar{s}, \bar{j})| e^{-\kappa_2(r+k)} + \kappa_3 |\hat{u}|_{(r,k)} \sup_{0 \leq \tau \leq r} \hat{\mu}^{-\ell}(\bar{s} + \tau) \\
&\leq \kappa_1 |\hat{z}(\bar{s}, \bar{j})| e^{-\kappa_2(r+k)} + \kappa_3 |\hat{u}|_{(r,k)} \hat{\mu}_k^{-\ell}(\bar{s}).
\end{aligned} \tag{G.46}$$

Now, using (G.45) with $s = \bar{s}$ and $j = \bar{j}$, we obtain:

$$|\hat{z}(\bar{s}, \bar{j})|_{\mathcal{A}} \leq \kappa_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\kappa_2(\bar{s}+\bar{j})} + \kappa_3 |\hat{u}|_{(\bar{s},\bar{j})} \sup_{0 \leq \tau \leq r} \hat{\mu}_k^{-\ell}(\tau). \tag{G.47}$$

Combining (G.46) and (G.47), and using Remark 2, we have

$$\begin{aligned}|\hat{y}(r,k)|_{\mathcal{A}} \leq \kappa_1\Big(&\kappa_1|\hat{z}(0,0)|_{\mathcal{A}}e^{-\kappa_2(\bar{s}+\bar{j})} \\ &+ \kappa_3 \sup_{0\leq\tau\leq r}|\hat{u}(\tau)| \sup_{0\leq\tau\leq r}\hat{\mu}_k^{-\ell}(\tau)\Big)e^{-\kappa_2(r+k)} \\ &+ \kappa_3 \sup_{0\leq\tau\leq r}|\hat{u}(\tau)|\hat{\mu}_k^{-\ell}(\bar{s}).\end{aligned}$$

Evaluating the above bound at $r = \bar{s}$ and $\tilde{j} \in \mathbb{Z}_{\geq 0}$ such that $(\bar{s}, \tilde{j}) \in \mathrm{dom}(y)$, we obtain:

$$\begin{aligned}|\hat{y}(\bar{s},\tilde{j})|_{\mathcal{A}} &\leq \kappa_1\Big(\kappa_1|\hat{z}(0,0)|_{\mathcal{A}}e^{-\kappa_2(\bar{s}+\bar{j})} \\ &\quad + \kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)| \sup_{0\leq\tau\leq\bar{s}}\hat{\mu}_k^{-\ell}(\tau)\Big)e^{-\kappa_2(\bar{s}+\tilde{j})} \\ &\quad + \kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)|\hat{\mu}_k^{-\ell}(\bar{s}) \\ &\leq \kappa_1\Big(\kappa_1|\hat{z}(0,0)|_{\mathcal{A}}e^{-\kappa_2(\bar{s}+\bar{j})} + \kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)|\Big)e^{-\kappa_2(\bar{s}+\tilde{j})} \\ &\quad + \kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)|\hat{\mu}_k^{-\ell}(\bar{s}) \\ &\leq \Big(\kappa_1|\hat{z}(0,0)|_{\mathcal{A}}e^{-\kappa_2(\bar{s}+\bar{j}+\tilde{j})} \\ &\quad + 2\kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)|\Big)\max\left\{\kappa_1 e^{-\kappa_2\bar{s}}, \hat{\mu}_k^{-\ell}(\bar{s})\right\},\end{aligned}$$

where we used the fact that $e^{-\kappa_2\tilde{j}} \leq 1$, and $\sup_{0\leq\tau\leq\bar{s}}\hat{\mu}_k^{-\ell}(\tau) \leq \mu_0^{-\ell} \leq 1$ since $\mu_0 \geq 1$ and $\ell > 0$. Using the result of Corollary G.1 it then follows that

$$|\hat{y}(\bar{s},\tilde{j})|_{\mathcal{A}} \leq \Big(\kappa_1|\hat{z}(0,0)|_{\mathcal{A}}e^{-\kappa_2(\bar{s}+\bar{j}+\tilde{j})} + 2\kappa_3 \sup_{0\leq\tau\leq\bar{s}}|\hat{u}(\tau)|\Big)\eta_k(\bar{s})$$

where $\eta_k(s) \coloneqq \max\{\kappa_1 e^{-\kappa_2 s},\ \xi_k^{-\ell}(s)\}$, $\xi_k(s) = \left(\frac{k-1}{T}s + 1\right)^{-\frac{k}{k-1}}$ for all $k > 1$ and $\xi_1(s) = e^{-\frac{s}{T}}$. Note that η_k is continuous and satisfies $\eta_k(s) \to 0$ as $s \to \infty$ since $\kappa_1 e^{-\kappa_2 s} \to 0$ and $\xi_k(s) \to 0$ as $s \to \infty$. Now, using the definition of $\hat{y}$, and letting

$\lambda \coloneqq 2\bar{s}$, $i \coloneqq \tilde{j} + \bar{j}$:

$$|\hat{z}(\lambda, i)|_{\mathcal{A}} \leq \Big(\kappa_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\kappa_2(\frac{\lambda}{2}+i)} + 2\kappa_3 \sup_{0\leq\tau\leq\bar{s}} |\hat{u}(\tau)|\Big)\eta_k(\lambda/2)$$

Since the choice of $(\bar{s}, \bar{j}) \in \mathrm{dom}(z)$ was arbitrary, z is complete, and the previous inequality holds for all $\tilde{j} \in \mathbb{Z}_{\geq 0}$, in particular we can use $s = 2\bar{s}$, $j = \bar{j}$, and $\tilde{j} = 0$ such that $(s, j) \in \mathrm{dom}(z)$. Thus, from the above inequality and using Remark 2, we obtain that there exists $\beta_k(r, s) \coloneqq r \cdot \eta_k(s) \in \mathcal{KL}$ such that

$$|\hat{z}(s, j)|_{\mathcal{A}} \leq \beta_k\Big(\bar{\kappa}_1 |\hat{z}(0,0)|_{\mathcal{A}} e^{-\bar{\kappa}_2(s+j)} + \bar{\kappa}_3 |\hat{u}|_{(s,j)}, s\Big),$$

with $\bar{\kappa}_1 \coloneqq \kappa_1$, $\kappa_2 \coloneqq \frac{\kappa_2}{2}$, $\bar{\kappa}_3 \coloneqq 2\kappa_3$. ■

Appendix H

Proofs of Chapter 10

Since the stability results of the zeroth-order hybrid dynamics $\mathcal{H}_0$ in Theorem 10.1 rely on the stability properties of the first-order dynamics $\mathcal{H}_1$, we first present the proof of Theorem 10.2.

Proof of Theorem 10.2

We first present the proof of our auxiliary lemmas.

Proof of Lemma 10.1

Suppose that $\mathcal{S}$ is a δ-gap synergistic family of diffeomorphisms adapted to ϕ. Then, we have that $\delta < \mu(\mathcal{S})$, meaning that

$$\delta < \left(\tilde{\phi}_q - \min_{p \in \mathcal{Q}} \tilde{\phi}_p\right)(z) \quad \forall q \in \mathcal{Q},$$

and all $z \in \text{Crit}\ \tilde{\phi}_q \setminus \mathcal{A}$. Then, it follows that for all $q \in \mathcal{Q}$ and $z \in \text{Crit}\tilde{\phi}_q \setminus \mathcal{A}$, there exists $p \in \mathcal{Q}$ such that (10.16) is satisfied.

Conversely, assume that for every $q \in \mathcal{Q}$ and $z \in \text{Crit}\tilde{\phi}_q \setminus \mathcal{A}$, there exists $p \in \mathcal{Q}$ such that (10.16) is satisfied. In particular, for all $q \in \mathcal{Q}$ and $z \in \text{Crit}\tilde{\phi}_q \setminus \mathcal{A}$ it follows that

$\delta < (\tilde{\phi}_q - \min_{p\in\mathcal{Q}} \tilde{\phi}_p)(z)$, which implies that

$$\delta < \min_{\substack{q\in\mathcal{Q} \\ z\in \mathrm{Crit}\ \tilde{\phi}_q \setminus \mathcal{A}}} (\tilde{\phi}_q - \min_{p\in\mathcal{Q}} \tilde{\phi}_p)(z).$$

This concludes the proof. ■

Lemma H.1. The HDS $\mathcal{H}_1$ is well-posed.

Proof. We prove that $\mathcal{H}_1$ satisfies the hybrid-basic conditions [199, Def. 2.20]. First, note that the flow map F_1 is continuous, by the continuity of $\sum_{i=1}^n \nabla_{E_i}\tilde{\phi}_q(\cdot)E_i(\cdot)$ in M for all $q \in \mathcal{Q}$, and the fact that $\mathcal{Q}$ is a discrete set. Second, define the function $\tilde{\mu} : M \times \mathcal{Q} \to \mathbb{R}$ by letting $\tilde{\mu}(z, q) \coloneqq (\tilde{\phi}_q - m)(z)$. Note that $\tilde{\mu}$ is continuous by following similar reasoning as in the continuity argument for F_1. Then, $\mathrm{gph}\ h = \{(z, q) \in M \times \mathcal{Q} \ : \ z \in M,\ \tilde{\mu}(z, q) = 0\}$ is closed since $\tilde{\mu}$ is continuous. It follows that h and G_1 are outer-semicontinuous. Boundedness of G_1 follows by compactness of $M \times Q$ and outer-semicontinuity of G_1. Now, C_1 and D_1 are closed, since they are sublevel and superlevel sets, respectively, of the continuous function $\tilde{\mu}$. The result follows via [33, Thm. 6.30]. ■

Proof of Lemma 10.2

Let $\phi \in C^\infty(M)$ be arbitrary. Assume that $\boldsymbol{\nabla}\phi|_z = 0$ at some $z \in M$. Then, by the representation of $\boldsymbol{\nabla}\phi$ in terms of the global orthonormal frame $\{E_i\}_{i=1}^n$, it follows that: $\sum_{i,j=1}^n \zeta^{ij}(z)\nabla_{E_i}\phi(z)E_j(z) = 0$. Thus, since the matrix with coefficients $\zeta^{ij}(z)$ is nonsingular for all $z \in M$, as it provides a local representation of the Riemannian metric, and given that $\{E_i\}_{i=1}^n$ is a frame, we obtain:

$$\nabla_{E_i}\phi(z) = 0,\ \forall i \in \{1, \cdots, n\}, \tag{H.1}$$

which implies that $\sum_{i=1}^{n} \nabla_{E_i}\phi(z)E_i(z) = 0$. Conversely, assume that $\sum_{i=1}^{n} \nabla_{E_i}\phi(z)E_i(z) = 0$. Then, equation (H.1) holds, and thus $0 = \sum_{i,j=1}^{n} \zeta^{ij}(z)\nabla_{E_i}\phi(z)E_j(z) = \text{grad}\phi|_z$. This concludes the proof. ■

Now, we consider the set of critical points of the warped cost functions that are not minimizers of ϕ:

$$\mathcal{E} \coloneqq \left\{(z,q) \in M \times \mathcal{Q} \ : \ z \in \text{Crit } \tilde{\phi}_q \setminus \mathcal{A}\right\}.$$

The following lemma shows that $\mathcal{E}$ is properly contained in D_1, thus enforcing jumps whenever $(z,q) \in \mathcal{E}$. This means that the HDS $\mathcal{H}_1$ must jump at critical points that are not minimizers of ϕ. In the following, we use A° to denote the topological interior of a set A.

Lemma H.2. Suppose that Assumption 10.4 is satisfied. Then $\mathcal{E} \subsetneq D_1^\circ$ and $G_1(\mathcal{E}) \subsetneq C_1^\circ$.

Proof. First, the fact that $\delta > 0$ combined with the continuity and positive semidefiniteness of $(\tilde{\phi}_q - m)(\cdot)$ ensures the existence of open subsets of C_1 and D_1 where $0 < (\tilde{\phi}_q - m)(z) < \delta$ and $(\tilde{\phi}_q - m)(z) > \delta$, respectively, proving that C_1° and D_1° are non-empty. Second, consider $(z,q) \in \mathcal{E}$. Lemma 10.1 guarantees the existence of $p \in \mathcal{Q}$ such that $\tilde{\phi}_p(z) + \delta < \tilde{\phi}_q(z)$. Given that $m(z) \leq \tilde{\phi}_q(z)$ for all $p \in \mathcal{Q}$, we deduce: $m(z) + \delta < \tilde{\phi}_q(z)$, implying $(z,q) \in D_1$. Thus, $\mathcal{E} \subseteq D_1$. Now, consider $(z,q) \in D_1 \setminus D_1^\circ$, which implies $m(z) = \tilde{\phi}_q(z) - \delta$. Assume, for contradiction, that $(z,q) \in \mathcal{E}$. By Lemma 10.1, there exists $p \in \mathcal{Q}$ such that $\tilde{\phi}_p(z) < \tilde{\phi}_q(z) - \delta = m(z)$, contradicting $m(z) \leq \tilde{\phi}_q(z)$ for all $q \in \mathcal{Q}$. Hence, $(z,q) \notin \mathcal{E}$, proving that $D_1 \setminus D_1^\circ$ contains elements not in $\mathcal{E}$, and therefore tha $\mathcal{E} \subsetneq D_1^\circ$. The fact that $G_1(\mathcal{E}) \subsetneq C_1^\circ$ follows by construction, since after a jump we have that $\tilde{\phi}_{q^+}(z^+) - m(z^+) = \tilde{\phi}_{q^+}(z) - m(z) = 0 < \delta$. ■

By leveraging the results of the previous lemmas we can now prove the first main theorem of the paper.

Proof of Theorem 10.2

Consider the Lyapunov function:

$$V(x) \coloneqq \tilde{\phi}_q(z) - \underline{\phi}, \tag{H.2}$$

which is continuous due to similar arguments to the ones used to prove the continuity of F_1 in Lemma H.1. Since $\underline{\phi} < \phi(z)$ for all $z \notin \mathcal{A}$, together with (A_2) in Definition 10.3, we have that $\tilde{\phi}_q(z) - \underline{\phi} \geq 0$ for all $(z, q) \in M \times \mathcal{Q}$ and $\tilde{\phi}_q(z) - \underline{\phi} = 0$ if and only if $(z, q) \in \mathcal{A} \times \mathcal{Q}$. Therefore, it follows that $V(x) \geq 0$ for all $(z, q) \in M \times \mathcal{Q}$ and $V(x) = 0$ if and only if $z \in \mathcal{A}$. Now, during the flows of $\mathcal{H}_1$, the Lie-derivative of V satisfies

$$\begin{aligned} \mathcal{L}_{F_1} V(x) &= -\sum_{i=1}^{n} \nabla_{E_i} \tilde{\phi}_q(z) E_i(z) \left(\tilde{\phi}_q - \underline{\phi} \right) \\ &= -\sum_{i=1}^{n} (E_i(z) \tilde{\phi}_q)^2 =: u_C(x), \ \forall x \in C_1, \end{aligned} \tag{H.3}$$

where we used the fact that $\mathcal{L}_X f(z) = \nabla_X f(z) = (X(z))(f)$ for all $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, and that $v(c) = 0$ for all $v \in T_z M$, every constant function c, and all $z \in M$, via [256, Lemma 3.4]. On the other hand, during the jumps of $\mathcal{H}_1$, using the definition of h and m in (10.14) and (10.15), it follows that:

$$\begin{aligned} V\left(x^+\right) - V(x) &= \left(\tilde{\phi}_{h(z)} - \tilde{\phi}_q \right)(z) \\ &= \left(m - \tilde{\phi}_q \right)(z) \leq -\delta =: u_D(x), \end{aligned} \tag{H.4}$$

for all $x \in D_1$. Since $u_C(x) \leq 0$ for all $z \in C_1$ and $u_D(x) < 0$ for all $x \in D_1$, it follows that $\mathcal{A}$ is stable under $\mathcal{H}_1$ via [199, Thm. 3.19].

To show the attractivity of $\mathcal{A}$ we employ the hybrid invariance principle [199, Thm. 3.23]. Indeed, since $u_C(x) \leq 0$ for all $x \in C_1$ and $u_D(x) < 0$ for all $x \in D_1$, and using $u_D^{-1}(0) = \emptyset$,

given $r \in V(\mathcal{A} \cup \mathcal{E}) \subset [0, \overline{\phi} - \underline{\phi}]$, solutions approach the largest weakly invariant set in $V^{-1}(r) \cap ((\mathcal{A} \cup \mathcal{E}) \times \mathcal{Q})$. Let Ω denote such an invariant set and assume that $r \neq 0$. By the definition of $\mathcal{E}$ and the synergistic family of diffeomorphisms, it follows that $\Omega \subset \mathcal{E}$. Additionally, by Lemma H.2, we obtain that $\Omega \subset D_1^\circ$. Since $D_1^\circ \cap C_1 = \emptyset$ by construction, for Ω to be invariant under $\mathcal{H}_1$, we would need to have that $\Omega = G(\Omega)$, but this would imply, via Lemma H.2, that $\Omega \subset C_1^\circ$, and thus that $\Omega \subset C_1^\circ \cap D_1^\circ = \emptyset \implies \Omega = \emptyset$. Therefore, we must have that $r = 0$, and thus that $\forall (z(0), q(0)) \in M \times \mathcal{Q}$ solutions approach the largest weakly invariant set in $V^{-1}(0) \cap ((\mathcal{A} \cup \mathcal{E}) \times \mathcal{Q}) = \mathcal{A} \times \mathcal{Q}$, which is $\mathcal{A} \times \mathcal{Q}$ it self. UGAS follows directly by the global attractivity and stability of $\mathcal{A}$. ■

Proof of Theorem 10.1

The proof uses tools recently developed for averaging on compact Riemannian manifolds [252] together with the framework for hybrid extremum seeking control introduced in [111].

First, since M is compact, we can select $\varepsilon_a \in \mathbb{R}_{>0}$ such that $\exp_z(\varepsilon_a \mathcal{D}(z)) \in \imath(M)$, with $\imath(M)$ the injectivity radius of M [252, Lemma 3.2]. This makes possible a Taylor expansion in normal coordinates along the geodesic dithers for every $\tilde{\phi}_q$, such that the average dynamics of $\mathcal{H}_0$ can be computed to be (see [111]) $\mathcal{H}_0^A = \{C_1, F_0^A, D_1, G_1\}$, where C_1, D_1, G_1 are defined in (10.20a), (10.20b) and (10.19) respectively, and $F_0^A : M \times \mathcal{Q} \to TM \times \mathbb{N}$ is the *average flow map*, given by:

$$F_0^A(x) \coloneqq \begin{pmatrix} -\sum_{i=1}^n \nabla_{E_i} \tilde{\phi}_q(z) E_i(z) + \sum_{i=1}^n \mathcal{O}(\varepsilon_a) E_i(z) \\ 0 \end{pmatrix}.$$

Hence, on closed subsets of M we have that

$$F_0^A(x) \in \overline{\text{con}}_z F_1(x + k\varepsilon_a \mathbb{B}, 0) + (k\varepsilon_a \mathbb{B}, 0), \tag{H.5}$$

for some $k > 0$, where F_1 was defined in (10.18). Here, the convex hull affects the state z only, and the Minkowski additions $(z + k\varepsilon_a\mathbb{B})$ are defined in a suitable ambient Euclidean space which always exists due to the Whitney Embedding Theorem [256, Thm 6.15]. By (H.5), any solution of the average dynamics $\mathcal{H}_0^A$ is also a solution of an inflated HDS generated from $\mathcal{H}_1$. Hence, and since $\mathcal{H}_1$ is a well-posed HDS via Lemma H.1, by [33, Thm. 7.21] we conclude that system $\mathcal{H}_0^A$ renders the set $\mathcal{A}$ GP-AS in the ambient Euclidean space as $\varepsilon_a \to 0^+$. Since $\mathcal{H}_0^A$ and $\mathcal{H}_1$ are nominally well-posed, all conditions to apply [111, Cor. 1] are satisfied. Therefore, together with the compactness of M, $\mathcal{H}_0$ renders the set $\mathcal{A} \times \mathbb{T}^n$ GP-AS in the ambient Euclidean space as $(\varepsilon_d, \varepsilon_a) \to 0^+$. Note that any solution z to $\mathcal{H}_0$ is constrained to M since the dithering is performed along geodesics on the manifold, and $\hat{f}_q(z, \chi) \in T_zM$ for all $(z, q, \chi) \in M \times \mathcal{Q} \times \mathbb{T}^n$. Thus, we obtain GP-AS of $\mathcal{A}$ in the sense of Definition 10.2. ■

Proof of Lemma 10.3

First, we compute the differential of $\mathcal{S}_q^{(2)}$ and, whenever (10.27) is satisfied, we show that it is full rank for all $z \in \mathrm{SO}(3)$. When $\phi(z) \leq \gamma$, the differential is trivially full-rank since $d\left(\mathcal{S}_q^{(2)}\right)_z = I$. When $\phi(z) > \gamma$, we obtain:

$$d\left(\mathcal{S}_q^{(2)}\right)_z = e^{k_q\alpha(\phi(z)-\gamma)X}\left[k_q\alpha'(\phi(z)-\gamma)\,(X\cdot z)\,d\phi_z + I\right]. \tag{H.6}$$

Since the linear operator $v \mapsto e^{k_q\alpha(\phi(z)-\gamma)X}v$ is invertible, because $e^{rX} \in \mathrm{SO}(3)$ for all $r \in \mathbb{R}$ and $X \in T_I\mathrm{SO}(3)$, to prove that $d\left(\mathcal{S}_q^{(2)}\right)_z$ is full-rank it suffices to show that $(\Psi_z + I)$ is invertible, where $\Psi_z \coloneqq k_q\alpha'(\phi(z)-\gamma)\,(X\cdot z)\,d\phi_z$. To this end, letting $\|\Psi_z\|_z$ denote the operator 2-norm induced by the inner product in the Hilbert space $\mathcal{V}_z \coloneqq (T_z\mathrm{SO}(3),\ \langle\cdot,\cdot\rangle_z)$, we obtain that: $\|\Psi_z\|_z \leq |k_q||\alpha'(\phi(z)-\gamma)|\|X\|_F\|\mathrm{grad}\phi|_z\|_F$. Then, whenever (10.27) is satisfied, it follows that $\|\Psi_z\|_z < 1$, which implies that $(I + \Psi_z)$ is invertible, and hence that $d\left(\mathcal{S}_q^{(2)}\right)_z$ is full rank for all z such that $\phi(z) > \gamma$. By the inverse function theorem

[256, Thm. 4.5], it follows that $\mathcal{S}_q^{(2)}$ is a local diffeomorphism everywhere. Now, note that $\mathcal{S}_q^{(2)}$ is a proper map[1] since it is continuous and SO(3) is a compact Hausdorff space. This fact, together with the compactness of SO(3), implies that $\mathcal{S}_q^{(2)}$ is surjective via [281, Lemma. 1]. Injectivity of $\mathcal{S}_q^{(2)}$ follows from the arguments presented in [242, Appendix, Proof Thm. 8], which we omit here for conciseness. Since $\mathcal{S}_q^{(2)}$ is bijective, as well as a local diffeomorphism everywhere, it follows that it is a global diffeomorphism. ■

Proof of Theorem 10.3

The proof employs the same concepts as the proof of Theorem 10.1. We provide some details for completeness. Specifically, we now consider the Taylor expansion of the flow-map $\tilde{F}_0$ in normal coordinates and we analyze the corresponding average hybrid dynamics $\tilde{\mathcal{H}}_0^A = \{\tilde{C}_1, \tilde{F}_0^A, \tilde{D}_1, \tilde{G}_1\}$, describing the evolution of the state $\tilde{x} \coloneqq (z, q, p) \in \mathbb{S}^2 \times \mathcal{Q} \times \mathcal{P}$. In this case, the average flow-map $\tilde{F}_0^A(\cdot)$ is given by

$$\tilde{F}_0^A(\tilde{x}) = \left\{ \tilde{\mathcal{F}}_{1,qp}(z) + \sum_{i=1}^{2} \mathcal{O}(\varepsilon_a) E_{i,p}(z) \right\} \times \{0\} \times \{0\},$$

where $\tilde{\mathcal{F}}_{1,qp}(z) \coloneqq -\sum_{i=1}^{2} \nabla_{E_{i,p}} \tilde{\phi}_q(z) E_{i,p}(z)$. The flow set $\tilde{C}_1$, the jump set $\tilde{D}_1$, and the jump map $\tilde{G}_1$ are the same as the sets $\tilde{C}_0$, $\tilde{D}_0$, and them map $\tilde{G}_0$ defined in Section 10.3, but disregarding the state $\chi \in \mathbb{T}^2$ from the main state of the system. Using this construction, (H.5) becomes

$$\tilde{F}_0^A(\tilde{x}) \in \overline{\text{con}}_z \tilde{F}_1(\tilde{x} + k\varepsilon_a \mathbb{B}, 0) + (k\varepsilon_a \mathbb{B}, 0)\,, \tag{H.7}$$

where $k > 0$, and $\tilde{F}_1(\tilde{x}) \coloneqq \{\tilde{F}_{1,qp}(z)\} \times \{0\} \times \{0\}$. Furthermore, let $\tilde{\mathcal{H}}_1$ be the first-order HDS with data $\tilde{\mathcal{H}}_1 = \{\tilde{C}_1, \tilde{F}_1, \tilde{D}_1, \tilde{G}_1\}$, and consider the same Lyapunov function of (H.2). During the flows of $\tilde{\mathcal{H}}_1$, it follows that $\mathcal{L}_{\tilde{F}_1} V(\tilde{x}) = -\sum_{i=1}^{n} (E_{i,p}(z) \tilde{\phi}_q)^2$ for all $\tilde{x} \in \tilde{C}_1$.

[1] A map $f : A \to B$ is proper if the preimage of each compact subset of B is compact.

During the jumps of $\tilde{\mathcal{H}}_1$, the change of the Lyapunov function $\Delta V(\tilde{x}) \coloneqq V(\tilde{x}^+) - V(\tilde{x})$ satisfies: $\Delta V(\tilde{x}) = -\delta$ whenever $\tilde{x} \in \tilde{D}_{1,s} \coloneqq \{(z,q) \in \mathbb{S}^2 \times \mathcal{Q} \ : \ \left(\tilde{\phi}_q - m\right)(z) \geq \delta\} \times \mathcal{P}$, and $\Delta V(\tilde{x}) = 0$ whenever $\tilde{x} \in \tilde{D}_1 \setminus \tilde{D}_{1,s}$. In words, the Lyapunov function decreases whenever there is a switch between warped cost functions, denoted by a change in q, and does not increase if the system only switches between families of vector fields, i.e., *only* when the state p changes. Now, by the structure of the flow and jump sets, after a jump that only changes p is triggered, the system can either exhibit a change in q, after which it necessarily flows, or directly flows. The converse is true if a jump that only changes q is triggered first. Then, leveraging the decrease of the Lyapunov function during flows and employing a similar reasoning as in the proof of Theorem 10.2, it follows that every solution of $\tilde{\mathcal{H}}_1$ converges to the largest weakest invariant set in $V^{-1}(0) \cap ((\mathcal{A} \cup \mathcal{E}) \times \mathcal{Q} \times \mathcal{P})$, which is $\tilde{\mathcal{A}} \coloneqq \mathcal{A} \times \mathcal{Q} \times \mathcal{P}$ itself. It follows that $\tilde{\mathcal{A}}$ is UGAS under $\tilde{\mathcal{H}}_1$ via the hybrid invariance principle [199, Thm. 3.23]. The GP-AS of $\mathcal{A}$ under $\tilde{\mathcal{H}}_0$ is obtained by using (10.22), the well-posedness of $\tilde{\mathcal{H}}_1$ and $\tilde{\mathcal{H}}_0^A$, and applying the same arguments at the end of the proof of Theorem 10.1. ■

Proof of Lemma 10.4

Using the fact that $d\left(e^{A\eta(z)}\right)_z = e^{A\eta(z)}(I + Azd\eta_z)$ for $A \in \mathbb{R}^{n\times n}$ and $\eta : \mathbb{R}^n \to \mathbb{R}$ we get: $\det\left(d\left(S_q^{(3)}\right)_z\right) = 1 + k_q\alpha'\left(\phi(z) - \gamma\right) d\phi_z\left(Xz\right)$, for all z such that $\phi(z) > \gamma$. Thus, whenever $|k_q| < \overline{k}^{(3)}$, $\det\left(d(S_q^{(3)})_z\right) \neq 0$ for all $z \in \mathbb{S}^2$. Note that $\mathcal{S}_q^{(3)}$ is proper, being both continuous and defined on the compact space $\mathbb{S}^2$. Then, by the fact that $\mathbb{S}^2$ is simply connected and $\det(dS_q^{(3)}) \neq 0$, it follows that $S_q^{(3)}$ is a diffeomorphism via [282, Thm. B]. strut ■

Appendix I

Proofs of Chapter 11

Proof of Lemma 11.1

We find a sufficient condition on α to ensure that the relations in (11.1) are satisfied, and confirm that they are implied by the assumption of the Lemma. First, we have that

$$\begin{aligned}
\nabla U_q(\theta) &= -\alpha \sin^2(\theta) + \frac{1}{2}\sin(\theta+\phi)\cos(\theta+\phi) \\
&= -\alpha \sin^2(\theta) + \frac{1}{4}\sin(2\,[\theta+\phi]) \\
&= \alpha^2 \sin(\theta) + \frac{1}{2}\sin(\theta+\varphi(q))\cos(\theta+\varphi(q)) \\
&= \alpha^2 \sin(\theta) + \frac{1}{4}\sin\left(2\,(\theta+\varphi(q))\right) \\
&= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos(2\varphi(q)) + \sin(2\varphi(q))\cos(2\theta)\right]
\end{aligned}$$

First, recall the sufficient condition to have that $(\theta_1, \theta_2) = (\pm\pi, 0)$ is locally exponentially stable. To do so, note that around that point the only solutions allowed to flow have $q = 0$. Hence, computing Jacobian of $F_{av}(x, 0)$ around $(\pm\pi, 0)$ we obtain

$$\frac{\partial F_{av}(x,0)}{\partial x}[(\pm\pi, 0)] = \begin{pmatrix} 0 & 1 \\ \alpha^2 - \frac{1}{2} & -\alpha\beta \end{pmatrix}$$

With $\alpha^2 < \frac{1}{2}$, and $\beta > 0$, it follows that the matrix is Hurwitz and thus that the point is locally exponentially stable. From now on, we assume that $\alpha^2 \in (0, 1/2)$ such that at least local exponential stability of $\pm\pi$ is attained.

First, we prove that $\nabla U_q(\theta)\mathbf{1}_{\{\theta\in\Theta_q\}} > 0 \ \forall\theta \in (-\pi, 0]$.

- $\underline{q = 0, \theta \in \Theta_0 \cap \{\theta < 0\}, \varphi(q) = 0}$:

$$\begin{aligned}\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos(0) + \cos(2\theta)\sin(0)\right] \\ &= \alpha^2 \sin(\theta) + \frac{1}{2}\sin(\theta)\cos(\theta) \\ &= \sin(\theta)\left(\alpha^2 + \frac{1}{2}\cos(\theta)\right)\end{aligned}$$

$(\theta \in \Theta_0 \cap \{\theta < 0\} \implies \sin(\theta) < 0)$

$$\nabla U_q(\theta) > 0 \impliedby \alpha^2 < -\frac{1}{2}\max_{\theta\in\Theta_0\cap\{\theta<0\}}\cos(\theta) = \frac{1}{2}\cos(3\pi/8)$$

- $\underline{q = -1, \theta \in \Theta_{-1} = [-\pi/8, 0], \varphi(q) = \pi/4}$

$$\begin{aligned}\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos\left(\frac{\pi}{2}\right) + \cos(2\theta)\sin\left(\frac{\pi}{2}\right)\right] \\ &= \alpha^2 \sin(\theta) + \frac{1}{4}\cos(2\theta) \\ &= \alpha^2 \sin(\theta) + \frac{1 - 2\sin^2(\theta)}{4} \\ &= \frac{1}{4} + \alpha^2 \sin(\theta) - \frac{1}{2}\sin^2(\theta) \\ &= -\frac{1}{2}\left(\sin^2(\theta) - 2\alpha^2\sin(\theta) - \frac{1}{2}\right) \\ &= -\frac{1}{2}\left((\sin(\theta) - \alpha^2)^2 - \alpha^4 - \frac{1}{2}\right) \\ &\geq -\frac{1}{2}\max_{[-\pi/8,0]}(\sin(\theta) - \alpha^2)^2 + \frac{1}{2}\left(\alpha^4 + \frac{1}{2}\right) \\ &= -\frac{1}{2}\left(-\sin(\pi/8) - \alpha^2\right)^2 + \frac{1}{2}\left(\alpha^4 + \frac{1}{2}\right)\end{aligned}$$

$$= -\frac{1}{2}\sin^2(\pi/8) - \sin(\pi/8)\alpha^2 + \frac{1}{4}$$
$$\nabla U_q(\theta) > 0 \impliedby \alpha^2 < \frac{\frac{1}{2} - \sin^2\left(\frac{\pi}{8}\right)}{2\sin(\pi/8)}$$

- $\underline{q = -2, \theta \in \Theta_{-2} = [-3\pi/8, -\pi/8], \varphi(q) = \pi/2}$

$$\begin{aligned}\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos(\pi) + \cos(2\theta)\sin(\pi)\right] \\ &= \alpha^2 \sin(\theta) - \frac{1}{4}\sin(2\theta) \\ &= \alpha^2 \sin(\theta) - \frac{1}{2}\sin(\theta)\cos(\theta) \\ &= -\sin(\theta)\left(\frac{1}{2}\cos(\theta) - \alpha^2\right)\end{aligned}$$

$(\theta \in \Theta_{-2} \implies -\sin(\theta) > 0)$

$$\nabla U_q(\theta) > 0 \impliedby \alpha^2 < \frac{1}{2}\min_{\theta\in\Theta_{-2}}\cos(\theta) = \frac{1}{2}\cos(3\pi/8)$$

- $\underline{q = -3, \theta \in \Theta_{-3} = [-5\pi/8, -3\pi/8], \varphi(q) = 3\pi/4}$,

$$\begin{aligned}\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos\left(\frac{3\pi}{2}\right) + \cos(2\theta)\sin\left(\frac{3\pi}{2}\right)\right] \\ &= \alpha^2 \sin(\theta) - \frac{1}{4}\cos(2\theta) \\ &= \alpha^2 \sin(\theta) - \frac{1 - 2\sin^2(\theta)}{4} \\ &= -\frac{1}{2}\left(\frac{1}{2} - 2\alpha^2\sin(\theta) - \sin^2(\theta)\right) \\ &= -\frac{1}{2}\left(\frac{1}{2} + \alpha^4 - \left(\alpha^2 + \sin(\theta)\right)^2\right) \\ &\geq \frac{1}{2}\left(-\frac{1}{2} - \alpha^4 + \min_{\theta\in\Theta_{-3}}\left(\alpha^2 + \sin(\theta)\right)^2\right) \\ &= \frac{1}{2}\left(-\frac{1}{2} - \alpha^4 + \left(\alpha^2 + \sin(-3\pi/8)\right)^2\right) \\ &= \frac{1}{2}\left(-\frac{1}{2} + 2\alpha^2\sin(-3\pi/8) + \sin^2(-3\pi/8)\right)\end{aligned}$$

$$= \frac{1}{2}\left(-\frac{1}{2} - 2\alpha^2 \sin(3\pi/8) + \sin^2(3\pi/8)\right)$$

$$\nabla U_q(\theta) > 0 \impliedby \alpha^2 < \frac{\sin^2(3\pi/8) - \frac{1}{2}}{2\sin(3\pi/8)}.$$

We now prove that $\nabla U_q(\theta)\mathbf{1}_{\{\theta\in\Theta_q\}} < 0 \ \forall\theta \in [0,\pi)$.

- $\underline{q = 0, \theta \in \Theta_0 \cap \{\theta > 0\}, \varphi(q) = 0}$

$$\begin{aligned}
\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos\left(\frac{3\pi}{2}\right) + \cos(2\theta)\sin\left(\frac{3\pi}{2}\right)\right] \\
&= -\frac{1}{2}\left(\frac{1}{2} + \alpha^4 - \left(\alpha^2 + \sin(\theta)\right)^2\right) \\
&\leq \frac{1}{2}\left(-\frac{1}{2} - \alpha^4 + \max_{\theta\in\Theta_1}\left(\alpha^2 + \sin(\theta)\right)^2\right) \\
&= \frac{1}{2}\left(-\frac{1}{2} - \alpha^4 + \left(\alpha^2 + \sin(\pi/8)\right)^2\right) \\
&= \frac{1}{2}\left(-\frac{1}{2} + \sin(\pi/8)^2 + 2\sin(\pi/8)\alpha^2\right) \\
\nabla U_q(\theta) < 0 &\impliedby \alpha^2 < \frac{\frac{1}{2} - \sin^2\left(\frac{\pi}{8}\right)}{2\sin(\pi/8)}
\end{aligned}$$

- $\underline{q = 2, \theta \in \Theta_2 = [\pi/8, 3\pi/8], \varphi(q) = \pi/2}$

$$\begin{aligned}
\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos(\pi) + \cos(2\theta)\sin(\pi)\right] \\
&= \alpha^2 \sin(\theta) - \frac{1}{4}\sin(2\theta) \\
&= \alpha^2 \sin(\theta) - \frac{1}{2}\sin(\theta)\cos(\theta) \\
&= \sin(\theta)\left(\alpha^2 - \frac{1}{2}\cos(\theta)\right)
\end{aligned}$$

$(\theta \in \Theta_2 \implies \sin(\theta) > 0)$

$$\leq \sin(\theta)\left(\alpha^2 - \frac{1}{2}\min_{\theta\in\Theta_2}\cos(\theta)\right)$$

$$\nabla U_q(\theta) < 0 \iff \alpha^2 < \frac{1}{2} \min_{\theta \in \Theta_2} \cos(\theta) = \frac{1}{2} \cos(3\pi/8)$$

- $\underline{q = 3, \theta \in \Theta_3 = [3\pi/8, 5\pi/8], \varphi(q) = \pi/4}$

$$\begin{aligned}
\nabla U_q(\theta) &= \alpha^2 \sin(\theta) + \frac{1}{4}\left[\sin(2\theta)\cos\left(\frac{\pi}{2}\right) + \cos(2\theta)\sin\left(\frac{\pi}{2}\right)\right] \\
&= \alpha^2 \sin(\theta) + \frac{1}{4}\cos(2\theta) \\
&= \alpha^2 \sin(\theta) + \frac{1 - 2\sin^2(\theta)}{4} \\
&= \frac{1}{4} + \alpha^2 \sin(\theta) - \frac{1}{2}\sin^2(\theta) \\
&= \frac{1}{2}\left(-\sin^2(\theta) + 2\alpha^2 \sin(\theta) + \frac{1}{2}\right) \\
&= \frac{1}{2}\left(-(\sin(\theta) - \alpha^2)^2 + \alpha^4 + \frac{1}{2}\right) \\
&\leq -\frac{1}{2} \min_{\theta \in \Theta_3} (\sin(\theta) - \alpha^2)^2 + \frac{1}{2}\left(\alpha^4 + \frac{1}{2}\right) \\
&= -\frac{1}{2}\left(\sin(3\pi/8) - \alpha^2\right)^2 + \frac{1}{2}\left(\alpha^4 + \frac{1}{2}\right) \\
&= -\frac{1}{2}\sin^2(3\pi/8) + \alpha^2 \sin(3\pi/8) + \frac{1}{4} \\
\nabla U_q(\theta) < 0 &\iff \alpha^2 < \frac{\sin^2(3\pi/8) - \frac{1}{2}}{2\sin(3\pi/8)}
\end{aligned}$$

Now, for any α we have that $\nabla U_0(\pm\pi) = -\alpha^2 \sin(\pm\pi) + \frac{1}{2}\sin\left(\pm\pi + \frac{\pi}{2}\right)\cos\left(\pm\pi\frac{\pi}{2}\right)$. Hence, provided that

$$\alpha^2 < \min\left\{\frac{1}{2}\cos(3\pi/8), \frac{\frac{1}{2} - \sin^2(\pi/8)}{2\sin(\pi/8)}, \frac{\sin^2(3\pi/8) - \frac{1}{2}}{2\sin(3\pi/8)}, \frac{1}{2}\right\} = \frac{1}{2}\cos(3\pi/8),$$

it follows that

$$\begin{aligned}
&\nabla U(x_1, q) = 0 \iff x_1 \in \{-\pi, \pi\}, \text{ and} \\
&\langle U(x_1, q),\, x_1 \rangle < 0 \quad \forall (x_1, q) \in \bigcup_{q \in Q} (\Theta_q \setminus \{-\pi, 0, \pi\}) \times \{q\}
\end{aligned}$$

■

Proof of Lemma 11.2

First, note that for every q we have that $\Theta_q \subset \mathbb{R}$ is a closed interval and $K_q^{\uparrow}$ is a finite set and hence closed. Therefore $\Theta_q \times \mathbb{R} \times \{q\} \times K_q^{\uparrow}$ is closed since finite product of closed sets is closed. Then, C is a finite union of closed sets which implies that $C \times \mathbb{S}^1$ is also closed.

Now, we have that $F_x(\cdot, \cdot, \varphi(\cdot))$ is continuous in $\mathbb{R}^2 \times Q \times \mathbb{S}^1$ since it is a composition of the continuous function $F_x : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{S}^1 \to \mathbb{R}^2$ and $\varphi : Q \to \mathbb{R}$, which is continuous since Q is finite. Continuity of F in $C \cup D \subset \mathbb{R}^2 \times \mathbb{R} \times \mathbb{S}^1$ follows directly by recalling that F_χ is continuous and noting that Cartesian product of continuous functions is continuous. Moreover, since F is continuous and single valued we obtain that it is locally bounded.

On the other hand, note that $\Theta_q \cap \Theta_{(q+\text{sign}(q)) \bmod 4} \neq \emptyset$ by design. Moreover, we have that $\{0\}$ is closed since it is a singleton set, and that $\Theta_q \cap \Theta_p$ is closed for all $p, q \in Q$, due to intersection of closed sets being closed. Consequently, ϑ_q is closed for every $q \in Q$. Noting that $D_p^{\uparrow}$ and $D_p^{\downarrow}$ are closed for every p, it follows that D is a finite union of the closed sets $\vartheta_q \times \left(D_p^{\uparrow} \cup D_p^{\downarrow}\right)$ and thus that $D \times \mathbb{S}^1$ is closed.

Finally, let $\nu : Q + \{-1, 1\} \to Q$ be defined by $\nu(\hat{q}) = \hat{q} \bmod 4$, and $\xi : Q \times R \to Q + \{-1, 1\}$ be given by $\xi(q, s) = q + \overline{\text{sign}}(s)$. It follows that $\nu(\cdot)$ is continuous, since $Q + \{-1, 1\}$ is finite, and thus locally bounded and outer-semicontinuous when seen as a set-valued map. Moreover, $\xi(\cdot, \cdot)$ is outer-semicontinous and locally bounded via [52, Prop. 5.51] due to $q \to \{q\}$ and $\overline{\text{sign}}(\cdot)$ being outer-semicontinuous and locally bounded. Thus, by [52, Prop. 5.52], $\nu \circ \xi : Q \times \mathbb{R} \rightrightarrows Q,\ (q, x_2) \mapsto (q + \text{sign}(x_2)) \bmod 4$, is outer-semicontinuous and locally bounded. Additionally, we have that $\overline{s}$ is outer-semicontinuous and locally bounded by construction. This, together with the continuity of $x_1 \mapsto x_1$, $x_2 \mapsto x_2$, and $\chi \mapsto \chi$, implies the outer-semicontinuity and locally boundedness of G.

Therefore, $\mathcal{H}$ satisfies the hybrid basic conditions [33, Asm. 6.5] which suffices to conclude that it is well posed [33, Thm. 6.30.]. ■

Proof of Lemma 11.3

Let $y = (x, q, k)$ such that the state of $\mathcal{H}$ can be written as $z = (y, \chi)$. Since χ does not play any role during jumps, to prove the lemma it suffices to focus on the discrete-time evolution of y. The discrete-time dynamics for y coincide with the average jump map:

$$G_{av}(y) = \begin{cases} \begin{pmatrix} x_1 \\ x_2 \\ \overline{\text{sign}}(x_2) \\ -\text{sign}(q) \end{pmatrix} & \text{if } y \in (D \cap (\{\vartheta_0\} \times \mathbb{R} \times Q \times \mathcal{K})) \\ \begin{pmatrix} x_1 \\ x_2 \\ \left(q + \overline{\text{sign}}(x_2)\right) \mod 4 \\ q + \text{sign}(q) \end{pmatrix} & \text{if } y \in \begin{pmatrix} D \cap \Big(\Big\{ x_1 \in \bigcup_{q \in Q \setminus \{0\}} \vartheta_q \Big\} \\ \times \mathbb{R} \times [Q \setminus \{0\}] \times \mathcal{K} \Big) \Big) \end{pmatrix} \\ \begin{pmatrix} x_1 \\ x_2 \\ \overline{s}(x_2) \\ q + \text{sign}(q) \end{pmatrix} & \text{if } y \in (D \cap (\{\vartheta_3\} \times \mathbb{R} \times \{0\} \times \mathcal{K})) \\ \begin{pmatrix} x_1 \\ x_2 \\ -\overline{s}(x_2) \\ q + \text{sign}(q) \end{pmatrix} & \text{if } y \in (D \cap (\{\vartheta_{-3}\} \times \mathbb{R} \times \{0\} \times \mathcal{K})) \end{cases} . \tag{I.1}$$

Now, note that D is constructed as the union of the sets $\left\{\vartheta_q \times \left(D_q^{\uparrow} \times D_q^{\downarrow}\right)\right\}_{q\in Q}$. Since $p \neq q \implies \vartheta_p \neq \vartheta_q$, by design, and it is always true that $x_1^+ = x_1$ during jumps regardless of the value of x_1, it follows that $G_{av}\left(\vartheta_q \times \left(D_q^{\uparrow} \times D_q^{\downarrow}\right)\right) \cap \vartheta_p \times \left(D_p^{\uparrow} \times D_p^{\downarrow}\right) = \emptyset$ whenever $p \neq q$. Thus, to show that solutions to $\mathcal{H}$ experience at most 2 consecutive jumps in-between intervals of flow, we only need to study the intersection of $G_{av}\left(\vartheta_q \times \left(D_q^{\uparrow} \times D_q^{\downarrow}\right)\right)$ with $(\{\vartheta_q\} \times \mathbb{R} \times Q \times \mathcal{K}) \cap (C \cup D)$.

Thus, for $y \in D$ such that $x_1 = \vartheta_0 = 0$, we have that

$$\begin{aligned}
G_{av}\left(\vartheta_0 \times D_0^{\uparrow}\right) &= G_{av}\left(\vartheta_0 \times \mathbb{R}_{\geq 0} \times \{-1\} \times \{-3\}\right) \\
&= G_{av}\left(\vartheta_0 \times (\{0\} \cup \mathbb{R}_{>0}) \times \{-1\} \times \{-3\}\right) \\
&= G_{av}\left([\vartheta_0 \times \{0\} \times \{-1\} \times \{-3\}] \cup [\vartheta_0 \times \mathbb{R}_{>0} \times \{-1\} \times \{-3\}]\right) \\
&= G_{av}\left(\vartheta_0 \times \{0\} \times \{-1\} \times \{-3\}\right) \cup G_{av}\left(\vartheta_0 \times \mathbb{R}_{>0} \times \{-1\} \times \{-3\}\right) \\
&= \left(\vartheta_0 \times \{0\} \times \{-1, 1\} \times \{1\}\right) \cup \left(\vartheta_0 \times \mathbb{R}_{>0} \times \{1\} \times \{1\}\right) \\
&= \vartheta_0 \times \left([\{0\} \times \{-1, 1\} \times \{1\}] \cup [\mathbb{R}_{>0} \times \{1\} \times \{1\}]\right) \qquad \text{(I.2)}
\end{aligned}$$

We consider the intersection of (I.2) with the part of the jump set D for which $\pi_{x_1}(D) = \vartheta_0$:

$$\begin{aligned}
&G_{av}\left(\vartheta_0 \times D_0^{\uparrow}\right) \cap \\
&\left(\vartheta_0 \times \left[D_0^{\uparrow} \cup D_0^{\downarrow}\right]\right) = \vartheta_0 \times \left[\left([\{0\} \times \{-1, 1\} \times \{1\}] \cup [\mathbb{R}_{>0} \times \{1\} \times \{1\}]\right) \cap D_0^{\uparrow}\right] \\
&\qquad\qquad \cup \vartheta_0 \times \left[\left([\{0\} \times \{-1, 1\} \times \{1\}] \cup [\mathbb{R}_{>0} \times \{1\} \times \{1\}]\right) \cap D_0^{\downarrow}\right] \\
&\qquad = \vartheta_0 \times \left[\left(\{0\} \times \{-1, 1\} \times \{1\}\right) \cap D_0^{\uparrow}\right] \\
&\qquad\qquad \cup \vartheta_0 \times \left[\left(\mathbb{R}_{>0} \times \{1\} \times \{1\}\right) \cap D_0^{\uparrow}\right] \\
&\qquad\qquad\qquad \cup \vartheta_0 \times \left[\left(\{0\} \times \{-1, 1\} \times \{1\}\right) \cap D_0^{\downarrow}\right] \\
&\qquad\qquad\qquad\qquad \cup \vartheta_0 \times \left[\left(\mathbb{R}_{>0} \times \{1\} \times \{1\}\right) \cap D_0^{\downarrow}\right] \\
&\qquad = \vartheta_0 \times \left[\left(\{0\} \times \{-1, 1\} \times \{1\}\right) \cap D_0^{\uparrow}\right]
\end{aligned}$$

$$
\begin{aligned}
&\cup\, \vartheta_0 \times [\mathbb{R}_{>0} \times \emptyset \times (\{1\} \cap \{-3\})] \\
&\quad \cup\, \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{1\}) \cap D_0^{\downarrow}\right] \\
&\quad\quad \cup\, \vartheta_0 \times [\emptyset \times \{1\} \times (\{1\} \cap \{3\})] \\
&= \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{1\}) \cap D_0^{\uparrow}\right] \\
&\quad \cup\, \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{1\}) \cap D_0^{\downarrow}\right] \\
&= \vartheta_0 \times [\{0\} \times \{-1\} \times (\{1\} \cap \{-3\})] \\
&\quad \cup\, \vartheta_0 \times [\{0\} \times \{1\} \times (\{1\} \cap \{3\})] \\
&= \vartheta_0 \times [\{0\} \times \{-1\} \times \emptyset] \\
&\quad \cup\, \vartheta_0 \times [\{0\} \times \{1\} \times \emptyset] \\
&= \emptyset
\end{aligned} \tag{I.3}
$$

$$
G_{av}\left(\vartheta_0 \times D_0^{\downarrow}\right) = \vartheta_0 \times \left(\left[\, \{0\} \times \{-1,1\} \times \{-1\} \,\right] \cup [\mathbb{R}_{<0} \times \{-1\} \times \{-1\}]\right)
$$

$$
\Longrightarrow
$$

$$
\begin{aligned}
&G_{av}\left(\vartheta_0 \times D_0^{\downarrow}\right) \cap \\
&\left(\vartheta_0 \times \left[D_0^{\uparrow} \cup D_0^{\downarrow}\right]\right) = \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\uparrow}\right] \\
&\qquad \cup\, \vartheta_0 \times \left[(\mathbb{R}_{<0} \times \{-1\} \times \{-1\}) \cap D_0^{\uparrow}\right] \\
&\qquad\quad \cup\, \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\downarrow}\right] \\
&\qquad\quad\quad \cup\, \vartheta_0 \times \left[(\mathbb{R}_{<0} \times \{-1\} \times \{-1\}) \cap D_0^{\downarrow}\right] \\
&\quad = \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\uparrow}\right] \\
&\qquad \cup\, \vartheta_0 \times [\emptyset \times \{-1\} \times (\{-1\} \cap \{-3\})] \\
&\qquad\quad \cup\, \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\downarrow}\right] \\
&\qquad\quad\quad \cup\, \vartheta_0 \times [\mathbb{R}_{<0} \times \emptyset \times (\{-1\} \cap \{3\})]
\end{aligned}
$$

$$
\begin{aligned}
&= \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\uparrow}\right] \\
&\quad \cup \vartheta_0 \times \left[(\{0\} \times \{-1,1\} \times \{-1\}) \cap D_0^{\downarrow}\right] \\
&= \vartheta_0 \times [\{0\} \times \{-1\} \times (\{-1\} \cap \{-3\})] \\
&\quad \cup \vartheta_0 \times [\{0\} \times \{1\} \times (\{-1\} \cap \{3\})] \\
&= \emptyset. \qquad \text{(I.4)}
\end{aligned}
$$

which means that whenever a solution experiences a jump after hitting $\vartheta_0 \times \left(D_0^{\uparrow} \cup D_0^{\downarrow}\right)$, it cannot jump again. On the other hand, for $y \in D$ with $x_1 \in \bigcup_{q \in Q \setminus \{0\}} \vartheta_q$, it follows that

$$
\begin{aligned}
G_{av}\left(\vartheta_p \times D_p^{\uparrow}\right) &= G_{av}\left(\vartheta_p \times \{x_2 \text{sign}(p) \geq 0\} \times \{p\} \times K_p^{\uparrow}\right) \\
&= \vartheta_p \times ([\{0\} \times \{[p + \{-1,1\}] \mod 4\} \times \{p + \text{sign}(p)\}]) \\
\cup \vartheta_p \times &([\{x_2 \text{sign}(p) > 0\} \times \{(p + \text{sign}(p)) \mod 4\} \times \{p + \text{sign}(p)\}]),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
&G_{av}\left(\vartheta_p \times D_p^{\uparrow}\right) \cap \\
&\left(\vartheta_p \times \left[D_p^{\uparrow} \cup D_p^{\downarrow}\right]\right) = \vartheta_p \times \left[\{0\} \times \{[p + \{-1,1\}] \mod 4\} \times \{p + \text{sign}(p)\} \cap D_p^{\uparrow}\right] \\
&\quad \cup \vartheta_p \times \left[(\{x_2 \text{sign}(p) > 0\} \times \{(p + \text{sign}(p)) \mod 4\} \times \{p + \text{sign}(p)\}) \cap D_p^{\uparrow}\right] \\
&\quad \cup \vartheta_p \times \left[\{0\} \times \{[p + \{-1,1\}] \mod 4\} \times \{p + \text{sign}(p)\} \cap D_p^{\downarrow}\right] \\
&\quad \cup \vartheta_p \times \left[(\{x_2 \text{sign}(p) > 0\} \times \{(p + \text{sign}(p)) \mod 4\} \times \{p + \text{sign}(p)\}) \cap D_p^{\downarrow}\right]
\end{aligned}
$$

(1. any set with $q \neq p$ has empty intersection with $D_p^{\uparrow}$)

$$
\begin{aligned}
&= \vartheta_p \times \left[\{0\} \times \{[p + \{-1,1\}] \mod 4\} \times \{\{p + \text{sign}(p)\}\} \cap D_p^{\downarrow}\right] \\
&\cup \vartheta_p \times \left[(\{x_2 \text{sign}(p) > 0\} \times \{(p + \text{sign}(p)) \mod 4\} \times \{p + \text{sign}(p)\}) \cap D_p^{\downarrow}\right]
\end{aligned}
$$

(2. any z with x_2, p such that $x_2\text{sign}(p) > 0$ does not belong to $D_p^{\downarrow}$)

$$\begin{aligned}
&= \vartheta_p \times \left[(\{0\} \times \{[p + \{-1, 1\}] \mod 4\} \times \{p + \text{sign}(p)\}) \cap D_p^{\downarrow}\right] \\
&= \vartheta_p \times \left[\{0\} \times \{(p + \text{sign}(p)) \mod 4\} \times \{p + \text{sign}(p)\} \cap K_p^{\downarrow}\right]
\end{aligned}$$

(3. $\{p + \text{sign}(p)\} \cap K_p^{\downarrow} = \emptyset$ by design)

$$= \emptyset. \tag{I.5}$$

Hence, whenever a solution experiences a jump after hitting $\bigcup_{q \in Q \setminus \{0\}} \vartheta_q \times D_0^{\uparrow}$, it cannot jump again. On the other hand, for $y \in D$ with $x_1 \in \bigcup_{q \in Q \setminus \{0\}} \vartheta_q$, it follows that

$$\begin{aligned}
G_{av}\left(\vartheta_p \times D_p^{\downarrow}\right) &= G_{av}\left(\vartheta_p \times \{x_2\text{sign}(p) \leq 0\} \times \left\{p^{\uparrow}\right\} \times K_p^{\downarrow}\right) \\
&= \vartheta_p \times \left(\left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\}\right]\right) \\
&\quad \cup \vartheta_p \times \left(\left[\{x_2\text{sign}(p) < 0\} \times \{p\} \times \{g(p)\}\right]\right)
\end{aligned}$$

where

$$p^{\uparrow} : Q \setminus \{0\} \to Q,\ p \mapsto (p + \text{sign}(p)) \mod 4, \qquad p^{\uparrow\uparrow} = \begin{cases} -3 & \text{if } p = 3 \\ 0 & \text{if } p = 2 \\ 3 & \text{if } p = 1 \\ -3 & \text{if } p = -1 \\ 0 & \text{if } p = -2 \\ 3 & \text{if } p = -3 \end{cases},$$

$$g(p) \coloneqq p^{\uparrow} + \text{sign}\left(p^{\uparrow}\right). \tag{I.6}$$

Therefore, we have

$$
\begin{aligned}
&G_{av}\left(\vartheta_p \times D_p^{\downarrow}\right) \\
&\cap\left(\vartheta_p \times \left[D_p^{\uparrow} \cup D_p^{\downarrow}\right]\right) = \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\uparrow}\right] \\
&\qquad\qquad \cup \vartheta_p \times \left[\left(\{x_2 \mathrm{sign}(p) < 0\} \times \{p\} \times \{g(p)\}\right) \cap D_p^{\uparrow}\right] \\
&\qquad\qquad\quad \cup \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\downarrow}\right] \\
&\qquad\qquad\qquad \cup \vartheta_p \times \left[\left(\{x_2 \mathrm{sign}(p) < 0\} \times \{p\} \times \{g(p)\}\right) \cap D_p^{\downarrow}\right]
\end{aligned}
$$

(1. Any $z = (x_1, x_2, p)$ satisfying $x_2 \mathrm{sign}(p) < 0$ is not in $D_p^{\uparrow}$)

$$
\begin{aligned}
&= \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\uparrow}\right] \\
&\qquad \cup \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\downarrow}\right] \\
&\qquad\quad \cup \vartheta_p \times \left[\left(\{x_2 \mathrm{sign}(p) < 0\} \times \{p\} \times \{g(p)\}\right) \cap D_p^{\downarrow}\right]
\end{aligned}
$$

$\left(2.\ \left\{p^{\uparrow}\right\} \cap \left\{p^{\uparrow\uparrow}, p\right\} = \emptyset\ \right)$

$$
\begin{aligned}
&= \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\uparrow}\right] \\
&\qquad\quad \cup \vartheta_p \times \left[\left(\{x_2 \mathrm{sign}(p) < 0\} \times \{p\} \times \{g(p)\}\right) \cap D_p^{\downarrow}\right]
\end{aligned}
$$

(3. $\{p\} \cap \left\{p^{\uparrow}\right\} = \emptyset$)

$$
\begin{aligned}
&= \vartheta_p \times \left[\{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \{g(p)\} \cap D_p^{\uparrow}\right] \\
&= \vartheta_p \times \{0\} \times \{p\} \times \left(\{g(p)\} \cap K_p^{\uparrow}\right) \\
&= \vartheta_p \times \{0\} \times \{p\} \times (\{g(p)\}) \in \vartheta_p \times D_p^{\uparrow}. \qquad \text{(I.7)}
\end{aligned}
$$

Equation (I.7) implies that if a solution hits $\vartheta_p \times D_p^{\downarrow}$ with $p \neq 0$ and jumps, it can only jump again whenever $x_2^+ = x_2 = 0$, since in that case $y^+ \in \vartheta_p \times D_p^{\uparrow}$. However, after such

jump, by (I.5), it cannot jump once more. Therefore, in that case there are at most 2 consecutive jumps. ■

Proof of Lemma 11.4

Note that solutions do not stop due to jumps since by construction $G_{av}(D) \subseteq (C \cup D)$. Moreover, note that F_{av} is globally Lipschitz. Thus, by Gronwall-Bellman inequality the flow map (11.10) does not generate finite escape times. Therefore, by [33, Prop. 2.10], all maximal solutions of $\mathcal{H}$ are complete. ■

Proof of Lemma 11.5

First, by Lemma 11.1, continuity of $U_q(\cdot)$ for all q, and the fact that $\Theta_{\hat{q}}$ is compact for all $p \in Q \setminus \{0\}$, we obtain that $\exists m_p \in \mathbb{R}_{>0}$ and $M_p \in \mathbb{R}_{>0}$ such that

$$\min_{\theta \in \Theta_p} \nabla U_p(\theta) = -m_p \text{sign}(p), \quad \max_{\theta \in \Theta_p} \nabla U_p(\theta) = -M_p \text{sign}(p) \tag{I.8}$$

Moreover, note that $p \in Q \setminus \{-3, 3\} \iff p^{\uparrow} \in Q \setminus \{0\}$. For the sake of simplicity of the proof assume that $p \geq 0$ which implies that $p^{\uparrow} > 0$; the same inequalities we present below hold by switching $m_{p^{\uparrow}}$ with $-m_{p^{\uparrow}}$, $M_{p^{\uparrow}}$ with $-M_{p^{\uparrow}}$, and appropriately changing the direction of the inequalities whenever $p^{\uparrow} < 0$. Therefore, we have that both $-m_{p^{\uparrow}} \text{sign}\left(p^{\uparrow}\right) = -m_p$ and $-M_{p^{\uparrow}} \text{sign}\left(p^{\uparrow}\right) = -M_{p^{\uparrow}}$ by using (I.8). Moreover, under the average dynamics $\mathcal{H}_{av}$ it follows that

$$m_{p^{\uparrow}} - \alpha\beta x_2 \leq \dot{x}_2 \leq M_{p^{\uparrow}} - \alpha\beta x_2 \quad \forall (x, q, k) \in \Theta_q \times \mathbb{R} \times \{q\} \times K_q^{\uparrow} \tag{I.9}$$

From (I.9), by using the comparison principle, we obtain that

$$m_{p^{\uparrow}} t + x_2(0,0)\left(1 + e^{-\alpha\beta t}\right) \leq x_2(t,0) \leq M_{p^{\uparrow}} t + x_2(0,0)\left(1 + e^{-\alpha\beta t}\right), \tag{I.10}$$

for all solutions to $y(t,0)$ to $\mathcal{H}_{av}$ with $y(0,0) \in \Theta_q \times \mathbb{R} \times \{q\} \times K^{\uparrow}$ and all $(t,0) \in \operatorname{dom}(x)$ such that $x_1(t,0) \in \Theta_{p^{\uparrow}}$. On the other hand, by assumption, we have that

$$x_1(0,0) = \vartheta_p \geq 0,\ x_2(0,0) \in \mathbb{R}_{\geq 0},\ q(0,0) = p^{\uparrow},\ k(0,0) = p^{\uparrow}, \tag{I.11}$$

where we have used the fact that $x_1^+ = x_1,\ x_2^+ = x_2$ during jumps, and that $x_1 \geq 0, x_2 \in \mathbb{R}_{\geq 0}\text{sign}(p)$ whenever $y \in \vartheta_p \times D_p^{\uparrow}$. Therefore, using (I.11) on (I.10), it follows that x_2 cannot change signs and additionally that it satisfies

$$m_{p^{\uparrow}} t \leq x_2(t,0).$$

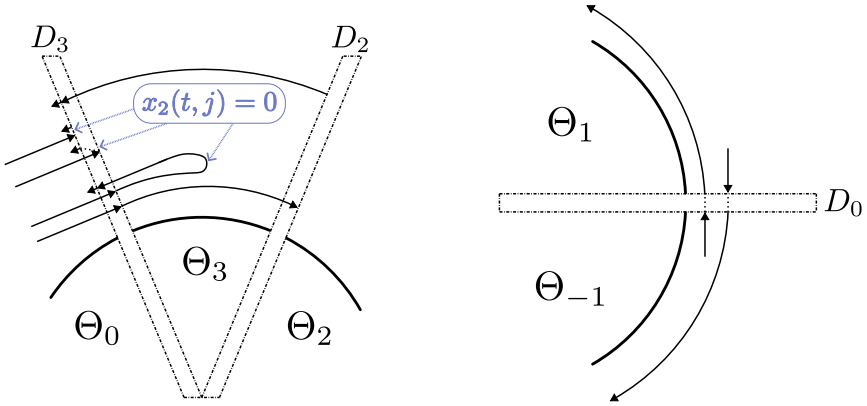
Therefore, using the definition of F_{av} in (11.10), and invoking the comparison principle, from (I.10) we obtain

$$m_{p^{\uparrow}} t \leq \dot{x}_1(t,0) \implies m_{p^{\uparrow}} \frac{t^2}{2} + \vartheta_p \leq x_1(t,0). \tag{I.12}$$

Since $\vartheta_{p^{\uparrow}} > \vartheta_p$ whenever $p^{\uparrow} \in Q \setminus \{0\},\ p^{\uparrow} > 0$, from (I.12) we can conclude that there exists $T > 0$ such that

$$x_1(T,0) = \vartheta_{p^{\uparrow}} \text{ for } T \leq T_p, \text{ and where } T_p^2 \coloneqq \frac{2\left|\vartheta_{p^{\uparrow}} - \vartheta_p\right|}{m_p}. \tag{I.13}$$

The same reasoning holds for $p^{\uparrow} \in Q \setminus \{0\},\ p^{\uparrow} < 0$. By the fact that $\dot{q} = 0$ and $\dot{k} = 0$, we obtain that $q(t,0) = p^{\uparrow}, k(t,0) = p^{\uparrow}$ for all $t \in [0,T]$. This, together with (I.13) and (I.12), implies that $y(T,0) = (\vartheta_p, c_{x_2}, p^{\uparrow}, p^{\uparrow}) \in \vartheta_{p^{\uparrow}} \times D_{p^{\uparrow}}^{\uparrow}$, where $c_{x_2} \geq 0$ is the non-negative value that x_2 has at the hybrid-time $(T,0)$, which depends on the initial condition $x_2(0,0)$. This concludes the proof. ■



(a) With $x_1 \in \bigcup_{q\in Q\setminus\{0\}} \vartheta_q$ **(b)** With $x_1 = \vartheta_0$

Figure I.1. Possible behaviors for trajectories hitting the jump set

Proof of Lemma 11.7

First, since $y(0,0) \in G_{av}\left(\vartheta_p \times D_p^{\uparrow}\right)$ we have that

$$x_1(0,0) = \vartheta_p \in \Theta_0,$$
$$x_2(0,0) \in \mathbb{R}_{\geq 0}\text{sign}(p),\ q(0,0) \in \text{sign}(p)\overline{s}(x_2),\ k(0,0) = \text{sign}(p)4.$$

Now, consider the behavior of the local Lyapunov function:

$$W(x) = U_0(x_1) + \frac{1}{2}|x_2|^2,$$

during the flows of $\mathcal{H}_{av}$. It follows, given $y \in \Theta_0 \times \mathbb{R} \times \{0\} \times K_q^{\uparrow} \subset C$, that

$$\dot{W}(x) = \nabla W(x)^{\top} F_{av}(x,q)$$

$$\left(y \in \Theta_0 \times \mathbb{R} \times \{0\} \times K_q^{\uparrow} \implies q = 0\right)$$

$$= \langle \nabla U_0(x_1),\, x_2\rangle + \langle x_2,\, -\nabla U_0(x_1) - \alpha\beta x_2\rangle$$
$$= -\alpha\beta\,\langle x_2,\, x_2\rangle$$

$$= -\alpha\beta|x_2|^2 \leq 0. \tag{I.14}$$

Now, note that $\dot{W} = 0$ only for solutions with $x_2 = 0$ identically. For thus solutions we thus have that $\dot{x}_2 = 0$ which from (11.10) implies that $\nabla U_0(x_1) = 0$ at all times. However, by Lemma 11.1 this happens if and only if $x_1 \in \{-\pi, \pi\}$. Therefore, from (I.14) we obtain that $\dot{W} < 0$ for all $y \in (\Theta_0 \times \mathbb{R} \times \{0\} \times \{-4, 4\}) \setminus \{-\pi, \pi\}$, and $\dot{W} = 0$ only if $(x_1(t,0), x_2(t,0)) \in \{-\pi, \pi\} \times \{0\}$ for all t. With this at hand we divide the remainder of the analysis by cases:

- Assume that $x_2(0,0) = 0$. Thus, it follows that $q(0,0) \in \text{sign}(p)\{0,3\}$. Since $k(0,0) = \text{sign}(p)4$ we obtain that $y(0,0) \in \Theta_0 \times \mathbb{R} \times \{0\} \times K_0^{\uparrow}$ and $y(0,0) \notin \left(\vartheta_p \times \left(D_p^{\uparrow} \cup D_p^{\downarrow}\right)\right) \cup \left(\Theta_p \times \mathbb{R} \times \{p\} \times K_p^{\uparrow}\right)$. Therefore, solutions must necessarily flow with $q = 0$. Now, assume that there exists $T > 0$ such that $x_1(T,0) = \vartheta_{p^{\uparrow}} = \vartheta_{-p} = -\vartheta_p$. Since $U_0(\cdot)$ is symmetric by design, we obtain that $U_0(x(T,0)) = U_0(-\vartheta_p) = U(\vartheta_p) = U(x_1(0,0))$, and hence that

$$\begin{aligned} W(x(T,0)) &= U(x_1(0,0)) + \frac{1}{2}|x_2(T,0)|^2 \\ &= U(x_1(0,0)) + \frac{1}{2}|x_2(T,0)|^2 + \frac{1}{2}|x_2(0,0)|^2 - \frac{1}{2}|x_2(0,0)|^2 \\ &= W(x_1(0,0)) + \frac{1}{2}\left(|x_2(T,0)|^2 - |x_2(0,0)|^2\right) \\ &\implies 2\left[W\left(x(T,0)\right) - W\left(x(0,0)\right)\right] = |x_2(T,0)|^2 - |x_2(0,0)|^2. \end{aligned} \tag{I.15}$$

  Since $x_1(0,0) \notin \{-\pi, \pi\}$, it follows from the discussion above that $\dot{W}(t,0) < 0$ for almost all t. Thus, we obtain that $W(x(T,0)) < W(x(0,0))$, which implies that $|x(T,0)|^2 < |x(0,0)|^2 = 0$, but this cannot happen. Therefore, if $x_2(0,0) = 0$ there cannot exist $T > 0$ such that $y(T,0) \in \vartheta_{p^{\uparrow}} \times D_{-p}^{\downarrow}$, and thus, solutions should flow with $x_1(t,0) \in \Theta_0$ for all $t > 0$. Hence, by the invariance principle, and via (I.14), it follows that $\lim_{t\to\infty}(x_1(t,0), x_2(t,0)) \in \{-\pi, \pi\} \times \{0\}$. Since

$(q(0,0),k(0,0)) \in \{0\} \times \{-4,4\}$, jumps are not triggered, and $\dot{q} = 0, \dot{k} = 0$ it follows that $(q(t,0),k(t,0)) \in \{0\} \times \{-4,4\}$ and thus that $\lim_{t\to\infty} y(t,0) \in \mathcal{A}_0$.

- If $x_2(0,0) \in \mathbb{R}_{>0}\text{sign}(p)$. On possibility is that the initial kinetic energy $\frac{1}{2}|x_2(0,0)|^2$ is big enough for there not to be enough dissipation during the flow with $x_1 \in \Theta_0$ and consequently that $x_1(T,0) \in \vartheta_{p^\uparrow}$ for some $T > 0$. If conversely the initial kinetic energy is small enough to have invariance of $\Theta_0 \times \mathbb{R} \times \{0\} \times \{-4,4\}$ by a similar argument to the one presented in the previous item, we obtain that $\lim_{t\to\infty} y(t,0) \in \mathcal{A}_0$.

■

Proof of Lemma 11.8

To prove Lemma 11.8 we study conditions to guarantee that $p^\uparrow + \text{sign}\left(p^\uparrow\right) \in K^\downarrow_{p^\downarrow}$ and $p^\uparrow + \text{sign}\left(p^\uparrow\right) \in K^\uparrow_p$ for all $p \in Q \setminus \{0\}$, enabling the successful emulation of the heavy-ball dynamics on the designed potential.

First, by using the definition of $D^\downarrow_p$ and G_{av} it follows that

$$\begin{aligned} &x_1(0,0) = \vartheta_p,\ x_2(0,0) = 0,\ q(0,0) \in \left\{p^{\uparrow\uparrow}, p\right\},\ k(0,0) = p^\uparrow + \text{sign}\left(p^\uparrow\right),\ \text{or} \\ &x_1(0,0) = \vartheta_p,\ x_2(0,0) \in \mathbb{R}_{<0}\text{sign}(p),\ q(0,0) = p,\ k(0,0) = p^\uparrow + \text{sign}\left(p^\uparrow\right). \end{aligned} \tag{I.16}$$

Now, we divide the analysis by cases depending on the value of $x_2(0,0)$.

- $x_2(0,0) = 0$. From (I.16) it follows that

$$y(0,0) \in \vartheta_p \times \{0\} \times \left\{p^{\uparrow\uparrow}, p\right\} \times \left\{p^\uparrow + \text{sign}\left(p^\uparrow\right)\right\}$$

. Note that by construction of the flow and jump sets of $\mathcal{H}$, no solution is allowed to flow nor jump from a state with $x_1 = \vartheta_p$ and $q = p^{\uparrow\uparrow}$. Therefore, we only need

to focus on the case with $y(0,0) = \vartheta_p \times \{0\} \times \{p\} \times \left\{p^{\uparrow} + \text{sign}\left(p^{\uparrow}\right)\right\}$. By design, $\underline{p^{\uparrow} + \text{sign}\left(p^{\uparrow}\right) \in K_p^{\uparrow}}$ for all $p \in Q \setminus \{0\}$. Thus, it follows that $y(0,0) \in \vartheta_p \times D_p^{\uparrow}$, and, due to $K_p^{\uparrow} \cap K_{p^{\uparrow}}^{\uparrow} = \emptyset$ for all $p \in Q$ and the fact that $q(0,0) = p$, that $y(0,0) \notin \left(\Theta_{p^{\uparrow}} \times \mathbb{R} \times \left\{p^{\uparrow}\right\} \times K_{p^{\uparrow}}^{\uparrow}\right)$. Moreover, although $y(0,0) \in \Theta_p \times \mathbb{R} \times \{p\} \times K_p^{\uparrow}$, since $x_1(0,0) = \vartheta_p$, $x_2(0,0) = 0$, and $\langle \nabla U_p(x_1),\, x_1 \rangle < 0$ by Lemma 11.1, there is no solution flowing from the particular $y(0,0)$ under study. In summary, we have that $y(0,0) \in \left(\vartheta_p \times D_p^{\uparrow}\right)$, $y(0,0) \notin \left(\Theta_{p^{\uparrow}} \times \mathbb{R} \times \left\{p^{\uparrow}\right\} \times K_{p^{\uparrow}}^{\uparrow}\right)$, and solutions will necessarily jump afterwards.

- $x_2(0,0) \in \mathbb{R}_{<0}\text{sign}(p)$. For this case, from (I.16), it follows that $y(0,0) = \vartheta_p \times \mathbb{R}_{<0}\text{sign}(p) \times \{p\} \times \left\{p^{\uparrow} + \text{sign}\left(p^{\uparrow}\right)\right\}$. Therefore, solutions will necessarily have to flow as $y(0,0) \in \left(\Theta_p \times \mathbb{R} \times \{p\} \times K_p^{\uparrow}\right)$ and $y(0,0) \notin D$. For the sake of simplicity, assume that $p \geq 0$. Then, by the comparison principle we have

$$m_p t + x_2(0,0)\left(1 + e^{-\alpha\beta t}\right) \leq x_2(t,0) \leq M_p t + x_2(0,0)\left(1 + e^{-\alpha\beta t}\right), \tag{I.17a}$$

$$\vartheta_p + m_p \frac{t^2}{2} + x_2(0,0)\left(t - \frac{1}{\alpha\beta}e^{-\alpha\beta t}\right) \leq x_1(t,0) \leq \vartheta_p + M_p \frac{t^2}{2} + x_2(0,0)\left(t - \frac{1}{\alpha\beta}e^{-\alpha\beta t}\right), \tag{I.17b}$$

where $m_p > 0$ and $M_p > 0$ are as defined in the proof of the previous lemma. The same inequalities we present below hold by switching m_p with $-m_p$, M_p with $-M_p$, and appropriately changing the direction of the inequalities whenever $p < 0$. Depending on the magnitude of $x_2(0,0)$, from (I.17) and the fact $x_2(0,0) < 0$, we consider the two following cases:

  - There exists $T_0 > 0$ such that $x_2(T_0,0) = 0$ and $x_1(T_0,0) \in (\vartheta_p, \vartheta_{p^{\downarrow}})$. In such case, $y(T_0,0) \in \Theta_p \times \mathbb{R}_{\geq 0} \times \{p\} \times \left\{K_p^{\uparrow}\right\}$. Then, using (11.10), by the comparison

principle we get that

$$m_p(t - T_0) \leq x_2(t, 0) \leq M_p(t - T_0), \tag{I.18a}$$

$$x_1(T_0, 0) + m_p \frac{(t - T_0)^2}{2} \leq x_1(t, 0) \leq x_1(T_0, 0) + m_p \frac{(t - T_0)^2}{2}, \tag{I.18b}$$

for $t > T_0$. Since $x_1(T_0, 0) \in \left(\vartheta_p, \vartheta_{p^\downarrow}\right)$, and $m_p > 0, M_p > 0$, by definition, from (I.18b) we obtain that there exists $T > T_0$ satisfying $(T - T_0)^2 \leq \frac{2\left|\vartheta_p - \vartheta_{p^\downarrow}\right|}{m_p}$, such that $x_1(T, 0) = \vartheta_p$ and $x_2(T, 0) \geq 0$. Since $\dot{q} = 0$ and $\dot{k} = 0$, it follows that $q(T, 0) = p,\ k(T, 0) = p^\uparrow + \text{sign}\left(p^\uparrow\right) \in K_p^\uparrow\ \forall p$. The same arguments hold for $p < 0$, by following the modifications mentioned above, and, therefore, it follows that $y(T, 0) \in \vartheta_p \times D_p^\uparrow$ for all $p \in Q \setminus \{0\}$.

– There exists $T > 0$ such that $x_1(T, 0) = \vartheta_{p^\downarrow}$ and $x_2(T, 0) \leq 0$, i.e., x_2 does not change sign during the flow interval. Since $\dot{q} = 0$ and $\dot{k} = 0$, it follows that $q(T, 0) = p,\ k(T, 0) = p^\uparrow + \text{sign}\left(p^\uparrow\right)$. Since $\underline{p^\uparrow + \text{sign}\left(p^\uparrow\right) \in K_{p^\downarrow}^\downarrow}$ for all $p \in Q \setminus \{0\}$ by design, it follows that $y(T, 0) = (\vartheta_{p^\downarrow}, c_{x_2}, p, p^\uparrow + \text{sign}\left(p^\uparrow\right)) \in \vartheta_{p^\downarrow} \times D_{p^\downarrow}^\downarrow$, where $c_{x_2} \geq 0$ is the non-positive value that x_2 has at the hybrid-time $(T, 0)$, which depends on the initial condition $x_2(0, 0)$. Since the same arguments hold for $p < 0$, by following the modifications mentioned above, it follows that $y(T, 0) \in \vartheta_{p^\downarrow} \times D_{p^\downarrow}^\downarrow$ for all $p \in Q \setminus \{0\}$.

■

Proof of Lemma 11.9

Together, Lemmas 11.5-11.8 show that $\mathcal{H}_{av}$ induces a finite amount of jumps before entering the region with $x_1 \in \Theta_0$ with low enough kinetic energy such that solutions will flow from there onward without jumping again. In particular, this implies that there cannot be maximal Zeno solutions to $\mathcal{H}_{av}$. ■

Proof of Theorem 11.1

Via Lemmas 11.5-11.9 we can conclude that the HDS $\mathcal{H}_{av}$ emulates continuous-time heavy-ball dynamics with friction coefficient $\alpha\beta$ and potential function given by $U(x_1, q)$ as defined in (11.6). Now, consider the following Lyapunov candidate function

$$V(x, q) = U(x_1, q) + \frac{|x_2|^2}{2}$$

which satisfies $V(x, q) \geq 0 \ \forall (x, q) \in C \cup D$, and $V(x, q) = 0 \iff (x, q) \in \mathcal{A}_0$ via the application of Lemma 11.1. Moreover, V is radially unbounded since V is continuous in x_1 and $x_1 \in \mathbb{S}^1$, and $V(x) \to \infty$ as $|x_2| \to \infty$. For all $(x, q) \in C$, during flows we have that

$$\begin{aligned}
\dot{V}(x, q, k) &= \nabla V^\top \begin{pmatrix} x_2 \\ -\nabla U_q(x_1) - \alpha\beta x_2 \end{pmatrix} \qquad \forall (x, q) \in C \\
&= \begin{pmatrix} \nabla U(x_1, q) & x_2 \end{pmatrix} \begin{pmatrix} x_2 \\ -\nabla U_q(x_1) - \alpha\beta x_2 \end{pmatrix} \\
&= \nabla U(x_1, q) x_2 - x_2 \nabla U_q(x_1) - \alpha\beta |x_2|^2 \\
&= \nabla U_q(x_1) x_2 - x_2 \nabla U_q(x_1) - \alpha\beta |x_2|^2 \\
&= -\alpha\beta |x_2|^2 \leq 0,
\end{aligned} \tag{I.19}$$

where we have used the fact that $(x, q, k) \in C \implies x \in \Theta_q \times \mathbb{R}$, and thus that $U(x_1, q) = U_q(x_1) \ \forall (x, q, k) \in C$. On the other hand, for every $(x, q, k) \in D$, during jumps we have that $V(x^+, q^+, k^+) = V(x, q, k)$.

Now, we show that no complete solution of $\mathcal{H}_{av}$ keeps V in a non-zero level set. To this end, consider a solution for which $\dot{V}(x, q, k) = 0$. Then, it follows that $x_2(t, j) = 0$ for almost all $(t, j) \in \mathrm{dom}\,(x)$, and thus that during flows $\dot{x}_2 = 0$ for almost all t. Therefore, since solutions of $\mathcal{H}_{av}$ satisfy $(\dot{x}, \dot{q}, \dot{k}) = F_{av}(x)$ for almost all t during intervals of flow,

it must be that $0 = -\nabla U_q(x_1(t,j))$. However, by Lemma 11.1, this can only happen if $(x_1(t,j), x_2(t,j), q(t,j), k(t,j)) \in \{-\pi, \pi\} \times \{0\} \times \{0\} \times \{-4, 4\}$ for all $(t,j) \in \mathrm{dom}(x)$. In particular, this means that $\dot{V} < 0$ for all $(x,q) \in C \setminus \{-\pi, \pi\} \times \{0\} \times \{-4, 4\}$. Therefore, no solution that flows can keep V constant in a non-zero level set. Since by Lemma 11.9 every maximal solution is non-Zeno, and in fact there is only a finite ammount of jumps, we obtain UGAS of $\mathcal{A}_0$ via the the hybrid invariance principle [33, Thm. 8.8.]. Global practical asymptotic stability of $\mathcal{A}_0 \times \mathbb{S}^1$ is obtained by applying [272, Thm. 2]. ■

Bibliography

[1] D. E. Ochoa and J. I. Poveda, "Momentum-based nash set-seeking over networks via multitime scale hybrid dynamic inclusions," *IEEE Transactions on Automatic Control*, vol. 69, no. 7, pp. 4245–4260, 2024.

[2] ——, "Accelerated continuous-time approximate dynamic programming via data-assisted hybrid control," *IFAC-PapersOnLine*, vol. 55, no. 12, pp. 561–566, 2022.

[3] Y. Chen, D. E. Ochoa, J. R. Marden, and J. I. Poveda, "High-order decentralized pricing dynamics for congestion games: Harnessing coordination to achieve acceleration," in *2023 American Control Conference (ACC)*, 2023, pp. 1086–1091.

[4] D. E. Ochoa, J. I. Poveda, C. Uribe, and N. Quijano, "Robust optimization over networks using distributed restarting of accelerated dynamics," *IEEE Control Systems Letters*, vol. 5, pp. 301–306, 2021.

[5] D. E. Ochoa, M. U. Javed, X. Chen, and J. I. Poveda, "Decentralized concurrent learning with coordinated momentum and restart," *Systems & Control Letters*, vol. 193, p. 105931, 2024.

[6] D. E. Ochoa, J. I. Poveda, A. Subbaraman, G. S. Schmidt, and F. R. Pour-Safaei, "Accelerated concurrent learning algorithms via data-driven hybrid dynamics and nonsmooth odes," in *Learning for Dynamics and Control*. PMLR, 2021, pp. 866–878.

[7] D. E. Ochoa and J. I. Poveda, "High-performance optimal incentive-seeking in transactive control for traffic congestion," *European Journal of Control*, vol. 68, p. 100696, 2022, 2022 European Control Conference Special Issue. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358022000899

[8] ——, "Dynamic Gains for Transient-Behavior Shaping in Hybrid Dynamic Inclusions," *63rd IEEE Conference on Decision and Control, to appear.*, 2024.

[9] D. E. Ochoa, N. Espitia, and J. I. Poveda, "Prescribed-time stability in switching systems with resets: A hybrid dynamical systems approach," *Systems & Control Letters*, vol. 193, p. 105910, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167691124001981

[10] D. E. Ochoa and J. I. Poveda, "Robust global optimization on smooth compact manifolds via hybrid gradient-free dynamics," *Automatica*, vol. 171, p. 111916, 2025.

[11] ——, "On Hybrid Vibrational Control: Robust Global Stabilization of the Kapitza's Pendulum," in Preparation.

[12] Y. Tang, K. Dvijotham, and S. Low, "Real-time optimal power flow," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2963–2973, 2017.

[13] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez, "Deep reinforcement learning for autonomous driving: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[14] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Timescale separation in autonomous optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 611–624, 2020.

[15] M. Colombino, E. Dall'Anese, and A. Bernstein, "Online optimization as a feedback controller: Stability and tracking," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 1, pp. 422–432, 2019.

[16] R. Goebel, R. Sanfelice, and A. R. Teel, *Hybrid Dynamical System*. Princeton, NJ: Princeton University Press, 2012.

[17] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: theory and insights," *Advances in neural information processing systems*, vol. 27, no. 1, p. 5312–5354, Jan. 2014.

[18] A. Wibisono, A. C. Wilson, and M. I. Jordan, "A variational perspective on accelerated methods in optimization," *Proceedings of the National Academy of Sciences*, vol. 113, no. 47, pp. E7351–E7358, 2016.

[19] W. Su, S. Boyd, and E. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and Insights," *Journal of Machine Learning Research*, vol. 17, no. 153, pp. 1–43, 2016.

[20] Y. E. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Russian Ac. of Sc.*, vol. 269, pp. 543–547, 1983.

[21] C. Cai and A. Teel, "Characterizations of Input-to-State stability for hybrid systems," *Systems & Control Letters*, vol. 59, pp. 47–53, 2009.

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

[23] B. Shi, S. S. Du, W. Su, and M. I. Jordan, "Acceleration via symplectic discretization of high-resolution differential equations," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[24] J. I. Poveda and N. Li, "Inducing Uniform Asymptotic Stability in Non-Autonomous Accelerated Optimization Dynamics via Hybrid Regularization," *58th IEEE Conference on Decision and Control, 2019.*, pp. 3000–3005, 2019.

[25] ——, "Robust hybrid zero-order optimization algorithms with acceleration via averaging in time," *Automatica*, vol. 123, p. 109361, 2021.

[26] J. Xu, Y. Tian, Y. Sun, and G. Scutari, "Accelerated primal-dual algorithms for distributed smooth convex optimization over networks," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 2381–2391.

[27] H. Attouch, Z. Chbani, and H. Riahi, "Fast proximal methods via time scaling of damped inertial dynamics," *SIAM Journal on Optimization*, vol. 29, no. 3, pp. 2227–2256, 2019.

[28] W. Krichene, A. Bayen, and P. L. Bartlett, "Accelerated mirror descent in continuous and discrete time," *Advances in neural information processing systems*, vol. 28, 2015.

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

[30] W. Wang, A. R. Teel, and D. Nešić, "Analysis for a class of singularly perturbed hybrid system via averaging," *Automatica*, vol. 48, no. 6, pp. 1057–1068, 2012.

[31] O'Donoghue and E. J. Candes, "Adaptive restart for accelerated gradient schemes," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 715–732, 2013.

[32] H. Attouch, Z. Chbani, and H. Riahi, "Combining fast inertial dynamics for convex optimization with tikhonov regularization," *Journal of Mathematical Analysis and Applications*, vol. 457, no. 2, pp. 1065–1094, 2018.

[33] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012.

[34] J. B. Rosen, "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games," *Econometrica*, vol. 33, pp. 520–534, 1965.

[35] P. Frihauf, M. Krstić, and T. Basar, "Nash Equilibrium Seeking in Noncooperative Games," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1192–1207, May 2012.

[36] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, "How to train your robot with deep reinforcement learning: lessons we have learned," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 698–721, 2021.

[37] J. Martinez-Piazuelo, D. E. Ochoa, N. Quijano, and L. F. Giraldo, "A multi-critic reinforcement learning method: An application to multi-tank water systems," *IEEE Access*, vol. 8, pp. 173 227–173 238, 2020.

[38] K. G. Vamvoudakis, Y. Wan, F. L. Lewis, and D. Cansever, "Handbook of Reinforcement Learning and Control," 2021.

[39] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. E. Dixon, *Reinforcement learning for optimal feedback control: A Lyapunov-based approach*. Springer, 2018.

[40] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[41] G. Chowdhary and E. Johnson, "Concurrent learning for convergence in adaptive control without persistency of excitation," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 3674–3679.

[42] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[43] D. Liberzon, *Calculus of variations and optimal control theory*. Princeton university press, 2011.

[44] K. Hornik, M. Stinchcombe, and H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural networks*, vol. 3, no. 5, pp. 551–560, 1990.

[45] K. G. Vamvoudakis, M. F. Miranda, and J. P. Hespanha, "Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2386–2398, 2015.

[46] R. Kamalapurkar, P. Walters, and W. E. Dixon, "Model-based reinforcement learning for approximate optimal regulation," *Automatica*, vol. 64, pp. 94–104, 2016.

[47] K. J. Astrom and B. Wittenmark, *Adaptive Control*. Addison-Wesley Publishing Company, 1989.

[48] K. Ciosek, "Imitation Learning by Reinforcement Learning," in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=1zwleytEpYx

[49] P. Rashidinejad, B. Zhu, C. Ma, J. Jiao, and S. Russell, "Bridging offline reinforcement learning and imitation learning: A tale of pessimism," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[50] K. Doya, "Reinforcement learning in continuous time and space," *Neural computation*, vol. 12, no. 1, pp. 219–245, 2000.

[51] J. I. Poveda and A. R. Teel, "Hybrid Mechanisms for Robust Synchronization and Coordination of Multi-Agent Networked Sampled-Data Systems," *Automatica,*, vol. 99, pp. 41–53, 2018.

[52] R. T. Rockafellar and R. J. Wets, *Variational Analysis*. Springer, 1998, vol. 317.

[53] S. R. Etesami, "Complexity and approximability of optimal resource allocation and nash equilibrium over networks," *SIAM Journal on Optimization*, vol. 30, no. 1, pp. 885–914, 2020.

[54] L. Fortunati, A. M. Manganelli, F. Cavallo, and F. Honsell, "You need to show that you are not a robot," *New Media & Society*, vol. 21, no. 8, pp. 1859–1876, 2019.

[55] J. Nash, "Non-Cooperative Games," *Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951.

[56] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. San Diego, CA: Academic Press, 1995.

[57] J. R. Marden, H. P. Young, G. Arslan, and J. S. Shamma, "Payoff-Based Dynamics for Multiplayer Weakly Acyclic Games," *SIAM J. Control. Optim.*, vol. 48, no. 1, pp. 373–396, 2009.

[58] M. Bianchi and S. Grammatico, "Fully Distributed Nash Equilibrium Seeking Over Time-Varying Communication Networks With Linear Convergence," *IEEE Cont. Syst. Letters*, vol. 5, no. 2, pp. 499–504, 2021.

[59] M. Stanković, K. H. Johansson, and D. M. Stipanović, "Distributed Seeking of Nash Equilibria With Applications to Mobile Sensor Networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 4, pp. 904–919, 2012.

[60] C. D. Persis and S. Grammatico, "Distributed averaging integral Nash equilibrium seeking on networks," *Automatica*, vol. 110, pp. 1–7, 2019.

[61] E. Altman, T. Başar, and R. Srikant, "Nash equilibra for combined flow control and routing in Networks: Asymptotic Behavior for a Large Number of Users," *IEEE Transactions on Automatic and Control*, vol. 47, no. 6, pp. 917–930, 2002.

[62] Z. Deng and X. Nian, "Distributed Generalized Nash Equilibrium Seeking Algorithm Design for Aggregative Games Over Weight-Balanced Digraphs," *IEEE Trans. Neural Networks and Learning Systems*, vol. 30, no. 3, pp. 695–706, 2019.

[63] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[64] W. H. Sandholm, "Potential games with continuous player sets," *Journal of Economic theory*, vol. 97, no. 1, pp. 81–108, 2001.

[65] J. Hofbauer and W. H. Sandholm, "Stable games and their dynamics," *Journal of Economic theory*, vol. 144, no. 4, pp. 1665–1693, 2009.

[66] P. Mertikopoulos and Z. Zhou, "Learning in games with continuous action sets and unknown payoff functions," *Mathematical Programming*, vol. 173, no. 1, pp. 465–507, 2019.

[67] B. Gao and L. Pavel, "On the Rate of Convergence of Continuous-Time Game Dynamics in N-Player Potential Games," *59th IEEE Conference on Decision and Control*, pp. 1678–1683, 2020.

[68] W. Sandholm, *Population Games and Evolutionary Dynamics*. Cambridge, MA: The MIT Press, 2010.

[69] J. R. Marden, G. Arslan, and J. Shamma, "Cooperative Control and Potential Games," *IIEEE Transactions on Systems, Man And Cybernetics*, vol. 39, pp. 1393–1407, 2009.

[70] J. R. Marden and J. S. Shamma, "Game theory and distributed control," in *Handbook of game theory with economic applications*. Elsevier, 2015, vol. 4, pp. 861–899.

[71] J. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Transactions on Automatic Control*, vol. 29, no. 1, pp. 42–50, 1984.

[72] E. Campos-Nanez, A. Garcia, and C. Li, "A game-theoretic approach to efficient power management in sensor networks," *Operations Research*, vol. 56, no. 3, pp. 552–561, 2008.

[73] G. Scutari, D. P. Palomar, J.-S. Pang, and F. Facchinei, "Flexible design of cognitive radio wireless systems," *IEEE Signal Processing Magazine*, vol. 26, no. 5, pp. 107–123, 2009.

[74] N. Li and J. R. Marden, "Designing games for distributed optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 230–242, 2013.

[75] J. Nash, "Non-cooperative games," *Annals of mathematics*, pp. 286–295, 1951.

[76] T. Roughgarden, "Algorithmic game theory," *Communications of the ACM*, vol. 53, no. 7, pp. 78–86, 2010.

[77] N. Quijano, C. Ocampo-Martinez, J. Barreiro-Gomez, G. Obando, A. Pantoja, and E. Mojica-Nava, "The role of population games and evolutionary dynamics in distributed control systems: The advantages of evolutionary game theory," *IEEE Control Systems Magazine*, vol. 37, no. 1, pp. 70–97, 2017.

[78] P. N. Brown and J. R. Marden, "Studies on robust social influence mechanisms: Incentives for efficient network routing in uncertain settings," *IEEE Control Systems Magazine*, vol. 37, no. 1, pp. 98–115, 2017.

[79] A. C. Pigou, *The Economics of Welfare*. New York: Macmillan, 1920.

[80] E. Provonsha and N. Sifuentes, "Road Pricing in London, Stockholm and Singapore, a way forward for New York City," Tri-State Transportation Campaign, Tech. Rep., 2017.

[81] T. Litman, "London Congestion Pricing - Implications for Other Cities," *CESifo DICE Report, ISSN 1613-6373, ifo Institut für Wirtschaftsforschung an der Universität München, München*, vol. 03, no. 3, pp. 17–21, 2005.

[82] E. Croci and A. Ravazzi, "Urban Road Pricing: A Comparative Study on the Experiences of London, Stockholm and Milan," *The Center for Research on Energy and Environmental Economics and Policy at Bocconi University, Working Paper Series - ISSN 1973-0381*, pp. 1–43, 2016.

[83] B. Schaller, "New York City's congestion pricing experience and implications for road pricing acceptance in the United States," *Transport Policy*, vol. 4, no. 17, pp. 266–273, 2010.

[84] W. Sandholm, "Evolutionary Implementation and Congestion Pricing," *Review of Economic Studies*, vol. 69, no. 667-689, 2002.

[85] U. Bhaskar, K. Ligett, and L. J. Schulman, "Achieving Target Equilibria in Network Routing Games without Knowing the Latency Functions," *IEEE Annual Symposium on Foundations of Computer Science*, pp. 31–40, 2014.

[86] Y. Nie, X. Wang, and K. Cheng, "Multi-Area Self-Adaptive Pricing Control in Smart City With EV user Participation," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–9, 2017.

[87] F. Farokhi and K. H. Johansson, "A picewise-constant congestion taxing policty for repeated routing games," *Transporation Research Part B*, vol. 78, pp. 123–143, 2015.

[88] S. Saharan, S. Bawa, and N. Kumar, "Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review," *Computer Communications*, vol. 150, pp. 603–625, 2020.

[89] C. Maheshwari, K. Kulkarni, M. Wu, and S. S. Sastry, "Dynamic tolling for inducing socially optimal traffic loads," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 4601–4607.

[90] J. I. Poveda, P. N. Brown, J. R. Marden, and A. R. Teel, "A Class of Distributed Adaptive Pricing Mechanisms for Societal Systems with Limited Information," *56th IEEE Conference on Decision and Control*, pp. 1490–1495, 2017.

[91] W. H. Sandholm, "Evolutionary implementation and congestion pricing," *The Review of Economic Studies*, vol. 69, no. 3, pp. 667–689, 2002.

[92] M. J. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Santa Monica, CA: RAND Corporation, 1955.

[93] B. O'donoghue and E. Candes, "Adaptive restart for accelerated gradient schemes," *Foundations of computational mathematics*, vol. 15, no. 3, pp. 715–732, 2015.

[94] D. Kim and J. A. Fessler, "Adaptive restart of the optimized gradient method for convex optimization," *Journal of Optimization Theory and Applications*, vol. 178, no. 1, pp. 240–263, 2018.

[95] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[96] D. S. Acharya and N. Nath, "Applications of Multi Agent Systems in Control Engineering: A State of the Art Survey," *International Journal of Innovative Research in Advanced Engineering*, vol. 2, no. 4, pp. 113–128, 2015.

[97] B. Gharesifard and J. Cortes, "Distributed Continuous-Time Convex Optimization on Weigh-Balanced Digraphs," *IEEE Transactions on Automatic Control*, vol. 59, pp. 781–786, 2014.

[98] A. Mokhtari, Q. Ling, and A. Ribeiro, "Network Newton Distributed Optimization Methods," *IEEE Transactions on Signal Processing*, vol. 65, pp. 146–161, 2017.

[99] J. Cortes and S. K. Niederlander, "Distributed Coordination for Nonsmooth Convex Optimization via Saddle-Point Dynamics," *Journal of Nonlinear Science*, vol. 29, pp. 1247–1272, 2019.

[100] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained Consensus and Optimization in Multi-Agent Networks," *IEEE Trans. Autom. Contr.*, vol. 55, pp. 922–938, 2010.

[101] A. C. Wilson, B. Recht, and M. I. Jordan, "A lyapunov analysis of accelerated methods in optimization," *Journal of Machine Learning Research*, vol. 22, no. 113, pp. 1–34, 2021.

[102] D. M. Hustig-Schultz and R. Sanfelice, "A Robust Hybrid Heavy Ball Algorithm for Optimization with High Performance," *American Control Conference*, pp. 151–156, 2019.

[103] C. A. Uribe, S. Lee, A. Gasnikov, and A. Nedić, "A dual approach for optimal algorithms in distributed optimization over networks," *Optimization Methods and Software*, vol. 0, no. 0, pp. 1–40, 2020.

[104] J. Zhang, C. A. Uribe, A. Mokhtari, and A. Jadbabaie, "Achieving acceleration in distributed optimization via direct discretization of the Heavy-Ball ODE," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 3408–3413.

[105] R. G. Sanfelice and A. R. Teel, "Dynamical properties of hybrid systems simulators," *Automatica*, vol. 46, pp. 239–248, 2010.

[106] H. Attouch and J. Peypouquet, "Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators," *Mathematical Programming*, vol. 174, pp. 391–432, 2018.

[107] D. Gadjov and L. Pavel, "A Passivity-Based Approach to Nash Equilibrium Seeking Over Networks," *IEEE Trans. Autom. Contr.*, vol. 64, no. 3, pp. 1077–1092, 2019.

[108] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.

[109] M. Ye and G. Hu, "Distributed Seeking of Time-Varying Nash Equilibrium for Non-cooperative Games," *IEEE Trans. Autom. Contr.*, vol. 60, pp. 3000–3005, 2015.

[110] S. Krilašević and S. Grammatico, "Learning generalized Nash equilibria in multi-agent dynamical systems via extremum seeking control," *Automatica*, vol. 133, 2021.

[111] J. I. Poveda and A. R. Teel, "A framework for a class of hybrid extremum seeking controllers with dynamic inclusions," *Automatica*, vol. 76, pp. 113–126, 2017.

[112] R. Kutadinata, W. H. Moase, and C. Manzie, "Dither Re-Use in Nash Equilibrium Seeking," *IEEE Trans. Autom. Control*, vol. 60, pp. 1433–1438, 2015.

[113] J. I. Poveda and N. Quijano, "Shahshahani gradient-like extremum seeking," *Automatica*, vol. 58, pp. 51–59, 2015.

[114] C. J. Maddison, D. Paulin, Y. W. Teh, B. O'Donoghue, and A. Doucet, "Hamiltonian Descent Methods," *arXiv:1809.05042*, 2018.

[115] B. Shi, S. S. Du, M. I. Jordan, and W. J. Su, "Understanding the acceleration phenomenon via high-resolution differential equations," *Mathematical Programming*, Jul 2021.

[116] M. Muehlebach and M. I. Jordan, "Optimization with momentum: Dynamical, control-theoretic, and symplectic perspectives," *Journal of Machine Learning Research*, vol. 22, no. 73, pp. 1–50, 2021.

[117] W. M. Moursi and L. Vandenberghe, "Douglas–Rachford Splitting for the Sum of a Lipschitz Continuous and a Strongly Monotone Operator," *Jour. of Opti. Theory and Appl.*, vol. 183, pp. 179–198, 2019.

[118] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2007.

[119] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas-Revised and Expanded Edition*. Princeton University Press, 2018.

[120] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, 2013.

[121] M. Arcak and N. C. Martins, "Dissipativity tools for convergence to Nash equilibria in population games," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 39–50, 2020.

[122] G. Bianchin, J. I. Poveda, and E. Dall'Anese, "Online optimization of switched lti systems using continuous-time and hybrid accelerated gradient flows," *Automatica*, vol. 146, p. 110579, 2022.

[123] D. Gadjov and L. Pavel, "On the exact convergence to Nash equilibrium in hypomonotone regimes under full and partial-decision information," *IEEE Transactions on Automatic Control*, 2022.

[124] M. Ye and G. Hu, "Solving Potential Games with Dynamical Constraint," *IEEE Transactions on Cybernetics*, vol. 46, no. 5, pp. 1156–1164, 2016.

[125] R. Kamalapurkar, B. Reish, G. Chowdhary, and W. E. Dixon, "Concurrent learning for parameter estimation using dynamic state-derivative estimators," *IEEE Trans. on Automatic Control*, vol. 62, no. 7, pp. 3594–3601, 2017.

[126] J. Casas, C.-H. Chang, and V. H. Duenas, "Switched Adaptive Concurrent Learning Control using a Stance Foot Model for Gait Rehabilitation using a Hybrid Exoskeleton," *IFAC-PapersOnLine*, vol. 55, no. 41, pp. 187–192, 2022.

[127] ——, "Switched Concurrent Learning Adaptive Control for Treadmill Walking Using a Lower Limb Hybrid Exoskeleton," *IEEE Trans. on Ctrl. Systems Technology*, 2023.

[128] G. V. Chowdhary and E. N. Johnson, "Theory and flight-test validation of a concurrent-learning adaptive controller," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 2, pp. 592–607, 2011.

[129] J. I. Poveda, M. Benosman, and K. G. Vamvoudakis, "Data-enabled extremum seeking: a cooperative concurrent learning-based approach," *Int. J. of Adaptive Control and Signal Processing*, vol. 35, no. 7, pp. 1256–1284, 2021.

[130] J. H. Le and A. R. Teel, "Concurrent learning in high-order tuners for parameter identification," in *2022 IEEE 61st Conf. on Decis. and Ctrl.* IEEE, 2022, pp. 2159–2164.

[131] T. Nguyen, R. Baraniuk, A. Bertozzi, S. Osher, and B. Wang, "Momentumrnn: Integrating momentum into recurrent neural networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1924–1936, 2020.

[132] M. Muehlebach and M. I. Jordan, "Optimization with Momentum: Dynamical, Control-Theoretic, and Symplectic Perspectives," *J. Mach. Learn. Res.*, vol. 22, no. 1, Jan. 2021.

[133] H. H. N. Nguyen, T. Nguyen, H. Vo, S. Osher, and T. Vo, "Improving Neural Ordinary Differential Equations with Nesterov's Accelerated Gradient Method," *Advances in Neural Information Processing Systems*, vol. 35, pp. 7712–7726, 2022.

[134] V. Roulet and A. d'Aspremont, "Sharpness, restart and acceleration," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[135] B. Wang, T. Nguyen, T. Sun, A. L. Bertozzi, R. G. Baraniuk, and S. J. Osher, "Scheduled restart momentum for accelerated stochastic gradient descent," *SIAM Journal on Imaging Sciences*, vol. 15, no. 2, pp. 738–761, 2022.

[136] Y. Yu, B. Açıkmeşe, and M. Mesbahi, "Mass–spring–damper networks for distributed optimization in non-Euclidean spaces," *Automatica*, vol. 112, p. 108703, 2020.

[137] N. M. Boffi and J.-J. E. Slotine, "A continuous-time analysis of distributed stochastic gradient," *Neural computation*, vol. 32, no. 1, pp. 36–96, 2020.

[138] C. Sun and G. Hu, "A continuous-time nesterov accelerated gradient method for distributed convex optimization with an exponential convergent bregman distance," in *2023 5th International Conference on Industrial Artificial Intelligence (IAI)*, 2023, pp. 1–6.

[139] D. E. Ochoa, J. I. Poveda, C. A. Uribe, and N. Quijano, "Robust optimization over networks using distributed restarting of accelerated dynamics," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 301–306, 2020.

[140] S. Z. Khong, Y. Tan, C. Manzie, and D. Nesic, "Multi-agent source seeking via discrete-time extremum seeking control," *Automatica*, vol. 50, no. 9, pp. 2312–2320, 2014.

[141] X. Chen and Y. Li, "Smooth formation navigation of multiple mobile robots for avoiding moving obstacles," *International Journal of Control, Automation, and Systems*, vol. 4, no. 4, pp. 466–479, 2006.

[142] W. Chen, C. Wen, S. Hua, and C. Sun, "Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative pe condition via consensus," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 91–106, 2013.

[143] F. Bullo, *Lectures on network systems*. CreateSpace, 2018, vol. 1.

[144] E. D. Sontag and Y. Wang, "On characterizations of the input-to-state stability property," *Systems & Control Letters*, vol. 24, no. 5, pp. 351–359, 1995.

[145] J. I. Poveda, K. G. Vamvoudakis, and M. Benosman, "CODES: Cooperative data-enabled extremum seeking for multi-agent systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 2988–2993.

[146] M. U. Javed, J. I. Poveda, and X. Chen, "Excitation Conditions for Uniform Exponential Stability of the Cooperative Gradient Algorithm over Weakly Connected Digraphs," *IEEE Control and Systems Letters, DOI:10.1109/LCSYS.2021.3049153*, pp. 1–6, 2021.

[147] A. C. Wilson, B. Recht, and M. I. Jordan, "A Lyapunov Analysis of Accelerated Methods in Optimization." *J. Mach. Learn. Res.*, vol. 22, pp. 113–1, 2021.

[148] U. Javed, J. I. Poveda, and X. Chen, "Scalable Resetting Algorithms for Synchronization of Pulse-Coupled Oscillators over Rooted Directed Graphs," *Automatica, under review.*, vol. 132, pp. 1–18, 2020.

[149] H. Ríos, D. Efimov, J. Moreno, W. Perruquetti, and J. Rueda-Escobedo, "Time-varying parameter identification algorithms: Finite and fixed-time convergence," *IEEE Trans. on Aut. Control*, vol. 62, no. 7, pp. 3671–3678, 2017.

[150] F. Tatari, M. Mazouchi, and H. Modares, "Fixed-Time System Identification Using Concurrent Learning," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 4892–4902, 2023.

[151] Z. Zang and R. R. Bitmead, "Transient bounds for adaptive control systems," in *29th IEEE conference on decision and control*. IEEE, 1990, pp. 2724–2729.

[152] X.-B. Gao, "Exponential stability of globally projected dynamic systems," *IEEE Transactions on Neural Networks*, vol. 14, no. 2, pp. 426–431, 2003.

[153] M. Abdelgalil, D. E. Ochoa, and J. I. Poveda, "Multi-time scale control and optimization via averaging and singular perturbation theory: From ODEs to hybrid dynamical systems," *Annual Reviews in Control*, vol. 56, p. 100926, 2023.

[154] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Mineola, New York: Dover Publications Inc., 2012, vol. 1.

[155] J. I. Poveda, M. Benosman, and K. Vamvoudakis, "Data-Enabled Extremum Seeking: A Cooperative Concurrent Learning-Based Approach," *International Journal of Adaptive Control and Signal Processing, DOI: 10.1002/acs.3189*, pp. 1–29, 2020.

[156] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement Learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[157] S. Boyd and S. S. Sastry, "Necessary and sufficient conditions for parameter convergence in adaptive control," *Automatica*, vol. 22, no. 6, pp. 629–639, 1986.

[158] K. S. Narendra and A. Annaswamy, "Persistent excitation in adaptive systems," *International Journal of Control*, vol. 45, no. 1, pp. 127–160, 1987.

[159] K. S. Narendra and A. M. Annaswamy, *Stable Adaptive Systems*. Courier Corporation, 2012.

[160] L. Praly, "Convergence of the gradient algorithm for linear regression models in the continuous and discrete time cases," *Research Report, PSL Research Unversity, Mines ParisTech*, 2017.

[161] G. Chowdhary and E. Johnson, "Concurrent Learning for Convergence in Adaptive Control without Persistence of Excitation," *49th IEEE Conference on Decision and Control*, pp. 3674–3679, 2010.

[162] G. Chowdhary, T. Yucelen, M. Muhlegg, and E. N. Johnson, "Concurrent Learning Adaptive Control of Linear Systems with Exponentially Convergent Bounds," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 4, pp. 280–301, 2012.

[163] G. Chowdhary, T. Wu, M. Cutler, N. K. Ure, and J. P. How, "Experimental Results of Concurrent Learning Adaptive Controllers," *AIAA Guidance, Navigation and Control Conference,*, pp. 1–14, 2012.

[164] R. Kamalapurkar, J. R. Klotz, and W. E. Dixon, "Concurrent Learning-based Approximate Feedback-Nash Equilibrium Solution of N-player Nonsero-sum Differential Games," *IEEE/CAA Journal of Automatica Sinica*, vol. 1, pp. 239–247, 2014.

[165] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic course*. Boston, MA.: Kluwer Academic Publishers, 2004, vol. 87.

[166] D. Ochoa, J. Poveda, C. Uribe, and N. Quijano, "Robust Resource Allocation with Momentum," *Submitted.*, 2019.

[167] A. R. Teel, J. I. Poveda, and J. Le, "First-Order Optimization Algorithms with Resets and Hamiltonian flows," *In proc. of IEEE Conference on Decision and Control*, pp. 5838–5843, 2019.

[168] J. I. Poveda and M. Krstić, "Non-Smooth Extremum Seeking Control with User-Prescribed Fixed-Time Convergence," *IEEE Transactions on Automatic Control, 10.1109/TAC.2021.3063700*, 2021.

[169] M. Lab, "U.S. is the world leader in traffic jams – USA Today," 2018, https://mobilitylab.org/2018/02/06/u-s-is-the-world-leader-in-traffic-jams/.

[170] R. Kutadinata, W. Moase, C. Manzie, L. Zhang, and T. Garoni, "Enhancing the performance of existing urban traffic light control through extremum-seeking," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 1–20, 2016.

[171] P. N. Brown and J. R. Marden, "Studies on Robust Social Influence Mechanisms," *IEEE Control Systems Magazine*, vol. 37, no. 1, pp. 98–115, 2017.

[172] A. Kleiner, B. Nebel, and V. A. Ziparo, "A mechanism for dynamic ride sharing based on parallel auctions," *Proc. 22nd Int. Join Conference on Artificial Intelligence*, pp. 266–272, 2011.

[173] A. M. Annaswamy, Y. Guan, H. E. Tseng, H. Zhou, T. Phan, and D. Yanakiev, "Transactive control in smart cities," *Proceedings of the IEEE*, vol. 106, no. 4, pp. 518–537, 2018.

[174] T. Phan, A. M. Annaswamy, D. Yanakiev, and E. Tseng, "A model-based dynamic toll pricing strategy for controlling highway traffic," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 6245–6252.

[175] G. Zhang, X. Ma, and Y. Wang, "Self-Adaptive Tolling Strategy for Enhanced High-Occupancy Toll Lane Operations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 306–317, 2014.

[176] N. Zheng, G. Rérat, and N. Geroliminis, "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 133–148, 2016.

[177] P. Kachroo, S. Gupta, S. Agarwal, and K. Ozbay, "Optimal control for congestion pricing: Theory, simulation, and evaluation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1234–1240, 2016.

[178] G. Bianchin, J. Cortés, J. I. Poveda, and E. Dall'Anese, "Time-Varying Optimization of LTI Systems via Projected Primal-Dual Gradient Flows," *IEEE Transactions on Control of Network Systems*, 2021, iEEE Transactions on Control of Network Systems, DOI 10.1109/TCNS.2021.3112762.

[179] C. Lombardi, L. Picado-Santos, and A. M. Annaswamy, "Model-Based Dynamic Toll Pricing: An Overview," *Applied Sciences*, vol. 11, no. 11, p. 4778, 2021.

[180] J. I. Poveda and M. Krstić, "Non-Smooth Extremum Seeking Control with User-Prescribed Convergence," *IEEE Transactions on Automatic Control*, vol. 66, no. 12, pp. 6156–6163, 2021.

[181] L. Engelson and P. O. Lindberg, "Congestion pricing of road networks with users having different time values," in *Mathematical and computational models for congestion charging*. Springer, 2006, pp. 81–104.

[182] K. B. Ariyur and M. Krstić, *Real-Time Optimization by Extremum-Seeking Control*. Hoboken, NJ: Wiley, 2003.

[183] H. Yu, S. Koga, T. R. Oliveira, and M. Krstić, "Extremum seeking for traffic congestion control with a downstream bottleneck," *Journal of Dynamic Systems, Measurement, and Control*, vol. 143, no. 3, p. 031007, 2021.

[184] L.-O. Sánchez Zamora, "Dynamic toll pricing for express lanes: a model-free feedback control approach," Master's thesis, Universitat Politècnica de Catalunya, 2020.

[185] K. Garg and D. Panagou, "Fixed-Time Stable Gradient Flows: Applications to Continuous-Time Optimization," *IEEE Transactions on Automatic and Control, DOI 10.1109/TAC.2020.3001436*, 2020.

[186] F. Zhu and S. V. Ukkusuri, "A reinforcement learning approach for distance-based dynamic tolling in the stochastic network environment," *Journal of Advanced Transportation*, vol. 49, no. 2, pp. 247–266, 2015.

[187] E. Moulay and W. Perruquetti, "Finite time stability and stabilization of a class of continuous systems," *Journal of Mathematical analysis and applications*, vol. 323, no. 2, pp. 1430–1443, 2006.

[188] K. Zimenko, D. Efimov, and A. Polyakov, "Stability analysis and stabilization of systems with hyperexponential rates," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 815–820, 2023.

[189] F. Lopez-Ramirez, D. Efimov, and W. P. c. A. Polyakov, "Finite-time and fixed-time input-to-state stability: Explicit and implicit approaches," *Systems & Control Letters*, vol. 144, 2020.

[190] J. I. Poveda, M. Krstić, and T. Başar, "Fixed-Time Seeking and Tracking of Time-Varying Nash Equilibria in Noncooperative Games," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 794–799.

[191] Y. Song, Y. Wang, J. Holloway, and M. Krstić, "Time-varying feedback for regulation of normal-form nonlinear systems in prescribed finite time," *Automatica*, vol. 83, pp. 243–251, 2017.

[192] Y. Orlov, "Time space deformation approach to prescribed-time stabilization: Synergy of time-varying and non-Lipschitz feedback designs," *Automatica*, vol. 144, p. 110485, 2022.

[193] P. Krishnamurthy, F. Khorrami, and M. Krstić, "A dynamic high-gain design for prescribed-time regulation of nonlinear systems," *Automatica*, vol. 115, p. 108860, 2020.

[194] ——, "Robust adaptive prescribed-time stabilization via output feedback for uncertain nonlinear strict-feedback-like systems," *European Journal of Control*, vol. 55, pp. 14–23, 2020.

[195] D. Tran and T. Yucelen, "Finite-time control of perturbed dynamical systems based on a generalized time transformation approach," *Systems & Control Letters*, vol. 136, p. 104605, 2020.

[196] V. Todorovski and M. Krstić, "Practical prescribed-time seeking of a repulsive source by unicycle angular velocity tuning," *Automatica*, vol. 154, p. 111069, 2023.

[197] B. T. Lopez and J.-J. E. Slotine, "Universal Adaptive Control of Nonlinear Systems," *IEEE Control Systems Letters*, vol. 6, pp. 1826–1830, 2022.

[198] Y. Song, H. Ye, and F. Lewis, "Prescribed-Time Control and Its Latest Developments," *IEEE Trans. on Systems, Man, and Cybernetics: Systems*, pp. 1–15, 2023.

[199] R. G. Sanfelice, *Hybrid Feedback Control*. Princeton University Press, 2021.

[200] N. Espitia and W. Perruquetti, "Predictor-Feedback Prescribed-Time Stabilization of LTI Systems With Input Delay," *IEEE Trans. Autom. Contr.*, vol. 67, no. 6, pp. 2784–2799, 2022.

[201] N. Espitia, A. Polyakov, D. Efimov, and W. Perruquetti, "Boundary time-varying feedbacks for fixed-time stabilization of constant-parameter reaction-diffusion systems," *Automatica*, vol. 103, pp. 398–407, 2019.

[202] W. Li and M. Krstić, "Prescribed-time output-feedback control of stochastic nonlinear systems," *IEEE Trans. on Automatic Control*, vol. 68, no. 3, pp. 1431–1446, 2022.

[203] A. Polyakov, "Nonlinear Feedback Design for Fixed-Time Stabilization of Linear Control Systems," *IEEE Transactions on Automatic and Control*, vol. 57, no. 8, pp. 2106–2110, 2012.

[204] Y. Dvir, A. Levant, D. Efimov, A. Polyakov, and W. Perruquetti, "Acceleration of finite-time stable homogeneous systems," *Int. Journal of Robust and Nonlinear Control*, vol. 28, no. 5, pp. 1757–1777, 2018.

[205] D. Efimov, A. Levant, A. Polyakov, and W. Perruquetti, "Supervisory Acceleration of Convergence for Homogeneous Systems," *Intl. Journal of Ctrl.*, vol. 91, no. 11, pp. 2524–2534, 2018.

[206] D. Efimov, A. Polyakov, A. Levant, and W. Perruquetti, "Convergence acceleration for observers by gain commutation," *Intl. Journal of Ctrl.*, vol. 91, no. 9, pp. 2009–2018, 2018.

[207] X. He, X. Li, and S. Song, "Prescribed-Time Stabilization of Nonlinear Systems via Impulsive Regulation," *IEEE Trans. on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 2, pp. 981–985, 2023.

[208] G. L. Slater and W. R. Wells, "Optimal evasive tactics against a proportional navigation missile with time delay." *Journal of Spacecraft and Rockets*, vol. 10, no. 5, pp. 309–313, 1973.

[209] D. Steeves, M. Krstić, and R. Vazquez, "Prescribed–time estimation and output regulation of the linearized Schrödinger equation by backstepping," *European Journal of Control*, vol. 55, pp. 3–13, 2020.

[210] A. Irscheid, N. Espitia, W. Perruqueti, and J. Rudolph, "Prescribed-time control for a class of semilinear hyperbolic PDE-ODE systems," in *Proc of the 4th IFAC Control Systems Governed by Partial Differential Equations (CPDE)*, 2022.

[211] S. Zekraoui, N. Espitia, and W. Perruquetti, "Prescribed-time predictor control of LTI systems with distributed input delay," in *60th IEEE Conf. on Decision and Control (CDC)*, Dec. 2021, pp. 1850–1855.

[212] N. Espitia, D. Steeves, W. Perruquetti, and M. Krstić., "Sensor delay-compensated prescribed-time observer for LTI systems," *Automatica*, vol. 135, p. 110005, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109821005318

[213] J. Holloway and M. Krstić, "Prescribed-time output feedback for linear systems in controllable canonical form," *Automatica*, vol. 107, pp. 77–85, 2019.

[214] F. Gao, Y. Wu, and Z. Zhang, "Global fixed-time stabilization of switched nonlinear systems: a time-varying scaling transformation approach," *IEEE Trans. on Circuits and Systems II: Exp. Briefs*, vol. 66, no. 11, pp. 1890–1894, 2019.

[215] Y. Orlov, R. I. V. Kairuz, and L. T. Aguilar, "Prescribed-time robust differentiator design using finite varying gains," *IEEE Control Systems Letters*, vol. 6, pp. 620–625, 2021.

[216] S. Liu, A. Tanwani, and D. Liberzon, "ISS and integral-ISS of switched systems with nonlinear supply functions," *Mathematics of Control, Signals, and Systems*, vol. 34, pp. 297–327, 2022.

[217] D. Liberzon, *Switching in Systems and Control*. Boston, MA.: Birkhauser, 2003.

[218] A. R. Teel, J. I. Poveda, and J. Le, "First-order optimization algorithms with resets and Hamiltonian flows," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5838–5843.

[219] G. Yang and D. Liberzon, "Input-to-state stability for switched systems with unstable subsystems: A hybrid Lyapunov construction," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 6240–6245.

[220] C. Prieur, I. Queinnec, S. Tarbouriech, L. Zaccarian *et al.*, "Analysis and synthesis of reset control systems," *Foundations and Trends® in Systems and Control*, vol. 6, no. 2-3, pp. 117–338, 2018.

[221] W. Wang, A. R. Teel, and D. Nešić, "Averaging in singularly perturbed hybrid systems with hybrid boundary layer systems," *51st IEEE Conference on Decision and Control*, vol. 6855-6860, pp. 6855–6860, 2012.

[222] A. R. Teel and D. Nešić, "Averaging Theory for a Class of Hybrid Systems," *Dynamics of Continuous, Discrete and Impulsive Systems*, vol. 17, pp. 829–851, 2010.

[223] A. R. Teel, F. Forni, and L. Zaccarian, "Lyapunov-based sufficient conditions for exponential stability in hybrid systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1591–1596, 2013.

[224] J. P. Hespanha and A. S. Morse, "Stabilization of switched systems with average dwell-time," *38th IEEE Conference on Decision and Control*, pp. 2655–2660, 1999.

[225] J. Hespanha and A. Morse, "Stability of switched systems with average dwell-time," *38th IEEE Conf. Decision Control*, vol. 3, pp. 2655–2660, 1999.

[226] G. Yang and D. Liberzon, "A Lyapunov-based small-gain theorem for interconnected switched systems," *Systems and Control Letters*, vol. 78, pp. 47–54, 2015.

[227] S. Li, J. Guo, and Z. Xiang, "Global stabilization of a class of switched nonlinear systems under sampled-data control," *IEEE Trans. on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 9, pp. 1912–1919, 2018.

[228] J. W. Grizzle, C. Chevallereau, R. W. Sinnet, and A. D. Ames, "Models, feedback control, and open problems of 3D bipedal robotic walking," *Automatica*, vol. 50, no. 8, pp. 1955–1988, 2014.

[229] Y. Zhao, Q. Tao, C. Xian, Z. Li, and Z. Duan, "Prescribed-time distributed Nash equilibrium seeking for noncooperation games," *Automatica*, vol. 151, p. 110933, 2023.

[230] A. Polyakov, D. Efimov, and B. Brogliato, "Consistent Discretization of Finite-Time and Fixed-Time Stable Systems," *SIAM Journal on Control and Optimization*, vol. 57, no. 1, 2019.

[231] J. Hauser, "A projection operator approach to the optimization of trajectory functionals," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 377–382, 2002.

[232] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[233] S. Grivopoulos and B. Bamieh, "Lyapunov-based control of quantum systems," *42nd IEEE International Conference on Decision and Control*, vol. 1, pp. 434–438 Vol.1, 2003.

[234] D. Gabay, "Minimizing a differentiable function over a differential manifold," *Journal of Optimization Theory and Applications*, vol. 37, no. 2, pp. 177–219, 1982.

[235] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[236] E. Sontag, "Stability and Stabilization: Discontinuities and the Effect of Disturbances," in *Nonlinear Analysis, Differential Equations and Control*, ser. NATO Science Series. Springer, 1999, vol. 528, pp. 551–598.

[237] U. Helmke and J. B. Moore, *Optimization and dynamical systems*. Springer Science & Business Media, 2012.

[238] S. P. Bhat and D. S. Bernstein, "A topological obstruction to continuous global stabilization of rotational motion and the unwinding phenomenon," *Systems & Control letters*, vol. 39, no. 1, pp. 63–70, 2000.

[239] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer, 2013, vol. 6.

[240] D. Angeli, "An almost global notion of input-to-state stability," *IEEE Trans. on Autom. Ctrl.*, vol. 49, no. 6, pp. 866–874, 2004.

[241] D. Efimov, "Global Lyapunov analysis of multistable nonlinear systems," *SIAM Journal on Ctrl. and Optim.*, vol. 50, no. 5, pp. 3132–3154, 2012.

[242] C. G. Mayhew and A. R. Teel, "Synergistic potential functions for hybrid control of rigid-body attitude," in *2011 American Ctrl. Conf. (ACC)*. IEEE, 2011, pp. 875–880.

[243] J.-M. Coron, "Global asymptotic stabilization for controllable systems without drift," *Mathematics of Control, Signals and Systems*, vol. 5, no. 3, pp. 295–312, 1992.

[244] M. Malisoff, M. Krichman, and E. Sontag, "Global stabilization for systems evolving on manifolds," *Journal of Dynamical and Control Systems*, vol. 12, no. 2, pp. 161–184, 2006.

[245] C. G. Mayhew and A. R. Teel, "On the topological structure of attraction basins for differential inclusions," *Systems & Control Letters*, vol. 60, no. 12, pp. 1045–1050, 2011.

[246] C. G. Mayhew, *Hybrid control for topologically constrained systems*. University of California, Santa Barbara, 2010.

[247] S. Berkane, A. Abdessameud, and A. Tayebi, "Hybrid global exponential stabilization on SO(3)," *Automatica*, vol. 81, pp. 279–285, 2017.

[248] P. Casau, R. G. Sanfelice, and C. Silvestre, "Adaptive backstepping of synergistic hybrid feedbacks with application to obstacle avoidance," in *2019 American Ctrl. Conf. (ACC)*, 2019, pp. 1730–1735.

[249] P. Casau, C. G. Mayhew, R. G. Sanfelice, and C. Silvestre, "Robust global exponential stabilization on the n-dimensional sphere with applications to trajectory tracking for quadrotors," *Automatica*, vol. 110, p. 108534, 2019.

[250] M. Krstić and H.-H. Wang, "Stability of extremum seeking feedback for general nonlinear dynamic systems," *Automatica*, vol. 36, no. 4, pp. 595–601, 2000.

[251] H.-B. Dürr, M. S. Stanković, K. H. Johansson, and C. Ebenbauer, "Extremum seeking on submanifolds in the Euclidian space," *Automatica*, vol. 50, no. 10, pp. 2591–2596, 2014.

[252] F. Taringoo, P. M. Dower, D. Nesic, and Y. Tan, "Optimization Methods on Riemannian Manifolds via Extremum Seeking Algorithms," *SIAM Journal on Ctrl. and Optim.*, vol. 56, no. 5, pp. 3867–3892, 2018.

[253] T. Strizic, J. I. Poveda, and A. R. Teel, "Hybrid gradient descent for robust global optimization on the circle," in *56th Conf. on Decision and Ctrl. (CDC)*. IEEE, 2017, pp. 2985–2990.

[254] R. Suttner, "Extremum Seeking Control for a Class of Mechanical Systems," *IEEE Trans. on Autom. Ctrl.*, 2022.

[255] R. Suttner and M. Krstić, "Nonlocal Nonholonomic Source Seeking Despite Local Extrema," *IEEE Trans. on Autom. Ctrl.*, 2023.

[256] J. M. Lee, *Introduction to Smooth Manifolds*. Springer, 2013.

[257] ——, *Introduction to Riemannian Manifolds*. Springer, 2018.

[258] C. Udriste, *Convex functions and optimization methods on Riemannian Manifolds*. Springer Science & Business Media, 2013, vol. 297.

[259] D. Angeli and D. Efimov, "Characterizations of input-to-state stability for systems with multiple invariant sets," *IEEE Trans. on Autom. Ctrl.*, vol. 60, no. 12, pp. 3242–3256, 2015.

[260] J. Milnor, *Lectures on the h-cobordism theorem*. Princeton University Press, 2015, vol. 2258.

[261] C. K. Lauand and S. Meyn, "Quasi-stochastic approximation: Design principles with applications to extremum seeking control," *IEEE Control Systems Magazine*, vol. 43, no. 5, pp. 111–136, 2023.

[262] B. C. Hall and B. C. Hall, *Lie groups, Lie algebras, and representations*. Springer, 2013.

[263] J. Gallier and J. Quaintance, *Differential geometry and Lie groups: A Computational Persepective*. Springer, 2020, vol. 12.

[264] M. Baradaran, J. I. Poveda, and A. R. Teel, "Global optimization on the sphere: A stochastic hybrid systems approach," *IFAC-PapersOnLine*, vol. 52, no. 16, pp. 96–101, 2019.

[265] P. Kapitza *et al.*, "Dynamic stability of a pendulum with an oscillating point of suspension," *Journal of experimental and theoretical physics*, vol. 21, no. 5, pp. 588–597, 1951.

[266] E. I. Butikov, "On the dynamic stabilization of an inverted pendulum," *American Journal of Physics*, vol. 69, no. 7, pp. 755–768, 2001.

[267] P. S. Landa, *Nonlinear oscillations and waves in dynamical systems*. Springer Science & Business Media, 2013, vol. 360.

[268] S. Meerkov, "Principle of vibrational control: theory and applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 4, pp. 755–762, 1980.

[269] W. Wang, A. R. Teel, and D. Nešić, "Analysis for a class of singularly perturbed hybrid systems via averaging," *Automatica*, pp. 1057–1068, 2012.

[270] C. G. Mayhew and A. R. Teel, "On the topological structure of attraction basins for differential inclusions," *Systems & Control Letters*, vol. 60, no. 12, pp. 1045–1050, 2011.

[271] E. Bernuau, W. Perruquetti, and E. Moulay, "Retraction obstruction to time-varying stabilization," *Automatica*, vol. 49, no. 6, pp. 1941–1943, 2013.

[272] W. Wang, A. Teel, and D. Nešić, "Analaysis for a class of singularly perturbed hybrid systems via averaging," *Automatica*, vol. 48, no. 6, pp. 1057–1068, 2012.

[273] C. Cai and A. R. Teel, "Characterizations of input-to-state stability for hybrid systems," *Systems & Control Letters*, vol. 58, no. 1, pp. 47–53, 2009.

[274] A. Beck, *First-order methods in optimization*. SIAM, 2017.

[275] A. R. Teel and J. I. Poveda, "A hybrid systems approach to global synchronization and coordination of multi-agent sampled-data systems," *IFAC-PapersOnLine*, vol. 48, no. 27, pp. 123–128, 2015.

[276] H. Zhang, Z. Li, Z. Qu, and F. L. Lewis, "On constructing Lyapunov functions for multi-agent systems," *Automatica*, vol. 58, pp. 39–42, 2015.

[277] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*. Princeton university press, 2009.

[278] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 2012.

[279] S. P. Bhat and D. S. Bernstein, "Finite-Time Stability of Continuous Autonomous Systems," *SIAM J. Control Optim.*, vol. 38, pp. 751–766, 2000.

[280] S. Parsegov, A. Polyakov, and P. Shcherbakov, "Nonlinear fixed-time control protocol for uniform allocation of agents on a segment," *IEEE Conference on Decision and Control*, pp. 7732–7737, 2012.

[281] C. W. Ho, "A note on proper maps," *Proceedings of the American Mathematical Society*, vol. 51, no. 1, pp. 237–241, 1975.

[282] W. B. Gordon, "On the diffeomorphisms of Euclidean space," *The American Mathematical Monthly*, vol. 79, no. 7, pp. 755–759, 1972.