Set-Valued Methods for Reachability Analysis and Estimation of Nonlinear Dynamical Systems

by

Mohammad Khajenejad

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved August 2021 by the
Graduate Supervisory Committee:

Sze Zheng Yong, Chair
Angelia Nedich
Kevin Reffett
Spring M. Berman
Georgios Fainekos
Hyunglae Lee

ARIZONA STATE UNIVERSITY

December 2021

ABSTRACT

The goal of this thesis research is to contribute to the design of set-valued methods, i.e., algorithms that leverage a set-theoretic framework that can provide a powerful means for control designs for general classes of uncertain nonlinear dynamical systems, and in particular, to develop set-valued algorithms for constrained reachability problems and estimation.

I propose novel fixed-order hyperball-valued observers for different classes of nonlinear systems, including Linear Parameter Varying, Lipschitz continuous and Decremental Quadratic Constrained nonlinearities, with unknown inputs that simultaneously find bounded sets of states and unknown inputs that contain the true states and inputs and are compatible with the measurement/outputs. In addition, I provide sufficient conditions for the existence and stability of the estimates, convergence of the estimation errors and optimality of the observers.

Moreover, I design state and unknown input observers as well as mode detectors for hidden mode switched linear and nonlinear systems with bounded-norm noise and unknown inputs. To address this, I propose a multiple-model approach to obtain a bank of mode-matched set-valued observers in combination with a novel mode observer, based on elimination. My mode elimination approach uses the upper bound of the norm of to-be-designed residual signals to remove inconsistent modes from the bank of observers. I also provide sufficient conditions for mode detectability.

Furthermore, I address the problem of designing interval observers for partially unknown nonlinear systems, using affine abstractions, nonlinear decomposition functions and a data-driven function over-approximation approach to over-estimate the unknown dynamic model. The proposed observer recursively computes the correct interval estimates. Then, using observed measurement signals, the observer iteratively shrinks the intervals. Moreover, the observer updates the over-approximation model of the unknown dynamics.

Finally, I propose a tractable family of remainder-from decomposition functions for a broad-range of dynamical systems. Moreover, I introduce a set-inversion algorithm that along with the proposed decomposition functions have several applications, e.g., in approximation of the reachable sets for bounded-error, constrained, continuous and/or discrete-time systems, as well as in guaranteed state estimation. Leveraging mixed-monotonicity, I provide novel set-theoretic approaches to address the problem of polytope-valued state estimation in bounded-error discrete-time nonlinear systems, subject to nonlinear observations/constraints.

To my love: Fatemeh!

ACKNOWLEDGEMENT

First of all, I would like to thank my advisor, Dr. Sze Zheng Yong, for his invaluable supervision, support and mentorship during the course of my Ph.D. degree. I cherish the significant impact he has had on me, not only in research and academics, but also in learning to build strong work ethics and principles. My gratitude extends to the faculty of Mechanical Engineering at Arizona State University (ASU) for their valuable instructions and the opportunity they have provided for me to undertake my studies.

Additionally, I would like to express gratitude to Prof. Kevin Reffett for his treasured support since the time I started my academic journey in the US at ASU as my advisor in Economics at the W.P. Carrey School of Business, and also for his ongoing reinforcement and encouragement as I transitioned to engineering. I would also like to thank my other graduate committee members, Prof. Angelia Nedich, Dr. Spring Berman, Dr. Georgios Fainekos and Dr. Hyunglae Lee for critiquing my work and providing their beneficial feedback which has been really influential in improving my research.

I would like to thank my lab mates and research team, Qiang, Zeyuan, Hassaan, Ruochen, Fatima and Tarun for the valuable times we spent together in the lab and on the excellent virtual group meetings, as well as for their friendship and helpfulness in social settings. Last but not least, my appreciation goes out to my family for their unconditional support: my parents and parents-in-law, Maman Maryam, Baba Mahmoud, Baba Jaan and Maman Jaan, my bothers, brother-in-law and her wife, Amin, Ahmad, Moein, Mohsen and Zahra, my loving wife, Fatemeh, my best pal, Taha, and my sweet angel, Marsa.

TABLE OF CONTENTS

Page

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER

1 INTRODUCTION 1

1.1 Contributions of the Research 11

1.2 Published/Under Review/ In Preparation Content 17

1.3 Other Research Contributions 19

1.4 Notation 19

2 SIMULTANEOUS INPUT AND STATE SET-VALUED $\mathcal{H}_\infty$-OBSERVERS FOR LINEAR PARAMETER-VARYING SYSTEMS 21

2.1 Problem Statement 22

2.2 Preliminary Material 23

2.2.1 System Transformation 23

2.3 Fixed-Order Simultaneous Input and State Set-Valued Observers ... 24

2.3.1 Set-Valued Observer Design 24

2.3.2 Optimal $\mathcal{H}_\infty$-Observer 27

2.4 Simulation Results 32

2.5 Conclusion 33

3 SIMULTANEOUS MODE, INPUT AND STATE SET-VALUED OBSERVERS WITH APPLICATIONS TO RESILIENT ESTIMATION AGAINST SPARSE ATTACKS 35

3.1 Problem Statement 35

3.2 Prpposed Observer Design 37

3.2.1 Overview of Multiple-Model Approach 37

3.2.2 Mode Elimination Approach 40
3.2.3 Tractable Computation of Thresholds 41
3.3 Mode Detectability 46
3.4 Simulation Results 49
3.5 Conclusion 52
4 SIMULTANEOUS STATE AND UNKNOWN INPUT SET-VALUED OBSERVERS FOR NONLINEAR DYNAMICAL SYSTEMS 53
4.1 Preliminary Material 54
4.1.1 Structural Properties 54
4.2 Problem Statement 59
4.3 More General Nonlinear System Model 61
4.4 Fixed-Order Simultaneous Input and State Set-Valued Observer Framework 62
4.4.1 System Transformation 63
4.4.2 Observer Structure 64
4.5 Observer Design and Analysis 65
4.5.1 Stable Observer Design 66
4.5.2 $\mathcal{H}_\infty$ Observer Design 69
4.5.3 Radii of Estimates and Convergence of Errors 72
4.6 Simulation Results and Comparison with Benchmark Observers 76
4.6.1 Single-Link Flexible-Joint Robotic System 76
4.6.2 Comparison with Benchmark Observers 80
4.7 Conclusion 85

5 SIMULTANEOUS MODE, STATE AND INPUT SET-VALUED OBSERVERS FOR SWITCHED NONLINEAR SYSTEMS 87
5.1 Problem Formulation 87
5.2 Proposed Observer Design 89
5.2.1 Overview of Multiple-Model Approach 90
5.2.2 Mode Observer 94
5.2.3 Mode Elimination Approach 94
5.2.4 Tractable Computation of Thresholds 96
5.3 Mode Detectability 99
5.4 Simulation Results 102
5.5 Conclusion 107
6 SIMULTANEOUS INPUT AND STATE INTERVAL OBSERVERS FOR NONLINEAR SYSTEMS 108
6.1 Preliminary Material 108
6.2 Problem Formulation 111
6.3 General Simultaneous Input and State Interval Observers (GSISIO) 113
6.3.1 Interval Observer Design 113
6.3.2 Observer Design 116
6.3.3 Uniform Boundedness of Estimates (Observer Stability) 119
6.4 Simulation Results 121
6.5 Conclusion 124
7 INTERVAL OBSERVERS FOR SIMULTANEOUS STATE AND MODEL ESTIMATION OF PARTIALLY KNOWN NONLINEAR SYSTEMS 125

7.1 Problem Formulation ... 125
7.2 State and Model Interval Observers (SMIO) ... 127
  7.2.1 Correctness of the Observer ... 131
  7.2.2 Observer Stability ... 134
7.3 Simulation Results ... 137
7.4 Conclusion ... 141
8 TIGHT REMAINDER-FORM DECOMPOSITION FUNCTIONS (WITH APPLICATIONS TO CONSTRAINED REACHABILITY AND INTERVAL OBSERVER DESIGN) ... 142
8.1 Preliminary Material ... 142
8.2 Problem Statement ... 149
8.3 Main Results ... 151
  8.3.1 Remainder-Form Decomposition Functions ... 151
  8.3.2 Error Bounds ... 154
  8.3.3 Linear Remainders ... 156
  8.3.4 Convergence Rate and Further Refinement ... 160
  8.3.5 Set-Inversion ... 161
8.4 Comparison with Other Inclusion Functions ... 164
  8.4.1 Comparison with the T_L Inclusion Functions ... 164
  8.4.2 Comparison with T_N, T_C and T_M Inclusion Functions ... 165
8.5 Applications ... 167
  8.5.1 Application to Constrained Reachability Analysis ... 167
  8.5.2 Application to State Estimation ... 168
8.6 Simulation Results ... 169

8.6.1 The Van Der Pol System 169
8.6.2 Example 3 in [130] 170
8.6.3 Example 2.11 in [57] 173
8.6.4 Continuous-Time System in [2] 174
8.6.5 The Unicycle System 175
8.6.6 Generic Transport Longitudinal Model 179
8.7 Conclusion 181
9 GUARANTEED STATE ESTIMATION VIA INDIRECT POLYTOPIC SET COMPUTATION FOR NONLINEAR DISCRETE-TIME SYSTEMS 183
9.1 Preliminaries 183
9.2 Problem Formulation 187
9.3 Indirect Polytopic Set Computation 189
9.3.1 Decomposition-Based ZB/CZ Propagation Step 189
9.3.2 Decomposition-Based CZ/ZB Update Step 191
9.3.3 Modifications to the Approach in [99] 193
9.4 Simulation Results 195
9.4.1 Example I 196
9.4.2 Example II (The Unicycle System) 198
9.5 Conclusion 199
REFERENCES 201

APPENDIX

A PROOFS 211

LIST OF TABLES

Table Page

3.1 Different Modes and Their T_2^q 50

5.1 Different Modes and Their T_2^q in Scenario I (i.e., with Bounded d_k) . . . 105

5.2 Different Modes and Their T_2^q in Scenario II (i.e., with Unbounded d_k) 105

9.1 Average Total Times (Seconds) and Average Total Volumes at Each Time Step for Five State Estimators in Example I: Each Average Is Taken Over 50 Simulations with Uniformly Sampled Noise and Initial State 197

9.2 Average Total Times (Seconds) and Average Total Volumes (10^{-5}) at Each Time Step for Five State Estimators in Example II: Each Average Is Taken Over 20 Simulations with Uniformly Sampled Noise and Initial State 200

LIST OF FIGURES

Figure Page

1.1 Some Set-theoretic Approaches in Control/Reachability Analysis 2

1.2 Safety-critical Cyber-physical Systems Are Susceptible to Attacks (Unknown Inputs) 4

2.1 Actual States x_1, x_2 and Their Estimates, as Well as Unknown Inputs d_1 and d_2 and Their Estimates 33

2.2 Actual Estimation Errors and Radii of Set-valued Estimates of States, $\|\tilde{x}_{k|k}\|, \delta^x_k$, and Unknown Inputs,$\|\tilde{d}_k\|, \delta^d_k$..... 33

3.1 $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes, as Well as the Number of Eliminated Modes in Time 51

3.2 State and Unknown Input Set-valued Estimates 51

4.1 Relationships Between Different Classes of Nonlinearities: $\Rightarrow$ Denotes Direct Implication, While $\rightarrow$ Denotes Implication with Addition Assumptions 59

4.2 Actual States x_1,x_2, as Well as Their Class 0 Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case (0) in Theorem 4.5.8), Class I Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case I in Theorem 4.5.8) and Class II Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case II in Theorem 4.5.8) 78

4.3 Actual States x_3,x_4 and Input d, as Well as Their Class 0 Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case (0) in Theorem 4.5.8), Class I Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case I in Theorem 4.5.8) and Class II Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case II in Theorem 4.5.8) 79

4.4 Estimation Errors and Their Upper Bounds for Class 0 (General Nonlinear), Class I (Lipschitz) and Class II (DQC*) Functions 80

4.5 Actual States x_1, x_2, and Their Estimates, as Well as Unknown Input d and Its Estimates in the Bounded Unknown Input Scenario, Obtained by Applying the Observer in [25] (Chen-Hu Estimate), the Observer in [23] (Chak-Stan-Shre Estimate) and Our Designed observer (SISO Estimate) 82

4.6 Estimation Errors in the Bounded Unknown Input Scenario for 50 Different Initial Values (Using Box Plots), Obtained by Applying the Observer in [25] (Chen-Hu Err.), the observer in [23] (Chak-Stan-Shre Err.) and Our Designed Observer (SISO Err.), as Well as the Computed Upper Bounds for the State and Input Errors (δ_k^x and δ_k^d) 83

4.7 Actual States x_1, x_2, and Their Estimates, as Well as Unknown Input d and Its Estimates in the Unbounded Unknown Input Scenario, Obtained by Applying the Observer in [25] (Chen-Hu Estimate), the Observer in [23] (Chak-Stan-Shre Estimate) and Our Designed Observer (SISO Estimate) 84

Figure Page

4.8 Estimation Errors in the Unbounded Unknown Input Scenario for 50 Different Initial Values (Using Box Plots), Obtained by Applying the Observer in [25] (Chen-Hu Err.), the Observer in [23] (Chak-Stan-Shre Err.) and Our Designed Observer (SISO Err.), as Well as the Computed Upper Bounds for the State and Input Errors (δ_k^x and δ_k^d) 85

5.1 Actual states x_1, x_2, and their estimates, as well as the unknown input d and its estimates, and the number of non-eliminated modes at each time step in the bounded unknown input scenario (Scenario I), when applying the observer in [141] (Zhen-Xu-Zhang Estimate) and our proposed observer (SMIS Estimate) 104

5.2 Actual States x_1, x_2, and Their Estimates, as Well as the Unknown Input d and its Estimates, and the Number of Non-eliminated Modes at Each Time Step in the Unbounded Unknown Input Scenario (Scenario II), When Applying the Observer in [141] (Zhen-Xu-Zhang Estimate) and Our Oroposed Observer (SMIS Estimate) 104

5.3 $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes in the Bounded Unknown Input Scenario (Scenario I) 106

5.4 $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes in the Unbounded Unknown Input Scenario (Scenario II) 106

6.1 Actual states and inputs, $x_{1,k}$, $x_{2,k}$, $d_{1,k}$, $d_{2,k}$, as well as their estimated maximal and minimal values 123

6.2 Estimation Errors, Estimate Interval Widths and Their Upper Bounds for the Interval-valued Estimates of States, $\|\tilde{x}_{k|k}\|$, $\|\Delta^x_k\|$, δ^x_k, and Unknown Inputs, $\|\tilde{d}_k\|$, $\|\Delta^d_k\|$, δ^d_k 123

7.1 Actual States, as Well as Their Estimated Maximal and Minimal Values 138

7.2 Actual Unknown Input, d_k, as Well as Its Estimated (Learned) Maximal and Minimal Values 139

7.3 Actual Estimation Errors, Interval Estimate Widths and Their Upper Bounds for the Interval Estimates of States and Unknown Inputs 140

7.4 Actual Unknown Dynamics Function $h(\zeta)$, Its Upper and Lower Framer Intervals at Time Step $k = 250$, as Well as Its Global Abstraction 141

8.1 Upper and Lower Bounds on x_1 and x_2 in System (8.32) (the Van Der Pol System), Applying $T_N(--)$, T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), the Best of T_N–T_R ($\cdot$-) and T_O ($+$), as Well as the True Trajectory ($-$) 170

8.2 Upper and Lower Bounds on x_1 and x_2 in System (8.33), Applying $T_N(--)$, T_{S_1} ($\triangle$), T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), T_R with Manufactured Redundancies ($\times\cdot$) and T_O ($+$), as Well as the True Trajectory ($-$) 172

8.3 Upper and Lower Bounds on x_1 and x_2 in System (8.36), (i.e., (8.33) Plus Observations), Applying $T_N(--)$, T_{S_1} ($\triangle$), T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), T_R with Manufactured Redundancies ($\times\cdot$) and T_O ($+$), as Well as the True Trajectory ($-$). 173

8.4 Upper and Lower Bounds on x_1 and x_2 in System (8.37), Applying $T_N(--)$, T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$) and the Best of T_N–T_R ($\cdot$-), as Well as the True Trajectory ($-$) 174

8.5 Upper and Lower Bounds for x_1, x_2 and x_3 in System (8.38), applying $T_N(--)$, T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), the Best of T_N–T_R ($\cdot$-) and T_O ($+$), as Well as the True Trajectory ($-$) 175

8.6 Upper and Lower Bounds on x_1 and x_2 in System (8.39) (the Unicycle System Without Observation), Applying T_N(−−), T_C (○), T_M (◇), T_L (□), T_R (∗), the Best of T_N–T_R (·-) and T_O (+), as Well as the True Trajectory (−) 177

8.7 Upper and Lower Bounds on x_1 and x_2 in System (8.39)–(8.40) (the Unicycle System with Observation), Applying T_N(−−), T_C (○), T_M (◇), T_L (□), T_R (∗), the Best of T_N–T_R (·-) and T_O (+), as Well as the True Trajectory (−) 178

8.8 Upper and Lower Bounds on x_1 and x_2 in System (8.41) (the GTM System Without Observation), Applying T_N(−−), T_C (○), T_M (◇), T_L (□), T_R (∗) and the Best of T_N–T_R (·-), as Well as the True Trajectory (−) 180

8.9 Upper and Lower Bounds on x_1 and x_2 in System (8.41)–(8.42) (the GTM System with Observation), Applying T_N(−−), T_C (○), T_M (◇), T_L (□), T_R (∗) and the Best of T_N–T_R (·-), as Well as the True Trajectory (−) 181

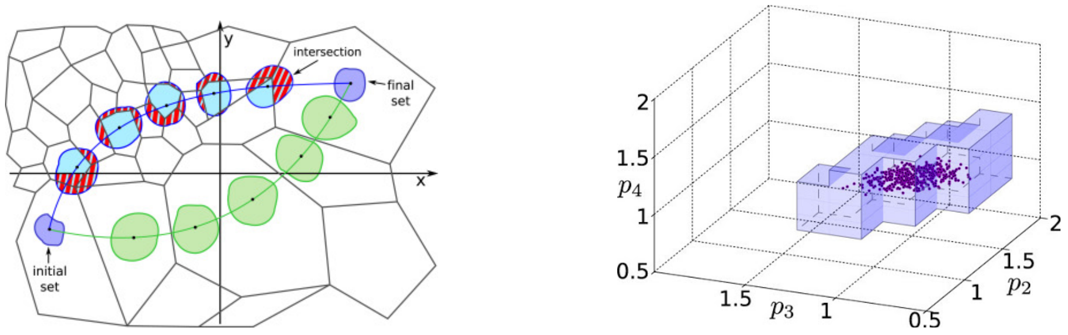
9.1 Results for Example I from the First Five Time Steps of Set-valued State Estimation, Using Five Different Approaches: Black Dots Are Obtained from Uniform Sampling of the Initial State and Noise Signals, and Propagating Through the System Dynamics 196

9.2 Results for Example II from the First Five Time Steps of Set-valued State Estimation, Using Five Different Approaches: Black Dots Are Obtained from Uniform Sampling of the Initial State and Noise Signals, and Propagating Through the System Dynamics 199

Chapter 1

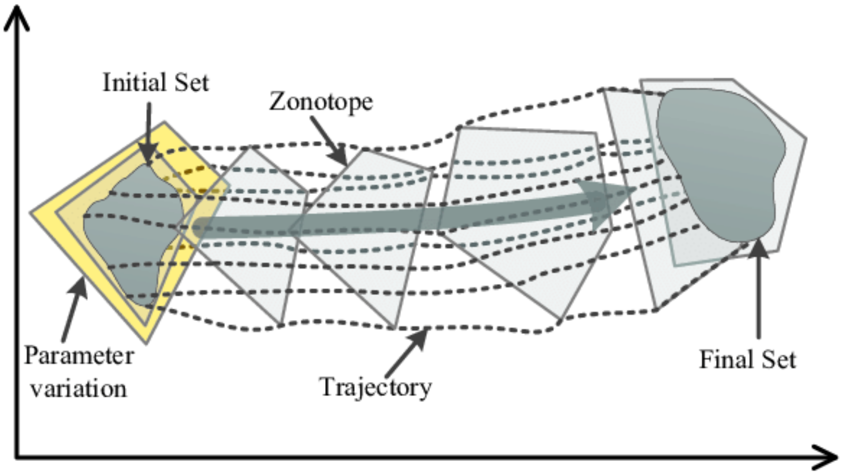
INTRODUCTION

Many problems in control systems and decision making can be naturally formulated, analyzed, and solved in a set-theoretic framework [16]. The reason for this is because control systems design often involve constraints, uncertainties, and design specifications, which can be naturally described using sets. Moreover, sets provide an appropriate framework to characterize system performance, e.g., for determining the domain of attraction, for computing reachable sets or for quantifying tracking/regulation control errors in feedback loops and estimation errors in inference and estimation problems. Thus, set-valued methods, i.e., algorithms that leverage a set-theoretic framework, can provide a powerful means for control designs for very general classes of uncertain nonlinear dynamical systems, and the goal of this thesis research is to contribute to the design of these tools, and in particular, to develop set-valued algorithms for constrained reachability problems and estimation (cf. Figure 1.1). In several engineering applications such as aircraft tracking, fault detection, attack (unknown input) detection and mitigation in Cyber-Physical Systems (CPS) and urban transportation [72, 136, 131], algorithms for unknown input reconstruction and state estimation have become increasingly indispensable and crucial to ensure their smooth and safe operation. Specifically, in safety-critical bounded-error systems, set/interval membership/reachability analysis methods have been applied to guarantee hard accuracy bounds. Further, in adversarial settings with potentially strategic unknown inputs that can be injected as counterfeit data signals by malicious agents into the sensor measurements and actuator signals to cause damage, steal energy etc [21, 44, 100, 109, 139], it is critical and desirable to simultaneously derive compatible



Set-based control

Set-based reachability



Set-based verification

Figure 1.1: Some Set-theoretic Approaches in Control/Reachability Analysis

estimates of states and unknown inputs, without assuming any *a priori* known bounds/intervals for the input signals. Given the strategic nature of these false data injection signals, they are not well-modeled by a zero-mean, Gaussian white noise nor by signals with known bounds. Hence, traditional Kalman filtering and unknown input observers do not apply. Nevertheless, reliable *set-valued* estimates of states and unknown inputs are indispensable and useful for the sake of attack identification, resilient control, etc. Similar state and input estimation problems can be found across a wide range of disciplines, from input estimation in physiological systems [36], to fault detection and diagnosis [91], to the estimation of mean areal precipitation [68].

Much of the research focus has been on simultaneous input and state estimation for stochastic systems with unknown inputs, assuming that the noise signals are Gaussian and white, via minimum variance unbiased (MVU) estimation approaches (e.g., [46, 47, 133, 135]), modified double-model adaptive estimation methods (e.g, [74]), or robust regularized least square approaches as in [3]. However, such Kalman filtering inspired approaches are not applicable for set-membership estimation problems in bounded-error settings, as is considered in this research, where *set-valued* uncertainties are considered and *sets* of states and unknown inputs that are compatible with measurements are desired (cf. [131] for a comprehensive discussion). In the context of attack-resilient estimation, numerous approaches were proposed for deterministic systems (e.g., [28, 45, 90, 107]), stochastic systems (e.g., [67, 134, 136]) and bounded-error systems [86, 89, 132], against false data injection attacks. However, these approaches mainly yield point estimates, i.e, the most likely or best single estimate, as opposed to set-valued estimates. On the other hand, the work in [89] only computes error bounds for the initial state and [86] assumes zero initial states and does not consider any optimality criteria.

In addition, unknown input observer designs for different classes of discrete-time

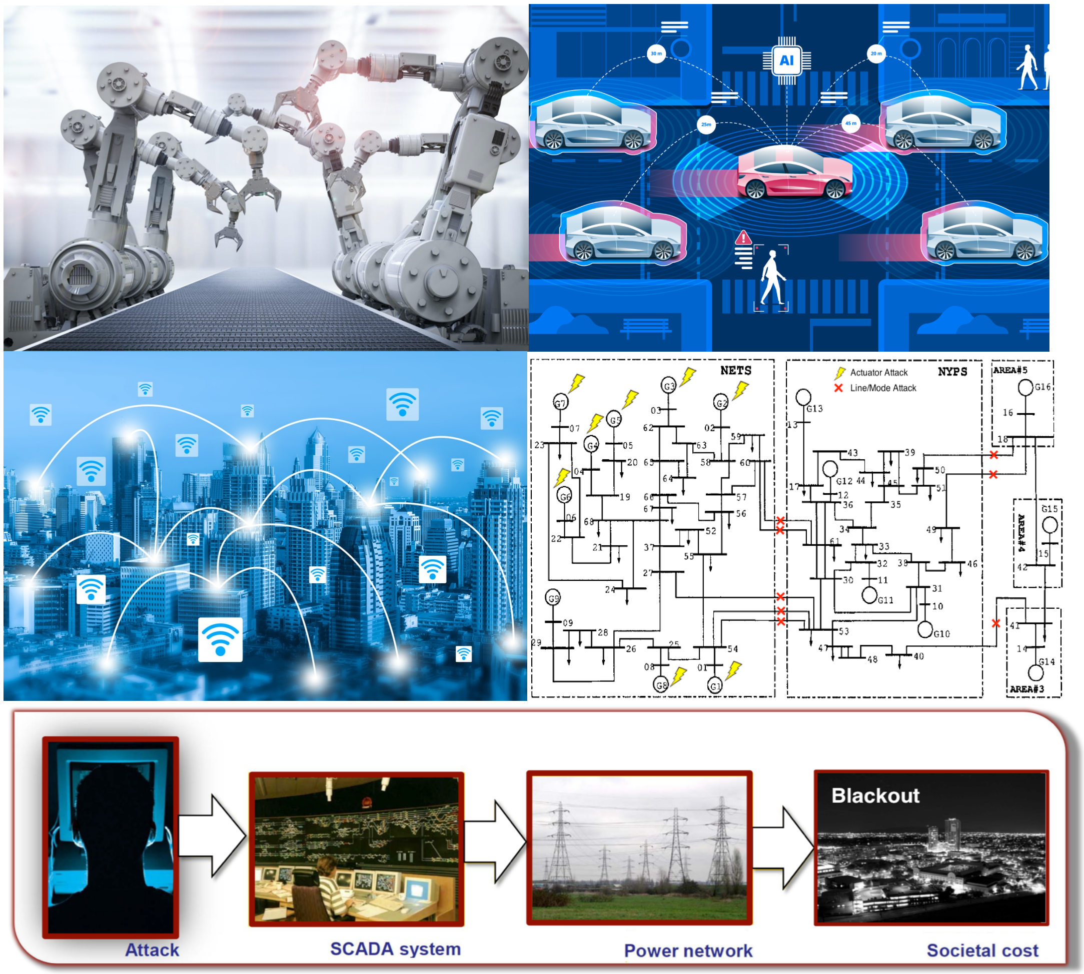


Figure 1.2: Safety-critical Cyber-physical Systems Are Susceptible to Attacks (Unknown Inputs)

nonlinear systems are relatively scarce. The method proposed in [123] leverages discrete-time sliding mode observers for calculating state and unknown input point estimates, assuming that the unknown inputs have *known* bounds and evolve as *known* functions of states, which may not be directly applicable when considering adversaries in the system. The authors in [69] proposed an LMI-based state estimation approach for globally Lipschitz nonlinear discrete-time systems, but did not consider unknown input reconstruction. An LMI-based approach was also used in [51] for simultaneous estimation of state and unknown input for a class of continuous-time dynamic systems with Lipschitz nonlinearities, but the authors did not address optimality nor stability properties for their observer, as well as only considered point estimates. The work in [4] designed an asymptotic observer to calculate point estimates for a class of continuous-time systems whose nonlinear terms satisfy an *incremental quadratic inequality* property. Similar work was done for the same class of discrete-time nonlinear systems in [23], while the set-valued state estimation approach in [99] uses mean value and first-order Taylor extensions to efficiently propagate constrained zonotopes through nonlinear mappings. However, none of them addressed unknown input estimation. Moreover, the restrictive assumption of bounded unknown inputs is needed in order to obtain convergent estimates.

Considering bounded unknown inputs, but with unknown bounds, the work in [25] applied second-order series expansions to construct observer for state estimation in nonlinear discrete-time systems. The authors also provided sufficient conditions for stability and optimality of the designed estimator. However, their method does not compute unknown input estimates. On the other hand, in a recent and interesting work in [22], the authors designed an observer for reconstruction of unknown exogenous inputs in nonlinear continuous-time systems with unknown and potentially unbounded inputs, providing sufficient LMI conditions for $\mathcal{L}_\infty$-stability of the observer. However, their observer does not simultaneously estimate the state, the unknown input estimates are point estimates and the optimality of their approach was not analyzed.

The author in [131] and references therein discussed the advantages of set-valued observers (when compared to point estimators) in terms of providing hard accuracy bounds, which are important to guarantee safety [17]. In addition, the use of *fixed-order* set-valued methods can help decrease the complexity of optimal observers [82], which grows with time. Hence, a fixed-order set-valued observer for linear time-invariant discrete time systems with bounded errors, was presented in [131], that simultaneously finds bounded sets of compatible states and unknown inputs that are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. In chapter 2 of this thesis, ([115]), we extend the approach in [131] to *linear parameter-varying* systems, while in chapters 3 and 5 ([116, 121]), we generalize the method to *switched linear and nonlinear* systems with unknown modes and sparse unknown inputs (attacks), respectively, in order to design simultaneous mode, state and input observers, and in chapter 4 ([118]), we further design novel set-valued observers for broader classes of nonlinear systems, where in all of them, the considered sets are *hyper-balls* in n-dimensional Euclidean space.

Moreover, considering interval-valued uncertainties, interval observer design has also been extensively studied in the literature [54, 66, 83, 14, 97, 96, 76, 77, 125, 41, 140]. However, relatively restrictive assumptions about the existence of certain system properties were imposed to guarantee the applicability of the proposed approaches, such as cooperativeness [97], linear time-invariant (LTI) dynamics [76], linear parameter-varying (LPV) dynamics that admits a diagonal Lyapunov function [125], monotone dynamics [83, 14], and Metzler and/or Hurwitz partial linearization of nonlinearities [96, 77]. The problem of designing an L_2/L_∞ unknown input interval observer for continuous-time LPV systems is studied in [42], where the required Metzler property is formulated as a part of a semi-definite program. However, this approach is not directly applicable for general discrete-time nonlinear systems. Moreover, in their setting, the unknown inputs do not affect the output (measurement) equation.

Leveraging *bounding functions*, the design of interval observers for a class of continuous-time nonlinear systems without unknown inputs has been addressed in [41]. However, no necessary and/or sufficient conditions for the existence of bounding functions or how to compute them have been discussed. Moreover, to conclude stability, somewhat restrictive assumptions on the nonlinear dynamics have been imposed. On the other hand, the authors in [140] studied interval state estimation for a class of uncertain nonlinear systems, by extracting a known nominal observable subsystem from the plant equations and designing the observer for the transformed system, but without providing guarantees that the derived functional bounds have finite values, i.e., are bounded sequences. Moreover, the derived conditions for the existence and stability of the observer are not *constructive*. More importantly, none of the aforementioned works consider unknown inputs (without known bounds/intervals) nor the reconstruction/estimation of the uncertain inputs. In chapter 6 of this research ([117]) , we design an observer that *simultaneously* returns interval-valued estimates of states and unknown inputs for a broad range of nonlinear systems [129], in contrast to existing interval observers in the literature that to the best of our knowledge, only return either state [54, 66, 83, 14, 97, 96, 76, 77, 125, 41, 140] or input [42] estimates.

Furthermore, dynamic models of many practical systems are often only partially known. Thus, the development of algorithms that can combine model learning and set membership estimation approaches is a critical and interesting problem. In such settings, set-valued data-driven approaches that use input-output data to *abstract* or over-approximate unknown dynamics or functions have gained increased popularity over the last few years [81, 20, 138, 13, 19], where the objective is to find *a set* *of dynamics* that frame/bracket the unknown system dynamics [81, 20], under the assumption that the unknown dynamics is univariate Lipschitz continuous [138], multivariate Lipschitz continuous [13] or Hölder continuous [19]. Nonetheless, to our knowledge, set-valued or interval observers for such data-driven models have not been considered in the literature. This, is the main focal point of chapter 7 in this document.

On the other hand, aiming to apply set-membership approaches to constrained nonlinear safety-critical systems and considering very useful and fundamental properties of monotone systems [52, 11], the significant idea of embedding the system dynamics into a higher dimensional monotone system [70, 43, 49], raised huge attention. One property that if satisfied, provides powerful means to obtain this goal is *mixed-monotonicity* (cf. Definition 6.1.7), that enables a system to admit *decomposition functions* and consequently higher dimensional dynamical *embedding systems*, which can be applied to extract properties of the original system. For instance, if mixed-monotonicity holds, by proving the nonexistence of equilibria of the embedding system except in a certain lower dimensional subspace, global asymptotic stability for the original system can be concluded [110, 29], by analyzing equilibria in the embedding space, forward invariant and attractive sets of the original system can be identified [1] and by evaluating the trajectory of the embedding system, reachable sets of the original system can be efficiently approximated, that is widely applicable e.g., in state estimation and abstraction-based control synthesis [1, 33, 32, 127]. Hence, finding a broad range of nonlinearities that satisfy mixed-monotonicity, as well as computing tractable and tight decomposition functions for such systems, along with considering uncertainties and constraints are all of great interest and critical problems.

Due to non-uniqueness of decomposition functions, several seminal studies have been done to address the critical challenge of constructing/identifying appropriate decomposition functions, when applying the theory of mixed-monotone systems, providing slightly different -but highly related- definitions and useful sufficient conditions for mixed-monotonicity [129, 128, 2, 79, 80, 32, 34, 29]. Moreover, the profound existing literature on interval arithmetic [57, 35, 84, 7, 65], equips us with the notion of inclusion functions and provides specific types of them e.g., natural, centered-form and mixed-form inclusions, which all of them, as well as their refinements [130, 5, 6, 102, 105] can be interpreted and applied as specific forms of decomposition functions. Particularly, the pioneer work in [35] studies the over-approximation of range of functions with higher than second order accuracies, using modifications of natural inclusions along with *subdivision principle*. The important study in [65] -to the best of our knowledge- establishes the applicability of natural inclusions to guaranteed state estimation. Later, several profound work have been done to propose *refined* interval arithmetic-based state bounding approaches, using new sources of information about the system such as state constraints, measurements/observations, manufactured redundant variables, second-order derivatives, etc, e.g. in [130, 5, 6, 102].

Concerning tightness of decomposition functions, the interesting and recent studies in [129, 2] provide tight decomposition functions for unconstrained mixed-monotone discrete and continuous-time dynamical systems, respectively, where the existence and computability of such tight decomposition functions rely on global solvability of nonlinear optimization programs, which is guaranteed in specific cases such as when the vector field is *Jacobian sign-stable*, or if all the *critical points* of the vector field can be precisely computed. On the other hand, the fascinating study in [128] provides sufficient conditions for mixed-monotonicity of (not necessarily Jacobian sign-stable) unconstrained discrete-time systems, along with proposing computable and constructive -but not necessarily tight- decomposition functions for differentiable and mixed-monotone vector fields, with a prior known bounds for their derivatives. In chapter 8 of this thesis, we further generalize the work in [128] to obtain computable and possibly tighter *mixed-monotone remainder-form decomposition functions* for not necessarily differentiable/smooth, constrained, continuous and discrete-time systems, affected by external and/or internal uncertainties such as bounded disturbance and/or uncertain parameters. We also study the applications of decomposition and inclusion functions to reachability analysis and state estimation.

In particular, A well-known strategy, which is common to most of the set-theoretic state estimation approaches is finding an enclosing set to the image set of the dynamics vector field, i.e., *propagation/prediction* step, as well as refining the obtained propagated set by finding an enclosure to its intersection with the set of states that are compatible/consistent with the observation/measurements, i.e., *update* step.

In case of linear systems with polytopic initial set, it is theoretically shown that tight (exact) enclosures can be obtained [48]. However, even for linear systems, the computational complexity of polytopic propagation is extensive and grows dramatically with time [104]. Hence, simpler sets such as parallelotopes [124, 27], ellipsoids [115, 93, 60], intervals [140, 61, 125, 62] or zonotopes [71, 31] have been used to characterize the enclosures. However, structural limitations of these sets sometimes leads to conservative enclosures. To address this, the work in [103] introduced *constrained zonotopes* to ease some of the limitations imposed by zonotopes, while *zonotope bundles* were proposed in [9] to describe the intersection of zonotopes without explicit computations.

Regarding nonlinear systems, obtaining efficient set-valued estimates is still very challenging, contrary to the linear case. A classical approach to tackle this problem has been to use interval arithmetic-based inclusion functions [84] to propagate the current enclosing sets through the nonlinear dynamics and then to apply interval-based set inversion techniques (e.g., SIVIA) to find upper approximations for the set of compatibles states with the current measurements [55, 56]. These approaches are computationally very efficient, but unfortunately, due to the nature of interval arithmetic, the resultant bounds are mostly conservative.

Alternatively, given linear observation functions, zonotopic propagation methods have been developed in [30, 5, 6], based on the first order Taylor expansion, the mean value extension or DC programming. However, significant errors are caused in update step due to the symmetry of zonotopes, even for linear measurements [103]. More recently, the interesting work in [99] proposed constrained zonotopic propagation and update algorithms for discrete-time nonlinear systems with linear observation functions, based on mean value and first order Taylor extensions. In chapter 9, we conclude this thesis, by addressing the problem of guaranteed state estimation of nonlinear systems in the presence of general convex set-valued, i.e., polytopic uncertainties, using mixed-monotonicity.

1.1 Contributions of the Research

In chapter 2, we propose a novel fixed-order set-valued observer for linear parameter-varying systems with unknown input and bounded noise signals that simultaneously finds bounded sets of states and unknown inputs that contain the true state and unknown input and are compatible/consistent with the measurement outputs. Specifically, we consider linear parameter-varying system dynamics that can be presented as a convex combination of linear time-invariant *constituent* dynamics. In addition, we provide necessary conditions for the boundedness of the set-valued estimates. We further prove the optimality of the filter in the minimum $\mathcal{H}_\infty$-norm sense, i.e., minimum average power amplification, by converting the corresponding problem into a tractable formulation using semi-definite programming with LMI constraints that is readily implementable using off-the-shelf optimization solvers. We also show that strong detectability of each constituent system is a necessary condition for the existence of such an $\mathcal{H}_\infty$-observer. Then, we provide some sufficient conditions for the convergence of upper bounds of the state and input estimation errors to steady state and for obtaining these steady state bounds. Finally, we demonstrate the effectiveness of our proposed set-valued observer through an illustrative example.

The goals of chapters 3 and 5 are to simultaneously consider state and unknown input estimation as well as mode detection for hidden mode switched linear and nonlinear systems with bounded-norm noise and unknown inputs, respectively. To address these, we propose a multiple-model approach that leverages the optimally designed set-valued state and input $\mathcal{H}_\infty$ observers in chapters 2 and 4 to obtain a bank of mode-matched set-valued observers in combination with a novel mode observer based on elimination. Our mode elimination approach uses the upper bound of the norm of to-be-designed residual signals to remove inconsistent modes from the bank of observers. Moreover, we provide sufficient conditions to guarantee that all false modes will be eventually eliminated.

Chapter 4 aims to bridge the gap between set-valued state estimation without unknown inputs and point-valued state and unknown input estimation for a broad range of nonlinear dynamical systems. In particular, we propose fixed-order set-valued observers for nonlinear discrete-time bounded-error systems that simultaneously find uniformly bounded sets of states and unknown inputs that contain the true state and unknown input, are compatible/consistent with measurement outputs and are optimal in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification. First, we introduce a novel class of nonlinear vector fields, Decremental Quadratic Constraint (DQC) systems, and show that they include a broad range of nonlinearities. We also derive some results on the relationship between DQC functions with some other classes of nonlinearities, such as incremental quadratic constraint, Lipschitz continuous and linear parameter-varying (LPV) functions. Then, we present our three-step recursive set-valued observer for nonlinear discrete-time systems. In particular, we derive sufficient conditions for the stability of the observer (i.e., the estimation errors are uniformly bounded) in the form of LMIs for general nonlinear systems, as well as less restrictive sufficient LMI conditions for stability of the observer for three classes of nonlinearities: (I) DQC, (II) Lipschitz continuous and (III) LPV systems. Furthermore, we design $\mathcal{H}_\infty$ observers, using additional LMIs for each of the aforementioned classes of systems. Finally, we derive sufficient conditions for convergence of the estimation errors for each class of functions.

In chapter 6, by leveraging a combination of nonlinear decomposition mappings [128, 32] and affine abstraction (bounding) functions [108], we design an observer that *simultaneously* returns interval-valued estimates of states and unknown inputs for a broad range of nonlinear systems [129], in contrast to existing interval observers in the literature that to the best of our knowledge, only return either state [54, 66, 83, 14, 97, 96, 76, 77, 125, 41, 140] or input [42] estimates. Moreover, we consider arbitrary unknown input signals with no assumptions of *a priori* known bounds/intervals, being stochastic with zero mean (as is often assumed for noise) or bounded. Further, we relax the assumption of a full-rank feedthrough matrix in [117], and extend the observer design by including a crucial *update step*, where starting from the intervals from the propagation step, the framers are iteratively updated by intersecting it with the state and input intervals that are compatible with the observations. As a result, the updated framers have decreased widths, i.e., tighter intervals can be obtained. In addition, we derive sufficient conditions for the existence of our observer that can be viewed as structural properties of the nonlinear systems, as an extension of the rank condition that is typically assumed in linear state and input estimation, e.g., [72, 136, 131]. We also provide several sufficient conditions in the form of Linear Matrix Inequalities

(LMI) for the stability of our designed observer (i.e., the uniform boundedness of the sequence of estimate interval widths). In addition, we show that given the state intervals and specific decomposition functions, our input interval estimates are *tight* and further provide upper bound sequences for the interval widths and derive sufficient conditions for their convergence and their corresponding steady-state values.

Chapter 7 bridges the gap between model-based set membership observer design approaches, e.g., [54, 66, 83, 14, 97, 96, 76, 77, 125, 41, 140, 78, 42, 131, 115, 116, 118], and data-driven function approximation methods (i.e., model learning methods), e.g., [81, 20, 138, 13, 19], to design interval observers for partially known nonlinear dynamical systems with bounded noise, where the state and observation vector fields belong to a fairly general class of nonlinear functions and the vector field of the unknown (input) dynamics is an *unknown function*. Our approach builds upon and extends the observer design approach in [117] by including a crucial *update step*, where starting from the intervals from the propagation step, the framers are iteratively updated by computing their intersection with the augmented state intervals that are compatible with the observations, resulting in tighter intervals (i.e., with decreased interval width) for the updated framers. In addition, our design incorporates a data-driven function approximation/abstraction approach based on [59] to recursively over-approximate the unknown dynamincs function from noisy observation data and interval estimates from the update step. Furthermore, by leveraging the combination of nonlinear decomposition/bounding functions [128, 32, 129, 117] and affine abstractions [108], we prove that our observer is correct, i.e., the framer property [77] holds and our estimation/abstraction of the unknown dynamics model becomes more precise and tighter over time. More importantly, we provide sufficient conditions, in the form of a finite number of constraint satisfaction checks, for the stability of our observer (i.e., for the uniform boundedness of the sequence of interval estimate widths), and compute the upper bounds for the interval widths of the sequence of estimates and derive their steady-state values.

Chapter 8 provides sufficient conditions for mixed-monotonicity of a broad range of nonlinear, bounded-error, constrained, discrete or continuous-time dynamical systems. The range of systems that we consider is broader compared to the ones considered in the literature in certain directions. Particularly, we relax the smoothness (differentiability) and bounded gradients requirements to one-sided boundedness of *generalized Clarke sub-differentials*, that basically holds for every locally Lipschitz vector field, we consider discrete-time as well as continuous-time systems together, the system can include bounded-error internal or external uncertainties in the form of uncertain parameters and/or process/measurement noise/disturbance and can be constrained under any locally Lipschitz nonlinear mapping between states, inputs and outputs/observations. Our sufficient conditions are constructive, i.e. we propose tight and computable remainder-form decomposition functions for such systems. The proposed decomposition functions are proven to be the best/tightest among the family of remainder-form decomposition functions that we construct. We show that the introduced decomposition function in [128] belongs to this family, but is not necessarily the tightest of them. Moreover, we obtain upper and lower bounds for the errors of approximation of range of a function using our proposed family of decomposition functions, showing that the best of them minimizes the lower bound, while the one given in [128], minimizes an upper bound of the errors. Further, we show that the error of the approximation could decay exponentially fast, using *subdivision principle*. We also slightly generalize the notion of decomposition functions to one-sided upper and lower decomposition functions. Furthermore, to deal with constraints, we develop a set-inversion algorithm applying our decomposition functions, where given the propagated interval of states, the constraint/measurement mapping and the measurement interval, the proposed algorithm returns the refined/updated interval of the states which is compatible/consistent with the measurements/observations. We then show that our decomposition functions along with the proposed set-inversion algorithm is applicable to solve constrained reachability as well as state estimation problems and hence is capable to improve further some of our existing results in state and input estimation [115, 116, 117, 113]. Moreover, the proposed set-inversion algorithm can be used with *any* applicable inclusion functions or the best of them, replacing our proposed decomposition functions. Finally, we illustrate the effectiveness of the proposed decomposition functions and set-inversion approach, using several examples, including discrete and continuous-time, constrained and unconstrained systems, where we compare our approach with multiple other inclusion/decomposition functions.

Finally, chapter 9 proposes novel methods for recursive state estimation (consisting of propagation and update steps) using polytopes (equivalently, constrained zonotopes or zonotope bundles) for nonlinear bounded-error discrete-time systems with nonlinear observation functions. Leveraging remainder-form mixed-monotone decomposition functions [63] and following the standard propagation and update approach, this chapter bridges the gap between constrained zonotope (CZ)/zonotope bundle (ZB)-based set-valued state estimation and nonlinear observation/constrained functions. In particular, for the propagation step, we transform the prior ZB/CZ's into the space of CZ/ZB generators, which are interval-valued, and further transform the vector field into two components, one that is proven to attain tight image sets, as well as a linear remainder function, for which a family of remainder-form mixed-monotone decomposition functions [63] can be obtained. Each of the decomposition functions produce enclosures of the state trajectories and thus, we can intersect them to obtain the desired propagated ZB/CZ enclosures.

Moreover, we show that a similar idea, i.e., transformation from the "state +

uncertainty" space to the space of generators of CZ/ZB's, can be used for the update step to find a family of enclosures to the *generalized nonlinear intersection* of the propagated set with the set of states that is compatible with the observations, where the final enclosures are proven to be ZB/CZ's. Furthermore, we prove that the mean value extension approach used in [99] to enclose a multiplication of an interval matrix to a constrained zonotope, can also be leveraged for the update step when the observation function is nonlinear. Finally, we compare our proposed approaches together and with the mean value extension-based approach in [99], implementing it on two examples, one with a linear and the other with a nonlinear observation function.

1.2 Published/Under Review/ In Preparation Content

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous input and state set-valued $\mathcal{H}_\infty$-observers for linear parameter-varying systems*". In American Control Conference (ACC), pages 4521–4526, 2019.

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous mode, input and state set-valued observers with applications to resilient estimation against sparse attacks*". In 2019 IEEE 58th Conference on Decision and Control (CDC), pages 1544–1550.

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous input and state interval observers for nonlinear systems with full-rank direct feedthrough*". In 2020 IEEE 59th Conference on Decision and Control (CDC), pages 5443–5448.

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous state and unknown input set-valued observers for nonlinear dynamical systems*". arXiv preprint arXiv:2001.10125, Submitted to Automatica, under review, 2020.

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous Mode, State and Input Set-Valued Observers for Switched Nonlinear Systems*". https://arxiv.org/pdf/2102.10793.pdf. Submitted to IFAC Journal of Nonlinear Analysis: Hybrid Systems, under review, 2021.

- **M. Khajenejad** and S.Z. Yong. "*Resilient State Estimation and Attack Mitigation in Cyber-Physical Systems*". Accepted for publication as a book chapter by Springer, 2021.

- **M. Khajenejad**, Z. Jin, and S.Z. Yong. "*Interval observers for simultaneous state and model estimation of partially known nonlinear systems*". In American Control Conference (ACC), pages 2848–2854, 2021

- **M. Khajenejad** and S.Z. Yong. "*Tight remainder-form decomposition functions with applications to constrained reachability and interval observer design*". https://arxiv.org/abs/2103.08638, Submitted to IEEE Transaction on Automatic Control, under review, 2021.

- **M. Khajenejad** and S.Z. Yong. "*Simultaneous input and state interval observers for nonlinear systems with rank-deficient direct feedthrough*". In European Control Conference (ECC), 2021, accepted.

- **M. Khajenejad**, Z. Jin, and S.Z. Yong. "*Set-Valued State and Unknown Terrain Estimation for Planetary Rovers*". Accepted for publication in the journal of Advanced Intelligent Systems, 2021.

- **M. Khajenejad**, F. Shoaib and S.Z. Yong. "*Guaranteed State Estimation via Remainder-Form Decomposition Function-Based Set Inclusion for Nonlinear Discrete-Time Systems*". In 60th IEEE Conference on Decision and Control (CDC), 2021 (accepted).

- **M. Khajenejad**, F. Shoaib and S.Z. Yong. "*Guaranteed State Estimation via Directly Implemented Polytopic Set Computation for Nonlinear Discrete-Time Systems*". In preparation.

1.3 Other Research Contributions

- Z. Jin, **M. Khajenejad**, and S.Z. Yong. "Data-driven model invalidation for unknown Lipschitz continuous systems via abstraction". In American Control Conference (ACC), pages 2975–2980. IEEE, 2020.
- S. M. Hassaan, **M. Khajenejad**, S. Jensen, Q. Shen and S. Z. Yong, "Incremental Affine Abstraction of Nonlinear Systems," in IEEE Control Systems Letters, vol. 5, no. 2, pp. 653–658, April 2021, doi: 10.1109/LCSYS.2020.3004503.
- **M. Khajenejad**, M.Cavorsi, R.Niu, Q.Shen and S.Z. Yong. "Tractable compositions of discrete-time control barrier functions with application to automotive safety control". In European Control Conference (ECC) 2021, accepted.
- **M. Khajenejad**, F. Afshinmanesh, A Marandi and BN Araabi. "*Intelligent Particle Swarm Optimization Using Q-Learning*". In Proceeding of IEEE Swarm Intelligence, pages 7–12, 2006.

1.4 Notation

$\mathbb{N}$, $\mathbb{N}_a$, $\mathbb{R}_{++}$, $\mathbb{R}^{n_z}$, $\mathbb{R}^{n\times m}$ and $\mathbb{D}^{n\times n} \subset \mathbb{R}^{n\times n}$ denote the set of positive integers, the first a positive integers, the set of positive real numbers, the n_z-dimensional Euclidean space, the space of n by m real matrices and the space of square diagonal matrices, respectively. Moreover, $\forall z, \underline{z}, \overline{z} \in \mathbb{R}^{n_z}$, $\underline{z} \le \overline{z} \Leftrightarrow \underline{z}_i \le \overline{z}_i, \forall i \in \{1,\dots,n_z\}$, where $\underline{z}_i$ denotes the i'th argument of the vector $\underline{z}$. Further, $\mathcal{Z} = [\underline{z}, \overline{z}] \triangleq [z \in \mathbb{R}^{n_z} | \underline{z} \le z \le \overline{z}]$ and $\|\overline{z} - \underline{z}\|_\infty$ are called an interval/box in $\mathbb{R}^{n_z}$ and the diameter of $\mathcal{Z}$, accordingly, where $\|z\|_\infty \triangleq \max_i |z_i|$ denotes the infinity norm of $z \in \mathbb{R}^{n_z}$. The set of all intervals in $\mathbb{R}^{n_z}$ is denoted by $\mathbb{IR}^{n_z}$.

Further, for vectors $v, w \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{p\times q}$, $\|v\| \triangleq \sqrt{v^\top v}$ and $\|M\|$ denote their (induced) 2-norm, and $v \leq w$ is an element-wise inequality. Moreover, the transpose, Moore-Penrose pseudoinverse, (i,j)-th element and rank of M are given by $M^\top$, $M^\dagger$, $M_{i,j}$ and $\mathrm{rk}(M)$. $M_{(r:s)}$ is a sub-matrix of M, consisting of its r-th through s-th rows, and we call M a non-negative matrix, i.e., $M \geq 0$, if $M_{i,j} \geq 0, \forall i \in \{1 \dots p\}, \forall j \in \{1 \dots q\}$. We also define $M^+, M^{++} \in \mathbb{R}^{p\times q}$ as $M^+_{i,j} = M_{i,j}$ if $M_{i,j} \geq 0$, $M^+_{i,j} = 0$ if $M_{i,j} < 0$, $M^{++} = M^+ - M$ and $|M| \triangleq M^+ + M^{++}$. Furthermore, $r = \mathit{rowsupp}(M) \in \mathbb{R}^p$, where $r(i) = 0$ if the i-th row of A is zero and $r(i) = 1$ otherwise, $\forall i \in \{1 \dots p\}$. For a symmetric matrix S, $S \succ 0$ and $S \prec 0$ ($S \succeq 0$ and $S \preceq 0$) are positive and negative (semi-)definite, respectively.

Moreover, for $\mathcal{Z}, \mathcal{W} \subset \mathbb{R}^n, R \in \mathbb{R}^{m\times n}, \mathcal{Y} \subset \mathbb{R}^m$, and $\mu : \mathbb{R}^n \to \mathbb{R}^m$, $R\mathcal{Z} \triangleq \{Rz | z \in \mathcal{Z}\}, \mathcal{Z} \oplus \mathcal{W} \triangleq \{z + w | z \in \mathcal{Z}, w \in \mathcal{W}\}$, $\mathcal{Z} \ominus \mathcal{W} \triangleq \{z - w | z \in \mathcal{Z}, w \in \mathcal{W}\}$, $\mu(\mathcal{Z}) \triangleq \{\mu(z) | z \in \mathcal{Z}\}$ and $\mathcal{Z} \cup_\mu \mathcal{Y} \triangleq \{z \in \mathcal{Z} | \mu(z) \in \mathcal{Y}\}$ denote the linear mapping, Minkowski sum, set subtraction, general (nonlinear) mapping and generalized (nonlinear) intersection, respectively. Furthermore, $\mathbb{B}^n_\infty \triangleq \{z \in \mathbb{R}^n | \|z\|_\infty \leq 1\}$ and $\mathbf{0}_n$ denote the ∞-norm hyperball and the zero vector in $\mathbb{R}^n$, respectively. For $z \in \mathbb{R}^n$, $\mathrm{diag}(z)$ is a diagonal matrix in $\mathbb{R}^{n\times n}$, with its diagonal elements being the corresponding elements of z. $\langle \cdot, \cdot \rangle$ denotes the inner product operator.

Chapter 2

SIMULTANEOUS INPUT AND STATE SET-VALUED $\mathcal{H}_\infty$-OBSERVERS FOR LINEAR PARAMETER-VARYING SYSTEMS

In this chapter [a] , we propose a novel fixed-order set-valued observer for linear parameter-varying systems with unknown input and bounded noise signals that simultaneously finds bounded sets of states and unknown inputs that contain the true state and unknown input and are compatible/consistent with the measurement outputs. Specifically, we consider linear parameter-varying system dynamics that can be presented as a convex combination of linear time-invariant *constituent* dynamics. In addition, we provide necessary conditions for the boundedness of the set-valued estimates. We further prove the optimality of the filter in the minimum $\mathcal{H}_\infty$-norm sense, i.e., minimum average power amplification, by converting the corresponding problem into a tractable formulation using semi-definite programming with LMI constraints that is readily implementable using off-the-shelf optimization solvers. We also show that strong detectability of each constituent system is a necessary condition for the existence of such an $\mathcal{H}_\infty$-observer. Then, we provide some sufficient conditions for the convergence of upper bounds of the state and input estimation errors to steady state and for obtaining these steady state bounds. Finally, we demonstrate the effectiveness of our proposed set-valued observer through an illustrative example.

---

[a] The content of this chapter is documented as a published paper in [115].

2.1 Problem Statement

System Assumptions. Consider the following linear parameter-varying discrete-time bounded-error system:

$$\begin{aligned} x_{k+1} &= \textstyle\sum_{i=1}^{N} \lambda_{i,k}(A^i x_k + B^i u_k + w_k^i) + G d_k, \\ y_k &= C x_k + \textstyle\sum_{i=1}^{N} \lambda_{i,k}(D^i u_k + v_k^i) + H d_k, \end{aligned} \tag{2.1}$$

where $\lambda_{i,k}$ is known and satisfies $0 \le \lambda_{i,k} \le 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$. $x_k \in \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$, $u_k \in \mathbb{R}^m$ is a known input vector, $d_k \in \mathbb{R}^p$ is an unknown input vector, and $y_k \in \mathbb{R}^l$ is the measurement vector. The process noise $w_k^i \in \mathbb{R}^n$ and the measurement noise $v_k^i \in \mathbb{R}^l$ are assumed to be bounded and ℓ_∞ sequences, with $\|w_k^i\| \le \eta_w$ and $\|v_k^i\| \le \eta_v$.We also assume an estimate $\hat{x}_0$ of the initial state x_0 is available, where $\|\hat{x}_0 - x_0\| \le \delta_0^x$. The matrices A^i, B^i, C, D^i, G and H are known for $i \in \{1, 2, \ldots, N\}$ and of appropriate dimensions, where G and H are matrices that encode the *locations* through which the unknown input or attack signal can affect the system dynamics and measurements and N is the number of *constituent* systems. Note that no assumption is made on H to be either the zero matrix (no direct feedthrough), or to have full column rank when there is direct feedthrough. Without loss of generality, we assume that $\text{rk}[G^\top \ H^\top] = p$, $n \ge l \ge 1$, $l \ge p \ge 0$, $m \ge 0$ and each $(A^i, B^i, C, D^i, G, H), i \in \{1, 2, \ldots, N\}$ represents a linear time-invariant constituent system:

$$\begin{aligned} x_{k+1}^i &= A^i x_k^i + B^i u_k + G d_k + w_k^i, \\ y_k^i &= C x_k + D^i u_k + H d_k + v_k^i. \end{aligned} \tag{2.2}$$

Unknown Input (or Attack) Signal Assumptions. The unknown inputs d_k are not constrained to be a signal of any type (random or strategic) nor to follow any model, thus no prior 'useful' knowledge of the dynamics of d_k is available (independent of $\{d_\ell\}$ $\forall k \neq \ell$, $\{w_\ell\}$ and $\{v_\ell\}$ $\forall \ell$). We also do not assume that d_k is bounded or has known bounds and thus, d_k is suitable for representing adversarial attack signals.

The simultaneous input and state set-valued observer design problem can be stated as follows:

Problem 2.1.1. *Given a linear parameter-varying discrete-time bounded-error system with unknown inputs* (2.1) *, design an optimal and stable filter that simultaneously finds bounded sets of compatible states and unknown inputs in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification.*

2.2 Preliminary Material

2.2.1 System Transformation

In order to decouple the output equation into two components, first a transformation is carried out for each of the constituent subsystems, one with a full rank direct feedthrough matrix and the other without direct feedthrough. Note that this similarity transformation is similar to the one in [131] and is not the same as the one in [135], which is no longer applicable as it was based on the noise error covariance.

Let $p_H \triangleq \text{rk}(H)$. Using singular value decomposition, we rewrite the direct feedthrough matrix H as $H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}$, where $\Sigma \in \mathbb{R}^{p_H \times p_H}$ is a diagonal matrix of full rank, $U_1 \in \mathbb{R}^{l \times p_H}$, $U_2 \in \mathbb{R}^{l \times (l-p_H)}$, $V_1 \in \mathbb{R}^{p \times p_H}$ and $V_2 \in \mathbb{R}^{p \times (p-p_H)}$, while $U \triangleq \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V \triangleq \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ are unitary matrices. When there is no direct feedthrough, Σ, U_1 and V_1 are empty matrices [b] , and U_2 and V_2 are arbitrary unitary matrices.

---

[b] Based on the convention that the inverse of an empty matrix is an empty matrix and the assumption that operations with empty matrices are possible.

Then, we decouple the unknown input into two orthogonal components:

$$d_{1,k} = V_1^\top d_k, \quad d_{2,k} = V_2^\top d_k. \tag{2.3}$$

Considering that V is unitary, we obtain

$$d_k = V_1 d_{1,k} + V_2 d_{2,k}, \tag{2.4}$$

and hence, we can represent the system (2.1) as:

$$\begin{aligned} x_{k+1} &= \textstyle\sum_{i=1}^N \lambda_{i,k}(A^i x_k + B^i u_k + w_k^i) + G_1 d_{1,k} + G_2 d_{2,k}, \\ y_k &= C x_k + \textstyle\sum_{i=1}^N \lambda_{i,k}(D^i u_k + v_k^i) + H_1 d_{1,k} \end{aligned} \tag{2.5}$$

where $G_1 \triangleq GV_1$, $G_2 \triangleq GV_2$ and $H_1 \triangleq HV_1 = U_1 \Sigma$. Next, the output y_k is decoupled using a nonsingular transformation $T = \begin{bmatrix} T_1^\top & T_2^\top \end{bmatrix}^\top \triangleq U^\top = \begin{bmatrix} U_1 & U_2 \end{bmatrix}^\top$ to get $z_{1,k} \in \mathbb{R}^{p_H}$ and $z_{2,k} \in \mathbb{R}^{l-p_H}$ given by

$$\begin{aligned} z_{1,k} &\triangleq T_1 y_k = U_1^\top y_k \\ &= C_1 x_k + \Sigma d_{1,k} + \textstyle\sum_{i=1}^N \lambda_{i,k} D_1^i u_k + \sum_{i=1}^N \lambda_{i,k} v_{1,k}^i \\ z_{2,k} &\triangleq T_2 y_k = U_2^\top y_k \\ &= C_2 x_k + \textstyle\sum_{i=1}^N \lambda_{i,k} D_2^i u_k + \sum_{i=1}^N \lambda_{i,k} v_{2,k}^i \end{aligned} \tag{2.6}$$

where $C_1 \triangleq U_1^\top C$, $C_2 \triangleq U_2^\top C$, $D_1^i \triangleq U_1^\top D^i$, $D_2^i \triangleq U_2^\top D^i$, $v_{1,k}^i \triangleq U_1^\top v_k^i$ and $v_{2,k}^i \triangleq U_2^\top v_k^i$. This transform is also chosen such that $\| \begin{bmatrix} {v_{1,k}^i}^\top & {v_{2,k}^i}^\top \end{bmatrix}^\top \| = \|U^\top v_k^i\| = \|v_k^i\|$.

2.3 Fixed-Order Simultaneous Input and State Set-Valued Observers

2.3.1 Set-Valued Observer Design

We consider a recursive three-step set-valued observer design. This design utilizes a similar framework as in [131] and contains an *unknown input estimation* step that uses the current measurement and the set of compatible states to estimate the set of compatible unknown inputs, a *time update* step which propagates the compatible set of states based on the system dynamics, and a *measurement update* step that uses the current measurement to update the set of compatible states. To sum up, our target is to design a three-step recursive set-valued observer of the form:

$$\begin{aligned}\textit{Unknown Input Estimation: } \hat{D}_{k-1} &= \mathcal{F}_d(\hat{X}_{k-1}, u_k),\\ \textit{Time Update:} \quad \hat{X}_k^\star &= \mathcal{F}_x^\star(\hat{X}_{k-1}, \hat{D}_{k-1}, u_k),\\ \textit{Measurement Update:} \quad \hat{X}_k &= \mathcal{F}_x(\hat{X}_k^\star, u_k, y_k),\end{aligned}$$

where $\mathcal{F}_d$, $\mathcal{F}_x^\star$ and $\mathcal{F}_x$ are to-be-designed set mappings, while $\hat{D}_{k-1}$, $\hat{X}_k^\star$ and $\hat{X}_k$ are the sets of compatible unknown inputs at time $k-1$, propagated, and updated states at time k, correspondingly. It is important to note that $d_{2,k}$ cannot be estimated from y_k since it does not affect $z_{1,k}$ and $z_{2,k}$. Thus, the only estimate we can obtain in light of (2.6) is a (one-step) delayed estimate of $\hat{D}_{k-1}$. The reader may refer to a previous work [133] for a complete discussion on when a delay is absent or when we can expect further delays. Similar to [26],[17],[131], as the complexity of optimal observers increases with time, only the fixed-order recursive filters will be considered. In particular, we choose set-valued estimates of the form:

$$\begin{aligned}\hat{D}_{k-1} &= \{d \in \mathbb{R}^p : \|d_{k-1} - \hat{d}_{k-1}\| \le \delta^d_{k-1}\},\\ \hat{X}_k^\star &= \{x \in \mathbb{R}^n : \|x_k - \hat{x}^\star_{k|k}\| \le \delta^{x,\star}_k\},\\ \hat{X}_k &= \{x \in \mathbb{R}^n : \|x_k - \hat{x}_{k|k}\| \le \delta^x_k\}.\end{aligned}$$

In other words, we restrict the estimation errors to balls of norm δ. In this setting, the observer design problem is equivalent to finding the centroids $\hat{d}_{k-1}$, $\hat{x}^\star_{k|k}$ and $\hat{x}_{k|k}$ as well as the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k of the sets $\hat{D}_{k-1}$, $\hat{X}_k^\star$ and $\hat{X}_k$, respectively. In addition, we limit our attention to observers for the centroids $\hat{d}_{k-1}$, $\hat{x}^\star_{k|k}$ and $\hat{x}_{k|k}$ that belong to the class of three-step recursive filters given in [47] and [135], defined as follows for each time k (with $\hat{x}_{0|0} = \hat{x}_0$):

Unknown Input Estimation:

$$\hat{d}_{1,k} = M_1(z_{1,k} - C_1\hat{x}_{k|k} - \textstyle\sum_{i=1}^{N} \lambda_{i,k} D_1^i u_k), \tag{2.7}$$

$$\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2\hat{x}_{k|k-1} - \textstyle\sum_{i=1}^{N} \lambda_{i,k} D_2^i u_k), \tag{2.8}$$

$$\hat{d}_{k-1} = V_1\hat{d}_{1,k-1} + V_2\hat{d}_{2,k-1}. \tag{2.9}$$

Time Update:

$$\hat{x}_{k|k-1}=\textstyle\sum_{i=1}^{N}\lambda_{i,k-1}(A^i\hat{x}_{k-1|k-1}+B^i u_{k-1})+G_1\hat{d}_{1,k-1}, \tag{2.10}$$

$$\hat{x}^\star_{k|k} = \hat{x}_{k|k-1} + G_2\hat{d}_{2,k-1}. \tag{2.11}$$

Measurement Update:

$$\begin{aligned}\hat{x}_{k|k} &= \hat{x}^\star_{k|k} + L(y_k - C\hat{x}^\star_{k|k} - \textstyle\sum_{i=1}^{N} \lambda_{i,k} D^i u_k) \\ &= \hat{x}^\star_{k|k} + \tilde{L}(z_{2,k} - C_2\hat{x}^\star_{k|k} - \textstyle\sum_{i=1}^{N} \lambda_{i,k} D_2^i u_k),\end{aligned} \tag{2.12}$$

where $L \in \mathbb{R}^{n\times l}$, $\tilde{L} \triangleq LU_2 \in \mathbb{R}^{n\times(l-p_H)}$, $M_1 \in \mathbb{R}^{p_H\times p_H}$ and $M_2 \in \mathbb{R}^{(p-p_H)\times(l-p_H)}$ are observer gain matrices that are designed according to Theorem 2.3.3. The main result in Theorem 2.3.3 is derived by minimizing the "volume" of the set of compatible states and unknown inputs, quantified by the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k. Note also that we applied $L = LU_2U_2^\top = \tilde{L}U_2^\top$ from Lemma 2.3.1 into (2.12). The state and input estimation errors are defined as $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}, \tilde{d}_{k-1} \triangleq d_{k-1} - \hat{d}_{k-1}, \tilde{d}_{1,k-1} \triangleq d_{1,k-1} - \hat{d}_{1,k-1}, \tilde{d}_{2,k-1} \triangleq d_{2,k-1} - \hat{d}_{2,k-1}$ respectively. In Lemmas 2.3.1 and 2.3.2, we will provide necessary conditions for boundedness of estimation errors and sufficient conditions for stability of the observer. All the proofs are provided in the Appendix.

Lemma 2.3.1 (Necessary Conditions for Boundedness of Set-Valued Estimates [131, Lemma 1])**.** *The input and state estimation errors, ($\tilde{d}_{k-1}$ and $\tilde{x}_{k|k}$), are bounded for all k (i.e., the set-valued estimates are bounded with radii $\delta^d_{k-1}, \delta^{x,\star}_k, \delta^x_k < \infty$), only if $M_1\Sigma = I$, $p \le l$, $M_2C_2G_2 = I$ and $LU_1 = 0$. Consequently,* $\mathrm{rk}(C_2G_2) = p - p_H$*,* $M_1 = \Sigma^{-1}$*,* $M_2 = (C_2G_2)^\dagger$ *and* $L = LU_2U_2^\top = \tilde{L}U_2^\top$*.*

Lemma 2.3.2 (Sufficient Conditions for Observer Stability)**.** *A sufficient condition for the stability of the set-valued observer is that* $(\overline{A}_k, C_2)$ *is uniformly detectable* [c] *for each* k*, where* $\overline{A}_k \triangleq (I - G_2M_2C_2)\hat{A}_k$ *and* $\hat{A}_k \triangleq \sum_{i=1}^{N} \lambda_{i,k}A^i - G_1M_1C_1$.

2.3.2 Optimal $\mathcal{H}_\infty$-Observer

In this section, we provide sufficient conditions for the *existence* of a set-valued observer for system (2.1) with any sequence $\{\lambda_{i,k}\}_{k=0}^{\infty}$ for all $i \in \{1, 2, \dots, N\}$ that satisfies $0 \leq \lambda_{i,k} \leq 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$ in the sense of $\mathcal{H}_\infty$ (i.e., minimizing the sum of squares of the state estimation error sequence). Furthermore, we introduce a relatively simple approach to find such an observer, which involves solving a semi-definite program with Linear Matrix Inequalities (LMI) as constraints. We will also show that given some structural conditions for the system, the upper bounds of the estimation errors for both states and unknown inputs are guaranteed to converge to steady state.

Theorem 2.3.3 ($\mathcal{H}_\infty$-Observer Design)**.** *Suppose Lemma 2.3.1 holds and there exist matrices* Y *and* $S \succ 0$ *with appropriate dimensions such that*

$$\begin{bmatrix} S & (\overline{A^i})^\top (S - C_2^\top Y^\top) & 0 & I \\ * & S & \begin{bmatrix} S - YC_2 & -Y \end{bmatrix} & 0 \\ * & * & \eta I & 0 \\ * & * & * & \eta I \end{bmatrix} \succ 0$$

for all $i \in \{1, 2, \dots, N\}$. *Then, there exists an* η *performance bounded* $\mathcal{H}_\infty$*-observer for system* (2.1) *with any sequence* $\{\lambda_{i,k}\}_{k=0}^{\infty}$ *for all* $i \in \{1, 2, \dots, N\}$ *that satisfies* $0 \leq \lambda_{i,k} \leq 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$ *when using* $\tilde{L} = S^{-1}Y$*, i.e.,* $\|T_{\tilde{x},w,v}\| \leq \eta^2$*, where* $T_{\tilde{x},w,v}$ *is the transfer function matrix that maps the noise signals* $\sum_{i=1}^{N} \lambda_{i,k} \begin{bmatrix} w_k^{i\top} & v_k^{i\top} \end{bmatrix}^T$ *to the updated state estimation error* $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$.

---

[c] For conciseness, the readers are referred to [10, Section 2] for the definition of uniform detectability. A spectral test can be found in [92].

Furthermore, the optimal filter gain $\tilde{L} = S^{\star -1}\tilde{Y}^{\star}$ *with* $\eta^{\star}$ $\mathcal{H}_{\infty}$*-performance can be obtained from the following semi-definite programming with LMI constraints:*

$$\begin{aligned}
(\eta^{\star}, S^{\star}, Y^{\star}) \in \arg\min_{\eta,S,Y} \quad & \eta \\
s.t \quad & \begin{bmatrix} S & (\overline{A^i})^\top (S - C_2^\top Y^\top) & 0 & I \\ * & S & \begin{bmatrix} S - YC_2 & -Y \end{bmatrix} & 0 \\ * & * & \eta I & 0 \\ * & * & * & \eta I \end{bmatrix} \succ 0, \\
& \forall i \in \{1, 2, .., N\}. \qquad (2.13)
\end{aligned}$$

Although Theorem 2.3.3 equips us with an approach for designing an $\mathcal{H}_{\infty}$-observer for the linear parameter-varying system in (2.1) when one exists, it would still be valuable to find a *structural* and conveniently testable property for the constituent linear time-invariant systems in (2.2) that is *necessary* for the existence of such an observer. Knowing such conditions would be beneficial in the sense that if they are *not* satisfied, the designer knows *a priori* that there does not exist any $\mathcal{H}_{\infty}$-observer for such an attacked system. This will be the goal of Theorem 2.3.4.

Theorem 2.3.4 (Necessary Conditions for the Existence of an $\mathcal{H}_{\infty}$-observer)**.** *There exists a simultaneous state and unknown input* $\mathcal{H}_{\infty}$*-observer for system* (2.1) *with any sequence* $\{\lambda_{i,k}\}_{k=0}^{\infty}$ *for all* $i \in \{1, 2, \ldots, N\}$ *that satisfies* $0 \le \lambda_{i,k} \le 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$, *only if each* (A^i, G, C, H) *is strongly detectable* [d] *for all* $i \in \{1, 2, \ldots, N\}$.

Next, we characterize the resulting radii δ_k^x and δ_{k-1}^d when using the proposed $\mathcal{H}_{\infty}$-observer.

Theorem 2.3.5 (Radii of Set-Valued Estimates)**.** *The radii* δ_k^x *and* δ_{k-1}^d *can be*

---

[d] For brevity, the readers may refer to [131] for the definition of strong detectability.

obtained as:

$$\delta^x_k = \delta^x_0 \theta^k + \overline{\eta} \textstyle\sum_{i=1}^{k} \theta^{i-1},$$

$$\delta^d_{k-1} = \beta \delta^x_{k-1} + \|V_2 M_2 C_2\| \eta_w + \big[\|(V_2 M_2 C_2 G_1 - V_1) M_1 T_1\| + \|V_2 M_2 T_2\|\big] \eta_v,$$

where

$$\beta \triangleq \max\nolimits_{i \in \{1,2,\ldots,N\}} \|V_1 M_1 C_1 + V_2 M_2 C_2 A_{e,i}\|,$$

$$\Psi \triangleq I - \tilde{L} C_2, \qquad \Phi \triangleq I - G_2 M_2 C_2,$$

$$A_{e,i} \triangleq \Psi \Phi (A^i - G_1 M_1 C_1), \qquad \theta \triangleq \max_{i \in \{1,2,\ldots,N\}} \|A_{e,i}\|.$$

The resulting fixed-order set-valued observer is summarized in Algorithm 1.

So far, we have designed an $\mathcal{H}_\infty$-observer for our linear parameter-varying system and provided necessary conditions for the boundedness of the set-valued estimates. It is worth mentioning that for the linear time-invariant case in [131], strong detectability of the system is also a sufficient condition for the convergence of the radii δ^x_k and δ^d_{k-1} to steady state. In our parameter-varying case, even if all constituent linear time-invariant systems are strongly detectable, there is no guarantee that the radii converge. The reason is that the convergence hinges on the stability of the product of *time-varying* matrices (cf. proof of Theorem 2.3.6), which is not guaranteed even if all the multiplicands are stable. In the next theorem, we discuss some sufficient conditions for the convergence of the radii to steady state.

Theorem 2.3.6 (Convergence)**.** *Suppose the conditions of Theorem 2.3.3 hold. Then, the radii δ^x_k and δ^d_{k-1} are convergent if $\|A_{e,i}\| < 1$ for all $i \in \{1, 2, \ldots, N\}$, where $A_{e,i}$ is defined in Theorem* (2.3.5)*. Moreover, the steady state radii is given by:*

$$\lim_{k\to\infty} \delta^x_k = \frac{\overline{\eta}}{1-\theta},$$
$$\lim_{k\to\infty} \delta^d_k = \frac{\overline{\eta}\beta}{1-\theta} + \eta_w \|V_2 M_2 C_2\| + \eta_v (\|V_2 M_2 T_2\| + \|R\|),$$

where $\overline{\eta} \triangleq (\|\Gamma\|\eta_v + \|\Psi\Phi\|\eta_w), R \triangleq V_2M_2C_2G_1M_1T_1 - V_1M_1T_1, \Gamma \triangleq -(\Psi\Phi G_1M_1T_1 + \Psi G_2M_2T_2 + \tilde{L}T_2)$.

Remark 2.3.7. *Alternatively, we can trade off between "optimality" of the observer and "convergence" of the radii. We can iteratively find* η *(e.g., by line search) that satisfies the following feasibility problem:*

$$\begin{aligned} &\text{Find } (S, Y) \\ &s.t \begin{bmatrix} S & * & 0 & I \\ (S - YC_2)\overline{A}^i & S & \begin{bmatrix} S - YC_2 & -Y \end{bmatrix} & 0 \\ * & * & \eta_0 I & 0 \\ * & * & * & \eta_0 I \end{bmatrix} \succ 0, \forall i \in \{1, 2, .., N\}, \end{aligned}$$

as well as the sufficient condition in Theorem 2.3.6, i.e., $\|A_{e,i}\| < 1$ *for all* $i \in \{1, 2, \ldots, N\}$*. Although the designed observer may not be optimum in minimum* $\mathcal{H}_\infty$ *sense when using this alternative method, we can guarantee the steady state convergence of the radii instead.*

Algorithm 1 Fixed-Order Input & State Set-Valued Observer

```
Initialize: $M_1 = \Sigma^{-1}$; $M_2 = (C_2 G_2)^{\dagger}$;
        $\Phi = I - G_2 M_2 C_2$;
        Compute $\tilde{L}$ via Theorem 2.3.3;
        $\Psi = I - \tilde{L} C_2$;
        $\theta \triangleq \max_{i \in \{1,2,\ldots,N\}} \|\Psi\Phi(A^i - G_1 M_1 C_1)\|$;
        $\hat{x}_{0|0} = \hat{x}_0 = \text{centroid}(\hat{X}_0)$;
        $\delta_0^x = \min_{\delta}\{\|x - \hat{x}_{0|0}\| \leq \delta, \forall x \in \hat{X}_0\}$;
        $\hat{d}_{1,0} = M_1(z_{1,0} - C_1 \hat{x}_{0|0} - D_1 u_0)$;
for $k = 1$ to $K$ do
    ▷ Estimation of $d_{2,k-1}$ and $d_{k-1}$
    $\hat{x}_{k|k-1} = \sum_{i=1}^{N} \lambda_{i,k} A^i \hat{x}_{k-1|k-1} + \sum_{i=1}^{N} \lambda_{i,k} B^i u_{k-1}$
        $+ G_1 \hat{d}_{1,k-1}$;
    $\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2 \hat{x}_{k|k-1} - \sum_{i=1}^{N} \lambda_{i,k} D_2^i u_k)$;
    $\hat{d}_{k-1} = V_1 \hat{d}_{1,k-1} + V_2 \hat{d}_{2,k-1}$;
    $\delta_{k-1}^d = \delta_{k-1}^x \|V_1 M_1 C_1 + V_2 M_2 C_2 \hat{A}_k\|$
        $+ \eta_v(\|(V_2 M_2 C_2 G_1 - V_1) M_1 T_1\| + \|V_2 M_2 T_2\|)$
        $+ \eta_w \|V_2 M_2 C_2\|$;
    $\hat{D}_{k-1} = \{d \in \mathbb{R}^l : \|d - \hat{d}_{k-1}\| \leq \delta_{k-1}^d\}$;
    ▷ Time update
    $\hat{x}_{k|k}^{\star} = \hat{x}_{k|k-1} + G_2 \hat{d}_{2,k-1}$;
    ▷ Measurement update
    $\hat{x}_{k|k} = \hat{x}_{k|k}^{\star} + \tilde{L}(z_{2,k} - C_2 \hat{x}_{k|k}^{\star} - \sum_{i=1}^{N} \lambda_{i,k} D_2^i u_k)$;
    $\delta_k^x = \delta_0^x \theta^k + \overline{\eta} \sum_{i=1}^{k} \theta^{i-1}$;
    $\hat{X}_k = \{x \in \mathbb{R}^n : \|x - \hat{x}_{k|k}\| \leq \delta_k^x\}$;
    ▷ Estimation of $d_{1,k}$
    $\hat{d}_{1,k} = M_1(z_{1,k} - C_1 \hat{x}_{k|k} - \sum_{i=1}^{N} D_1^i u_k)$;
end for
```

2.4 Simulation Results

In this section, we consider a convex combination of two constituent linear time-invariant strongly detectable subsystems that have been used in the literature as a benchmark for some state and input filters (e.g., [26]):

$$A^1 = \begin{bmatrix} 0.9 & .5 \\ -0.3 & 1 \end{bmatrix}; A^2 = \begin{bmatrix} 0.85 & .55 \\ -0.35 & 1 \end{bmatrix}; C = \begin{bmatrix} 1 & .2 \\ 1.1 & 1.9 \end{bmatrix};$$
$$G = \begin{bmatrix} -0.02 & 0.04 \\ 0.01 & -0.05 \end{bmatrix}; H = \begin{bmatrix} 1.1 & 2 \\ 2.2 & 4 \end{bmatrix}; B^1 = B^2 = I_{2\times 2}; D = 0_{2\times 2}.$$

The unknown inputs used in this example are as given in Figure 2.1, while the initial state estimate and noise signals (drawn uniformly) have bounds $\delta^x_0 = 0.5$, $\eta_w = 0.02$ and $\eta_v = 10^{-4}$. We also picked uniformly random coefficients, $\lambda_{i,k}$, that satisfies $0 \leq \lambda_{i,k} \leq 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$. Based on the results of Theorem 2.3.3 and by solving the corresponding semi-definite programming problem using YALMIP [73] and MOSEK [12] as the solver, we find $S^\star = \begin{bmatrix} 0.2745 & 0.1933 \\ 0.1933 & 0.4200 \end{bmatrix}$, $Y^\star = \begin{bmatrix} 0.0010 \\ 0.1613 \end{bmatrix}$ and the $\mathcal{H}_\infty$-observer gain as $\tilde{L} = S^{\star\,-1} Y^\star = \begin{bmatrix} -0.3946 \\ 0.5656 \end{bmatrix}$. Then, applying Algorithm 1, we summarized the set-valued state and unknown input results in Figures 2.1 and 2.2. The radii are observed to be convergent to steady state in Figure 2.2.

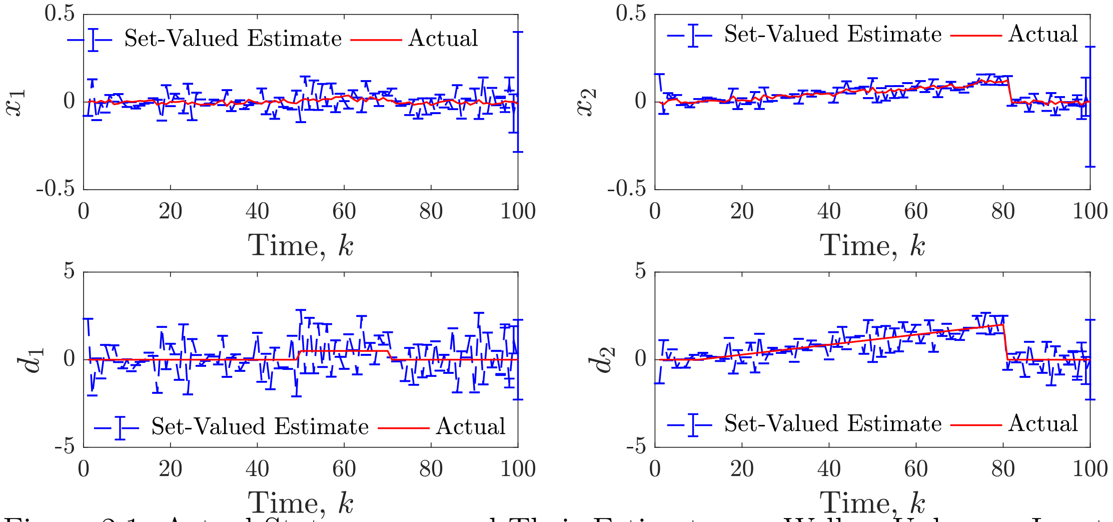


Figure 2.1: Actual States x_1, x_2 and Their Estimates, as Well as Unknown Inputs d_1 and d_2 and Their Estimates

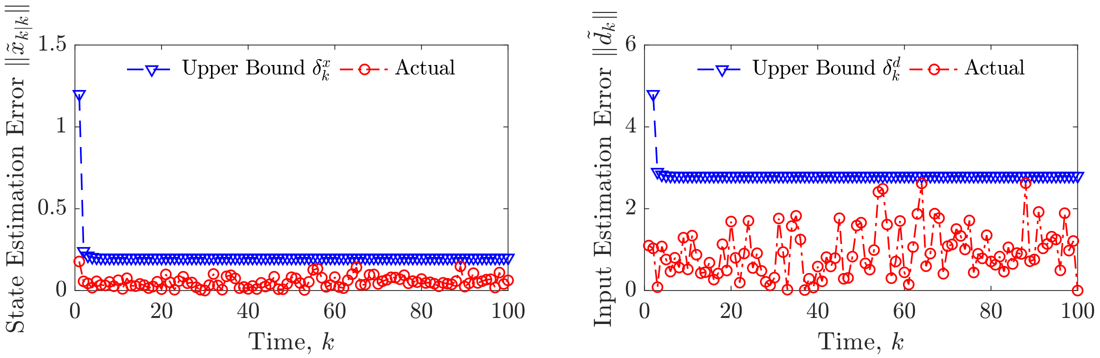


Figure 2.2: Actual Estimation Errors and Radii of Set-valued Estimates of States, $\|\tilde{x}_{k|k}\|, \delta_k^x$, and Unknown Inputs, $\|\tilde{d}_k\|, \delta_k^d$.

2.5 Conclusion

We presented a fixed-order set-valued $\mathcal{H}_\infty$-observer for linear parameter-varying bounded-error discrete-time dynamic systems, which can be expressed as a convex combination of strongly detectable linear time-invariant constituent systems. We provided sufficient conditions for the optimality of the designed observer, which can be obtained from a semi-definite programming problem with LMI constraints. We also showed that the strong detectability of the constituent linear time-invariant systems is necessary for the existence and stability of such an observer and for the boundedness of the set-valued estimates. In addition, we came up with sufficient structural conditions for the convergence of the radii of the set-valued state and input estimates and derived the steady state radii. Finally, we demonstrated the effectiveness of our proposed approach using an illustrative example.

Chapter 3

SIMULTANEOUS MODE, INPUT AND STATE SET-VALUED OBSERVERS WITH APPLICATIONS TO RESILIENT ESTIMATION AGAINST SPARSE ATTACKS

The goal of this chapter [a] is to simultaneously consider state and unknown input estimation as well as mode detection for hidden mode switched linear systems with bounded-norm noise and unknown inputs. To address this, we propose a multiple-model approach that leverages the optimally designed set-valued state and input $\mathcal{H}_\infty$ observers in our previous work [131] to obtain a bank of mode-matched set-valued observers in combination with a novel mode observer based on elimination. Our mode elimination approach uses the upper bound of the norm of to-be-designed residual signals to remove inconsistent modes from the bank of observers. In particular, we provide a tractable method to calculate an upper bound signal for the residual's norm and prove that the upper bound signal is a convergent sequence. Moreover, we provide sufficient conditions to guarantee that all false modes will be eventually eliminated.

3.1 Problem Statement

Consider a hidden mode switched linear system with bounded-norm noise and unknown inputs (i.e., a hybrid system with linear and noisy system dynamics in each mode, and the mode and some inputs are not known/measured):

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k^q + G^q d_k^q + w_k, \\ y_k &= Cx_k + Du_k^q + H^q d_k^q + v_k, \end{aligned} \tag{3.1}$$

---

[a] The content of this chapter is documented as a published paper in [116].

where $x_k \in \mathbb{R}^n$ is the continuous system state and $q \in \mathbb{Q} = \{1, 2, \ldots, Q\}$ is the hidden discrete state or *mode*. For each (fixed) mode q, $u_k^q \in U_k^q \subset \mathbb{R}^m$ is the *known* input, $d_k^q \in \mathbb{R}^p$ the unknown but *sparse* input or attack signal, i.e., every vector d_k^q has precisely $\rho \in \mathbb{N}$ nonzero elements where ρ is a known parameter, $y_k \in \mathbb{R}^l$ is the output, whereas $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are process and measurement 2-norm bounded disturbances with known parameters η_w and η_v as their 2-norm bounds respectively. The matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $G^q \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{l\times n}$, $D \in \mathbb{R}^{l\times m}$ and $H^q \in \mathbb{R}^{l\times p}$ are known and no prior 'useful' knowledge or assumption of the dynamics of d_k^q, except *sparsity* is assumed.

More precisely, G^q and H^q represent the different hypothesis for each mode $q \in \mathbb{Q}$, about the sparsity pattern of the unknown inputs, which in the context of sparse attacks corresponds to which actuators and sensors are attacked or not attacked. In other words, we assume that $G^q = G\mathbb{I}_G^q$ and $H^q = H\mathbb{I}_H^q$ for some input matrices $G \in \mathbb{R}^{n\times t_a}$ and $H \in \mathbb{R}^{l\times t_s}$, where t_a and t_s are the number of vulnerable actuator and sensor signals respectively. Note that $\rho_a^q \leq t_a \leq m$ and $\rho_s^q \leq t_s \leq l$, where ρ_a^q (ρ_s^q) is the number of attacked actuator (sensor) signals and clearly cannot exceed the number of vulnerable actuator (sensor) signals, which in turn cannot exceed the total number of actuators (sensors). Furthermore, we assume that the total number of unknown inputs/attacks in each mode is known and equals $\rho = \rho_a + \rho_s$ (sparsity assumption). Moreover, the *index matrix* $\mathbb{I}_G^q \in \mathbb{R}^{t_a\times\rho}$ ($\mathbb{I}_H^q \in \mathbb{R}^{t_s\times\rho}$) represents the sub-vector of $d_k \in \mathbb{R}^\rho$ that indicates signal magnitude attacks on the actuators (sensors).

Note that the approach in our paper can be easily extended to handle mode-dependent A, B, C, D, w_k, v_k, η_w and η_v but is omitted to simplify the notation. Moreover, throughout the paper, we assume, without loss of generality, that for each possible mode q, the system (A, G^q, C, H^q) is strongly detectable [131, Definition 1], since this is a necessary and sufficient condition for obtaining meaningful set-valued state and input estimates when the mode is known.

Using the modeling framework above, the simultaneous state, unknown input and hidden mode estimation problem is threefold and can be stated as follows:

Problem 3.1.1. *Given a switched linear hidden mode discrete-time bounded-error system with unknown inputs* (3.1),

1. *Design a bank of mode-matched observers that for each mode optimally finds the set estimates of compatible states and unknown inputs in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification, conditional on the mode being true.*
2. *Develop a mode observer via elimination and the corresponding criterion to eliminate false modes.*
3. *Find sufficient conditions for eliminating all false modes.*

3.2 Prpposed Observer Design

In this section, we propose a multiple-model approach for simultaneous mode, state and unknown input estimation for (3.1), where the goal of the observer is to find compatible set estimates $\hat{D}_k$, $\hat{X}_k$ and $\hat{\mathbb{Q}}_k$ for unknown inputs, states and modes at time step k, respectively.

3.2.1 Overview of Multiple-Model Approach

The multiple-model design approach consists of three components: (i) designing a bank of mode-matched set-valued observers, (ii) designing a mode observer for eliminating incompatible modes using residual detectors, and (iii) a global fusion observer that outputs the desired set-valued mode, input and state estimates.

Mode-Matched Set-Valued Observer

First, we design a bank of mode-matched observers, which consists of Q simultaneous state and input $\mathcal{H}_\infty$ set-valued observers based on the optimal fixed-order observer design in [131], which we briefly summarize here. For each mode-matched observer corresponding to mode q, following the approach in [131, Section 3.1], we consider set-valued fixed-order estimates of the form:

$$\hat{D}^q_{k-1} = \{d_{k-1} \in \mathbb{R}^p : \|d_{k-1} - \hat{d}^q_{k-1}\| \leq \delta^{d,q}_{k-1}\}, \tag{3.2}$$

$$\hat{X}^q_k = \{x_k \in \mathbb{R}^n : \|x_k - \hat{x}^q_{k|k}\| \leq \delta^{x,q}_k\}, \tag{3.3}$$

where their centroids are obtained with the following three-step recursive observer that is optimal in $\mathcal{H}_\infty$-norm sense:

Unknown Input Estimation:

$$\begin{aligned} \hat{d}^q_{1,k} &= M^q_1(z^q_{1,k} - C^q_1\hat{x}^q_{k|k} - D^q_1 u^q_k) \\ \hat{d}^q_{2,k-1} &= M^q_2(z^q_{2,k} - C^q_2\hat{x}^q_{k|k-1} - D^q_2 u^q_k) \\ \hat{d}^q_{k-1} &= V^q_1\hat{d}^q_{1,k-1} + V^q_2\hat{d}^q_{2,k-1} \end{aligned} \tag{3.4}$$

Time Update:

$$\begin{aligned} \hat{x}^q_{k|k-1} &= A\hat{x}^q_{k-1|k-1} + Bu^q_{k-1} + G^q_1\hat{d}^q_{1,k-1} \\ \hat{x}^{\star,q}_{k|k} &= \hat{x}^q_{k|k-1} + G^q_2\hat{d}^q_{2,k-1} \end{aligned} \tag{3.5}$$

Measurement Update:

$$\hat{x}^q_{k|k} = \hat{x}^{\star,q}_{k|k} + \tilde{L}^q(z^q_{2,k} - C^q_2\hat{x}^{\star,q}_{k|k} - D^q_2 u^q_k) \tag{3.6}$$

where $\tilde{L}^q \in \mathbb{R}^{n\times(l-p_{H^q})}$, $M^q_1 \in \mathbb{R}^{p_{H^q}\times p_{H^q}}$ and $M^q_2 \in \mathbb{R}^{(p-p_{H^q})\times(l-p_{H^q})}$ are observer gain matrices that are chosen in the following theorem from [131] to minimize the "volume" of the set of compatible states and unknown inputs, quantified by the radii $\delta^{d,q}_{k-1}$ and $\delta^{x,q}_k$.

Theorem 3.2.1. *[131, Lemma 2 & Theorem 4] Suppose the system* (A, G^q, C, H^q) *is strongly detectable,* $M_1^q \Sigma^q = I$ *and* $M_2^q C_2^q G_2^q = I$*. Then, for each mode* q*, there exists a stable and optimal (in* $\mathcal{H}_\infty$*-norm sense) observer with gain* $\tilde{L}^q$*, where the input and state estimation errors,* $\tilde{d}_{k-1}^q \triangleq d_{k-1}^q - \hat{d}_{k-1}^q$ *and* $\tilde{x}_{k|k}^q \triangleq x_k - \hat{x}_{k|k}^q$*, are bounded for all* k *(i.e., the set-valued estimates are bounded with radii* $\delta_{k-1}^{d,q}, \delta_k^{x,q} < \infty$*), and the observer gains and the set estimates are given in [131, Theorem 2 & Algorithm 1].*

Mode Estimation Observer

To estimate the set of compatible modes, we consider an elimination approach that compares residual signals against some thresholds. Specifically, we will eliminate a specific mode q, if $\|r_k^q\|_2 > \hat{\delta}_{r,k}^q$, where the residual signal r_k^q is defined as follows and the thresholds $\hat{\delta}_{r,k}^q$ will be derived in Section 5.2.3.

Definition 3.2.2 (Residuals)**.** *For each mode* q *at time step* k*, the residual signal is defined as:*

$$r_k^q \triangleq z_{2,k}^q - C_2^q \hat{x}_{k|k}^{\star,q} - D_2^q u_k^q.$$

Global Fusion Observer

Then, combining the outputs of both components above, our proposed global fusion observer will provide mode, unknown input and state set-valued estimates at each time step k as:

$$\hat{\mathbb{Q}}_k = \{q \in \mathbb{Q} \,\Big|\, \|r_k^q\|_2 \leq \hat{\delta}_{r,k}^q\},$$
$$\hat{D}_{k-1} = \cup_{q \in \hat{\mathbb{Q}}_k} D_{k-1}^q, \;\; \hat{X}_k = \cup_{q \in \hat{\mathbb{Q}}_k} X_k^q.$$

The multiple-model approach is summarized in Algorithm 2.

Algorithm 2 Simultaneous Mode, State and Input Estimation

```
$\hat{\mathbb{Q}}_0 = \mathbb{Q}$;
for $k = 1$ to $N$ do
    for $q \in \hat{\mathbb{Q}}_{k-1}$ do
        ▷ Mode-Matched State and Input Set-Valued Estimates
        Compute $T_2^q, M_1^q, M_2^q, \tilde{L}^q, \hat{x}_{k|k}^{\star,q}, \hat{X}_k^q, \hat{D}_{k-1}^q$ via Theorem 3.2.1;
        $z_{2,k}^q = T_2^q y_k$;
        ▷ Mode Observer via Elimination
        $\hat{\mathbb{Q}}_k = \hat{\mathbb{Q}}_{k-1}$;
        Compute $r_k^q$ via Definition 3.2.2 and $\hat{\delta}_{r,k}^q$ via Theorem 3.2.7;
        if $\|r_k^q\|_2 > \hat{\delta}_{r,k}^q$ then $\hat{\mathbb{Q}}_k = \hat{\mathbb{Q}}_k \backslash \{q\}$;
        end if
    end for
    ▷ State and Input Estimates
    $\hat{X}_k = \cup_{q \in \hat{\mathbb{Q}}_k} \hat{X}_k^q$;  $\hat{D}_k = \cup_{q \in \hat{\mathbb{Q}}_k} \hat{D}_k^q$;
end for
```

3.2.2 Mode Elimination Approach

The idea is simple. If the residual signal of a particular mode exceeds its upper bound conditioned on this mode being true, we can conclusively rule it out as incompatible. To do so, for each mode q, we first compute an upper bound ($\hat{\delta}_{r,k}^q$) for the 2-norm of its corresponding residual at time k, conditioned on q being the *true* mode. Then, comparing the 2-norm of residual signal in Definition 3.2.2 with $\hat{\delta}_{r,k}^q$, we can eliminate mode q if the residual's 2-norm is strictly greater than the upper bound. This can be formalized using the following proposition and theorem.

Proposition 3.2.3. *Consider mode q at time step k, its residual signal r_k^q (as defined in Definition 3.2.2) and the unknown true mode q^*. Then,*

$$
\begin{aligned}
&r_k^q = r_k^{q|*} + \Delta r_k^{q|q*}, \mathit{where}\\
&r_k^{q|*} \triangleq z_{2,k}^{q*} - C_2^q \hat{x}_{k|k}^{\star,q} - D_2^q u_k^q = T_2^{q*} y_k - C_2^q \hat{x}_{k|k}^{\star,q} - D_2^q u_k^q,\\
&\Delta r_k^{q|q*} \triangleq (T_2^q - T_2^{q*}) y_k,
\end{aligned}
$$

where $r_k^{q|*}$ *is the true mode's residual signal (i.e.,* $q = q^*$*), and* $\Delta r_k^{q|q^*}$ *is the* residual error.

Proof. This follows directly from plugging the above expressions into the right hand side term of Definition 3.2.2. □

Theorem 3.2.4. *Consider mode* q *and its residual signal* r_k^q *at time step* k*. Assume that* $\delta_{r,k}^{q,*}$ *is any signal that satisfies* $\|r_k^{q|*}\|_2 \leq \delta_{r,k}^{q,*}$*, where* $r_k^{q|*}$ *is defined in Proposition 3.2.3. Then, mode* q *is not the true mode, i.e., can be eliminated at time* k*, if* $\|r_k^q\|_2 > \delta_{r,k}^{q,*}$*.*

Proof. To use contradiction, suppose q is the true mode. By uniqueness of the true mode $q = q^*$, so $T_2^q = T_2^{q*}$ and by Proposition 3.2.3, $\Delta r_k^{q|q*} = 0$ and hence $\|r_k^q\|_2 = \|r_k^{q|*}\|_2 \leq \delta_{r,k}^{q,*}$, which contradicts with the assumption. □

3.2.3 Tractable Computation of Thresholds

Theorem 3.2.4 provides a sufficient condition for mode elimination at each time step. To apply this sufficient condition, we need to compute an upper bound for $\|r_k^{q|*}\|_2$, i.e., our $\delta_{r,k}^{q,*}$ signal (cf. Theorem 3.2.7) and show that it is bounded in the following lemmas.

Lemma 3.2.5. *Consider any mode* q *with the unknown true mode being* q^**. Then, at time step* k*, we have*

$$
r_k^{q|*} = C_2^q \tilde{x}_{k|k}^{\star,q} + v_{2,k}^q = \mathbb{A}_k^q t_k, \tag{3.7}
$$

where $t_k \triangleq \begin{bmatrix} \tilde{x}_{0|0}^\top & w_0^\top & \dots & w_{k-1}^\top & v_0^\top & \dots & v_k^\top \end{bmatrix}^\top \in \mathbb{R}^{(n+l)(k+1)}$,

$$\begin{aligned}\mathbb{A}_k^q \triangleq & [C_2^q \overline{A}^q A_e^{q\,k-1} | C_2^q \overline{A}^q A_e^{q\,k-2} B_{e,w}^q | C_2^q \overline{A}^q A_e^{q\,k-2} B_{e,w}^q \dots \\ & C_2^q \overline{A}^q A_e^{q\,k-1-i} B_{e,w}^q | \dots | C_2^q \overline{A}^q A_e^q B_{e,w}^q | C_2^q B_{e,w}^{\star,q} | \\ & C_2^q \overline{A}^q A_e^{q\,k-2} B_{e,v_1}^q | C_2^q \overline{A}^q A_e^{q\,k-2} (B_{e,v_1}^q + A_e^q B_{e,v_2}^q) \dots \\ & C_2^q \overline{A}^q A_e^{q\,k-1-i} (B_{e,v_1}^q + A_e^q B_{e,v_2}^q) | \dots | \\ & C_2^q \overline{A}^q A_e^q (B_{e,v_1}^q + A_e^q B_{e,v_2}^q) | C_2^q (B_{e,v_1}^{q,\star} + \overline{A}^q B_{e,v_2}^q) | \\ & C_2^q B_{e,v_2}^{q,\star} + T_2^q] \in \mathbb{R}^{(l-p_{H^q}) \times (n+l)(k+1)},\end{aligned}$$

with $\overline{A}^q \triangleq (I - G_2^q M_2^q C_2^q)(A - G_1^q M_1^q C_1^q)$, $A_e^q \triangleq (I - \tilde{L}^q C_2^q)\overline{A}^q$, $B_{e,w}^{\star,q} \triangleq (I - G_2^q M_2^q C_2^q)$, $B_{e,v1}^{\star,q} \triangleq -(I - G_2^q M_2^q C_2^q)(G_1^q M_1^q T_1^q)$, $B_{e,w}^q \triangleq (I - \tilde{L}^q C_2^q) B_{e,w}^{\star,q}$, $B_{e,v1}^q \triangleq (I - \tilde{L}^q C_2^q) B_{e,v1}^{\star,q}$ *and* $B_{e,v2}^q \triangleq (I - \tilde{L}^q C_2^q) B_{e,v2}^{\star,q} - \tilde{L}^q T_2^q$, $B_{e,v2}^{\star,q} \triangleq -G_2^q M_2^q T_2^q$.

Proof. Considering (3.7), the first equality comes from Definition 3.2.2 and $z_{2,k}^q = C_2^q x_k + D_{2,k}^q u_k^q + v_{2,k}^q$ from [131], assuming that q is the true mode, and the second equality is implied by the first equality and the fact in [131, Appendix C] that

$$\begin{aligned}\tilde{x}_{k|k}^{\star,q} &= \overline{A}^q A_e^{q\,k-1} \tilde{x}_{0|0} + \overline{A}^q A_e^{q\,k-2} \begin{bmatrix} B_{e,w}^q B_{e,v1}^q \end{bmatrix} \vec{w}_0 \\ &\quad + B_{e,w}^{\star,q} w_{k-1} + (B_{e,v1}^{\star,q} + \overline{A}^q B_{e,v2}^q) v_{k-1} + B_{e,v2}^{\star,q} v_k \\ &\quad + \textstyle\sum_{i=1}^{k-2} \overline{A}^q A_e^{q\,k-1-i} \begin{bmatrix} B_{e,w}^q & B_{e,v1}^q + A_e^q B_{e,v2}^q \end{bmatrix} \vec{w}_i, \\ \vec{w}_k &\triangleq \begin{bmatrix} w_k^\top & v_k^\top \end{bmatrix}^\top. \end{aligned}$$

□

Lemma 3.2.6. *For each mode q at time step k, there exists a generic finite valued upper bound $\delta_{r,k}^q < \infty$ for $\|r_k^{q|*}\|_2$.*

Proof. Consider the following optimization problem for $\|r_k^{q|*}\|_2$ by leveraging Lemma 3.2.5:

$$\begin{aligned}
&\delta^q_{r,k} \triangleq \max_{t_k} \|\mathbb{A}^q_k t_k\|_2 \qquad (3.8)\\
&s.t.\ t_k = \begin{bmatrix} \tilde{x}^\top_{0|0} & w^\top_0 & \dots & w^\top_{k-1} & v^\top_0 & \dots & v^\top_k \end{bmatrix}^\top,\\
&\|\tilde{x}_{0|0}\|_2 \leq \delta^x_0,\ \|w_i\|_2 \leq \eta_w,\ \|v_j\|_2 \leq \eta_v,\\
&i \in \{0, ..., k-1\},\ j \in \{0, ..., k\}.
\end{aligned}$$

The objective 2-norm function is continuous and the constraint set is an intersection of level sets of lower dimensional norm functions, which is closed and bounded, so is compact. Hence, by Weierstrass Theorem [15, Proposition 2.1.1], the objective function attains its maxima on the constraint set and so a finite-valued upper bound exists. □

Clearly $\delta^q_{r,k}$ in Lemma 3.2.6 is the *tightest* possible residual norm's upper bound and potentially can eliminate the most possible number of modes, so is the best choice if we can calculate it. But, notice that although it was straight forward to show that a finite-valued $\delta^q_{r,k}$ exists, but since the optimization problem in Lemma 3.2.6 is a *norm maximization* (not minimization) over the intersection of level sets of lower dimensional norm functions, i.e., a non-concave maximization over intersection of quadratic constraints, it is an NP-hard problem [18]. To tackle with this complexity, we provide an over-approximation for $\delta^q_{r,k}$ in the following Theorem 3.2.7, which we call $\hat{\delta}^q_{r,k}$.

Theorem 3.2.7. *Consider mode q. At time step k, let*

$$
\begin{aligned}
&\hat{\delta}_{r,k}^{q} \triangleq \min\{\delta_{r,k}^{q,inf}, \delta_{r,k}^{q,tri}\},\\
&\delta_{r,k}^{q,inf} \triangleq \|\mathbb{A}_k^q t_k^\star\|_2,\\
&\delta_{r,k}^{q,tri} \triangleq \delta_0^{x,q}\|C_2^q\overline{A}^q A_e^{qk-1}\|_2 + \eta_w\|C_2^q\overline{A}^q A_e^{qk-2}\|_2 +\\
&\quad \textstyle\sum_{i=1}^{k-2}[\eta_w\|C_2^q\overline{A}^q A_e^{qi}B_{e,w}^q\|_2 + \eta_v\|C_2^q\overline{A}^q A_e^{qi}(B_{e,v_1}^q + A_e^q B_{e,v_2}^q)\|_2]\\
&\quad + \eta_v(\|C_2^q\overline{A}^q A_e^{qk-2}B_{e,v_1}^q\|_2 + \|C_2^q(B_{e,v_1}^{q,\star} + \overline{A}^q B_{e,v_2}^q)\|_2)\\
&\quad + \|C_2^q B_{e,v_2}^{q,\star} + T_2^q\|_2) + \eta_w\|C_2^q B_{e,w}^{\star,q}\|_2,
\end{aligned}
$$

where $t_k^\star$ is a vertex of the following hypercube:

$$
\mathcal{X}_k^q \triangleq \Big\{x \in \mathbb{R}^{(n+l)(k+1)} \;\Big|\; |x(i)| \leq \begin{cases} \delta_0^x, 1 \leq i \leq n \\ \eta_w, n+1 \leq i \leq n(k+1) \\ \eta_v, n(k+1)+1 \leq i \leq (n+l)(k+1) \end{cases} \Big\},
$$

i.e.,

$$
t_k^\star(i) \in \begin{cases} \{-\delta_0^x, \delta_0^x\}, 1 \leq i \leq n, \\ \{-\eta_w, \eta_w\}, n+1 \leq i \leq n(k+1), \\ \{-\eta_v, \eta_v\}, n(k+1)+1 \leq i \leq (n+l)(k+1). \end{cases}
$$

Then, $\hat{\delta}_{r,k}^q$ is an over-approximation for $\delta_{r,k}^q$ in Lemma 3.2.6.

Proof. Consider the optimization problem

$$
\begin{aligned}
&\delta_{r,k}^{q,inf} \triangleq \max_{t_k} \|\mathbb{A}_k^q t_k\|_2 \qquad (3.9)\\
&s.t.\ t_k = \begin{bmatrix} \tilde{x}_{0|0}^\top & w_0^\top & \dots & w_{k-1}^\top & v_0^\top & \dots & v_k^\top \end{bmatrix},\\
&\qquad \|\tilde{x}_{0|0}\|_\infty \leq \delta_0^x,\ \|w_i\|_\infty \leq \eta_w,\ \|v_j\|_\infty \leq \eta_v,\\
&\qquad \forall i \in \{0, ..., k-1\},\ \forall j \in \{0, ..., k\}.
\end{aligned}
$$

Comparing (A.61) and (3.9), the two problems have the same objective functions, while since $\|.\|_\infty \leq \|.\|_2$, the constraint set for (A.61) is a subset of the one for (3.9). Hence $\delta^q_{r,k} \leq \delta^{q,inf}_{r,k}$. Also, it is easy to see that $\hat{\delta}^q_{r,k} \leq \delta^{q,tri}_{r,k}$, using triangle and sub-multiplicative inequalities. Moreover, (3.9) is a *maximization* of a convex objective function over a convex constraint (hypercube $\mathcal{X}^q_k$). By a famous result [101, Corollary 32.2.1], in such a problem, the objective function attains its maxima on some of the extreme points of the constraint set, which in this case are the vertices of the hypercube $\mathcal{X}^q_k$. $\square$

It can be easily seen as a corollary of Theorem 3.2.7 that:

Corollary 3.2.8. $\eta^t_k \triangleq \|t^\star_k\|_2 = \sqrt{n{\delta^x_o}^2 + kn\eta^2_w + (k+1)l\eta^2_v}$.

Theorem 3.2.7 enables us to obtain an upper bound for $\|r^{q|*}_k\|_2$, by enumerating the objective function in (3.9) at vertices of the hypercube $\mathcal{X}^q_k$ and choosing the largest value as $\delta^{q,inf}_{r,k}$. Moreover, we can easily calculate $\delta^{q,tri}_{r,k}$; then, the upper bound is chosen as the minimum of the two as $\hat{\delta}^q_{r,k}$.

Remark 3.2.9. *Although simulation results indicate that especially in earlier time steps, $\delta^{q,inf}_{r,k}$ may have smaller values than $\delta^{q,tri}_{r,k}$, but if we only consider $\delta^{q,inf}_{r,k}$ as the over-approximation and do not use $\delta^{q,tri}_{r,k}$, then we will face two difficulties. First, as time increases, the number of required enumerations (i.e., the number of hypercube's vertices which is $2^{(n+l)(k+1)}$) increases with an exponential rate. Second and more importantly, as Lemma 3.3.4 will indicate later, $\delta^{q,inf}_{r,k}$ goes to infinity as time increases, so it will be unlikely to eliminate any mode when the time step is large, i.e., asymptotically speaking, $\delta^{q,inf}_{r,k}$ will be useless. In contrast, again by Lemma 3.3.4, $\delta^{q,tri}_{r,k}$ converges to some steady-state value, so it can be always used as an over-approximation for $\delta^q_{r,k}$ in the mode elimination process.*

3.3 Mode Detectability

In addition to the nice properties regarding the stability and boundedness of the mode-matched set estimates of state and input obtained from [131], we now provide some sufficient conditions for the system dynamics, which guarantee that regardless of the observations, after some large enough time steps, *all* the false (i.e., not true) modes can be eliminated, when applying Algorithm 1. To do so, first, we define the concept of mode detectability as well as some assumptions for deriving our sufficient conditions for mode detectability.

Definition 3.3.1 (Mode Detectability)**.** *System* (3.1) *is called mode detectable if there exists a natural number* $K > 0$*, such that for all time steps* $k \geq K$*, all false modes are eliminated.*

Assumption 3.3.2. *There exist known* $R_y, R_x \in \mathbb{R}$ *such that* $\forall k, y_k \in Y \triangleq \{y \in \mathbb{R}^l | \, \|y\|_2 \leq R_y\}$ *and* $x_k \in X \triangleq \{x \in \mathbb{R}^n | \, \|x\|_2 \leq R_x\}$*, i.e., there exist known bounds for the whole observation/measurement and state spaces, respectively.*

Assumption 3.3.3. *The unknown input/attack signal has an* unlimited energy*, i.e.,*

$$\lim_{k \to \infty} \|d^{q*}_{0:k}\|_2 = \infty,$$

where $d^{q*}_{0:k} \triangleq \begin{bmatrix} d^{q*\top}_k & d^{q*\top}_{k-1} & \dots & d^{q*\top}_0 \end{bmatrix}^\top$.

Note that Assumption 3.3.3 is not restrictive because otherwise, the unknown input/attack signal must vanish asymptotically, which means that the true mode (with no unknown inputs) can be inferred asymptotically.

In order to derive the desired sufficient conditions for mode detectability in Theorem 3.3.7, we first present the following Lemmas 3.3.4–3.3.6. For the sake of clarity, the proofs of these results are given in the Appendix.

Lemma 3.3.4. *For each mode* q*,*

$$\lim_{k\to\infty} \delta^{q,inf}_{r,k} = \infty. \tag{3.10}$$

$$\lim_{k\to\infty} \hat{\delta}^{q}_{r,k} = \lim_{k\to\infty} \delta^{q,tri}_{r,k} \leq \lim_{k\to\infty} \overline{\delta}^{q,tri}_{r,k} = \overline{\delta}^{q,tri}_{r} < \infty, \tag{3.11}$$

where

$$\begin{aligned}
\overline{\delta}^{q,tri}_{r,k} &\triangleq \delta_0^{x,q} \| C_2^q \overline{A}^q A_e^{q\,k-1} \|_2 + \eta_w \| C_2^q \overline{A}^q A_e^{q\,k-2} \|_2 \\
&+ \eta_w [\| C_2^q \overline{A}^q A_e^q \|_2 \| B^q_{e,w} \|_2 \sum_{i=0}^{k-3} (\| A_e^q \|_2^i) + \| C_2^q B^{\star,q}_{e,w} \|_2] \\
&+ \eta_v [\| C_2^q \overline{A}^q A_e^q \|_2 \| B^q_{e,v_1} + A_e^q B^q_{e,v_2} \|_2 \sum_{i=0}^{k-3} \| A_e^q \|_2^i] \\
&+ \eta_v [\| C_2^q B^{q,\star}_{e,v_2} + T_2^q \|_2 + \| C_2^q (B^{q,\star}_{e,v_1} + \overline{A}^q B^q_{e,v_2}) \|_2] + \eta_v \| C_2^q \overline{A}^q A_e^{q\,k-2} B^q_{e,v_1} \|_2, \\
\overline{\delta}^{q,tri}_{r} &\triangleq \eta_w [\| C_2^q B^{q,\star}_{e,w} \|_2 + \frac{\| C_2^q \overline{A}^q A_e^q \|_2}{(1-\theta^q)} + \| B^q_{e,w} \|_2] + \eta_v [\| B^q_{e,v_1} + A_e^q B^q_{e,v_2} \|_2 \\
&+ \| C_2^q B^{q,\star}_{e,v_2} + T_2^q \|_2 + \| C_2^q (B^{q,\star}_{e,v_1} + \overline{A}^q B^q_{e,v_2}) \|_2], \;\; \theta^q \triangleq \| A_e^q \|_2,
\end{aligned}$$

with $\overline{A}^q$, A_e^q, $B^q_{e,w}$, $B^{q,\star}_{e,w}$, B^q_{e,v_1}, $B^{q,\star}_{e,v_1}$, B^q_{e,v_2} *and* $B^{q,\star}_{e,v_2}$ *given in Lemma 3.2.5.*

Lemma 3.3.5. *Suppose that Assumption 3.3.2 holds. Consider two different modes* $q \neq q' \in Q$ *and their corresponding upper bounds for their residuals' norms,* $\delta^q_{r,k}$ *and* $\delta^{q'}_{r,k}$*, at time step* k*. At least one of the two modes* $q \neq q'$ *will be eliminated if*

$$\| C_2^q \hat{x}^{\star,q}_{k|k} - C_2^{q'} \hat{x}^{\star,q'}_{k|k} + D_2^q u_k^q - D_2^{q'} u_k^{q'} \|_2 > \delta^q_{r,k} + \delta^{q'}_{r,k} + R_z^{q,q'} \tag{3.12}$$

where $R_z^{q,q'} \triangleq R_y \| T_2^q - T_2^{q'} \|_2$.

Lemma 3.3.6. *Consider any mode* q *with the unknown true mode being* q^**. Then, at time step* k*, we have*

$$r_k^q = \begin{bmatrix} \mathbb{T}_k^{q,q^*} & \mathbb{B}_k^{q,q^*} & \mathbb{D}_k^{q,q^*} \end{bmatrix} \begin{bmatrix} t_k^\top & u_{0:k}^{q^*\top} & d_{0:k}^{q*\top} \end{bmatrix}^\top,$$

where $u_{0:k}^{q^*} \triangleq \begin{bmatrix} u_k^{q*\top} & u_{k-1}^{q*\top} & \dots u_0^{q*\top} \end{bmatrix}^\top$,

$$\mathbb{T}_k^{q,q^*} \triangleq (T_2^{q^*} - T_2^q) \begin{bmatrix} CA^k & CA^{k-1} & \dots & C & I \end{bmatrix} + \mathbb{A}_k^q,$$
$$\mathbb{B}_k^{q,q^*} \triangleq (T_2^{q^*} - T_2^q) \begin{bmatrix} D & CB & CAB & \dots & CA^{k-1}B \end{bmatrix},$$
$$\mathbb{D}_k^{q,q^*} \triangleq (T_2^{q^*} - T_2^q) \begin{bmatrix} H & CG & CAG \dots & CA^{k-1}G \end{bmatrix},$$

with t_k *given in Lemma 3.2.5 and* $d_{0:k}^{q*}$ *in Assumption 3.3.3.*

Theorem 3.3.7 (Sufficient Conditions for Mode Detectability)**.** *System* (3.1) *is mode detectable, i.e., all false modes will be eliminated after some large enough time step* K*, using Algorithm 1, if the assumptions in Theorem 3.2.1 and either of the following hold:*

i. Assumption 3.3.2 and $\forall q, q' \in Q$*,* $q \neq q'$*,*

$$\sigma_{min}(W^{q,q'}) > \frac{\overline{\delta}_r^{q,tri} + \overline{\delta}_r^{q',tri} + R_y'^{q,q'}}{\sqrt{R_x^2 + \eta_v^2}};$$

ii. Assumption 3.3.3 and $T_2^q \neq T_2^{q'}$ *holds* $\forall q, q' \in Q, q \neq q'$*,*

where $W^{q,q'} \triangleq \begin{bmatrix} (C_2^q - C_2^{q'}) & (T_2^q - T_2^{q'}) & -I & I & D_2^q & -D_2^{q'} \end{bmatrix}$.

3.4 Simulation Results

We consider a system that has been used as a benchmark for many state and input filters/observers (e.g.,[137]):

$$A = \begin{bmatrix} 0.5 & 2 & 0 & 0 & 0 \\ 0 & 0.2 & 1 & 0 & 1 \\ 0 & 0 & 0.3 & 0 & 1 \\ 0 & 0 & 0 & 0.7 & 1 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}; G = \begin{bmatrix} 1 \\ 0.1 \\ 0.1 \\ 1 \\ 0 \end{bmatrix}; H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$B = 0_{5\times 1}; C = I_5; D = 0_{5\times 1}.$$

The unknown inputs used in this example are as given in Figure 3.2, while the initial state estimate and noise signals have bounds $\delta_x = 0.5$, $\eta_w = 0.02$ and $\eta_v = 10^{-4}$. We assume possible attacks on the actuator and four of five sensors, i.e., $t_a = 1$ and $t_s = 4$. Moreover, we assume that there are $\rho = 4$ attacks, so we should consider $Q = \binom{5}{4} = 5$ modes. Table 3.1 indicates different modes, their attack location(s) and the matrix T_2^q for each mode q, where, as can be observed, the second set of sufficient conditions in Theorem 3.3.7 holds, i.e., $T_2^q \neq T_2^{q'}$ for all $q \neq q'$, so we expect that after some large enough time, all the false modes be eliminated, i.e., at most one (true) mode remains at each time step, which can be seen in Figure 3.1, where the number of eliminated modes at each time step is exhibited.

Moreover, for each specific mode q, the signals $\|r_k^q\|_2$, $\|r_k^{q|*}\|_2$, $\delta_{r,k}^{q,tri}$ and $\delta_{r,k}^{q,inf}$ are depicted in Figure 3.1. As can be seen, up to some large enough time, at different time intervals for different modes, one of the upper bounds may be tighter than the other, or vice-versa, so it is reasonable that we consider a minimum of them as the computed upper bound in our mode elimination algorithm. Furthermore, for all modes, $\delta_{r,k}^{q,tri}$ is eventually convergent while $\delta_{r,k}^{q,inf}$ diverges, as we proved in Lemma 3.3.4. So, after some large enough time, $\delta_{r,k}^{q,tri}$ can be used as our upper-bound, while $\delta_{r,k}^{q,inf}$ becomes useless. The corresponding set-valued estimates are provided in Figure 3.2.

Table 3.1: Different Modes and Their T_2^q

Mode	Attack location(s)	T_2^q
$q = 1$	Actuator & Sensors 1,2,3	$[0.2518\ -0.1068\ -0.2409\ -0.5862\ 0.7236]^\top$
$q = 2$	Actuator & Sensors 1,2,4	$[0.0080\ 0.7604\ -0.1522\ -0.5862\ -0.6313]^\top$
$q = 3$	Actuator & Sensors 1,3,4	$[-0.5357\ 0.7289\ 0.1984\ -0.3774\ 0.0009]^\top$
$q = 4$	Actuator & Sensors 2,3,4	$[0.7092\ -0.5570\ -0.1797\ -0.3295\ 0.2143]^\top$
$q = 5$	Sensors 1,2,3,4	$[0.1679\ -0.5682\ 0.5198\ -0.4883\ 0.3747]^\top$

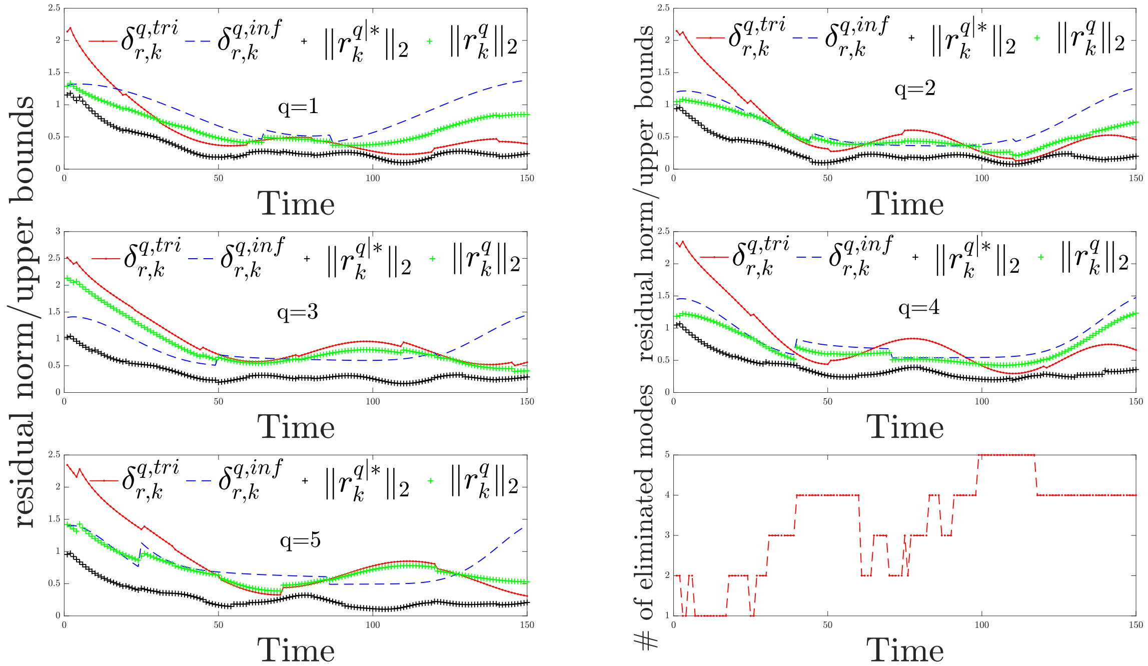


Figure 3.1: $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes, as Well as the Number of Eliminated Modes in Time

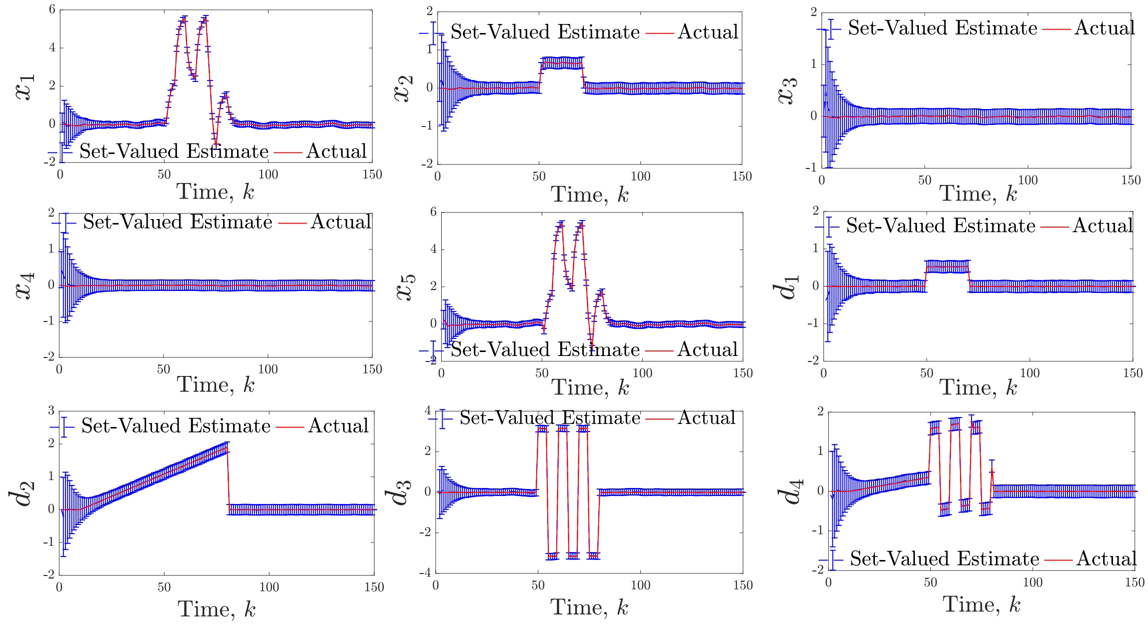


Figure 3.2: State and Unknown Input Set-valued Estimates

3.5 Conclusion

We proposed a residual-based approach for hidden mode switched linear systems with bounded-norm noise and unknown attack signals. The proposed approach at each time step, removes the inconsistent modes and their corresponding observers from a bank of estimators, which includes mode-matched observers. Each mode-matched observer, conditioned on its corresponding mode being true, simultaneously finds bounded sets of states and unknown inputs that include the true state and inputs. Our mode elimination criterion required a bounded upper bound for the residual's norm, for which we proved its existence and computed it by over-approximating the value function of a non-concave NP-hard norm-maximization problem by expanding its constraint set and converting it into a convex maximization over a convex set with finite number of extreme points. Such a problem can be solved by enumerating the objective function on the extreme points of the constraint set and comparing the corresponding values. Moreover, we proved the convergence of the upper bound signal and derived sufficient conditions for eventually eliminating all false modes using our mode elimination algorithm. Finally, we demonstrated the effectiveness of our observer using an illustrative example.

Chapter 4

SIMULTANEOUS STATE AND UNKNOWN INPUT SET-VALUED OBSERVERS FOR NONLINEAR DYNAMICAL SYSTEMS

In this chapter [a] , we propose fixed-order set-valued observers for nonlinear bounded-error dynamical systems with unknown input signals that simultaneously find bounded sets of states and unknown inputs that include the true states and inputs. Sufficient conditions in the form of Linear Matrix Inequalities (LMIs) for the stability of the proposed observers are derived for general nonlinear systems and furthermore, less restrictive sufficient conditions are provided for three classes of nonlinear systems: (I) Linear Parameter-Varying (LPV), (II) Lipschitz continuous, and (III) Decremental Quadratic Constrained (DQC) systems. This includes a new DQC property that is at least as general as the incremental quadratic constrained property for nonlinear systems. In addition, we design the optimal $\mathcal{H}_\infty$ observer among those that satisfy the stability conditions, using semi-definite programs with additional LMIs constraints. Furthermore, sufficient conditions are provided for the upper bounds of the estimation errors to converge to steady state values and finally, the effectiveness of the proposed set-valued observers is demonstrated through illustrative examples, where we compare the performance of our observers with some existing observers.

---

[a] The content of this chapter is documented as a submitted and under review paper in [118].

4.1 Preliminary Material

4.1.1 Structural Properties

Here, we briefly introduce the structural properties that we will consider for our different classes of systems, so that we will be able to refer to them later when needed.

Definition 4.1.1 (Strong Detectability [131])**.** *The following bounded-error Linear Time Invariant (LTI) system:*

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Gd_k + w_k, \\ y_k &= Cx_k + Du_k + Hd_k + v_k, \end{aligned} \tag{4.1}$$

i.e., the tuple (A, G, C, H)*, is strongly detectable if* $y_k = 0 \ \forall \, k \geq 0$ *implies* $x_k \to 0$ *as* $k \to \infty$*, for all initial states and input sequences* $\{d_i\}_{i \in \mathbb{N}}$*, where* A, B, G, C, D, H *are known constant matrices with appropriate dimensions, and* x_k*,* u_k*,* y_k*,* d_k*,* w_k *and* v_k *are system state, known input, output, unknown input, bounded norm process noise and measurement noise signals, respectively.*

Remark 4.1.2. *Several necessary and sufficient rank conditions are provided in [131, Theorem 1] to check the strong detectability of system* (4.1)*, in other words,* (A, G, C, H)*, including*

$$\operatorname{rk} \mathcal{R}_S(z) \triangleq \operatorname{rk} \begin{bmatrix} zI - A & -G \\ C & H \end{bmatrix} = n + p, \forall z \in \mathbb{C}, |z| \geq 1.$$

It is worth mentioning that all the aforementioned conditions are equivalent to the system being minimum-phase (i.e., the invariant zeros of $\mathcal{R}_S(z)$ *are stable). Moreover, strong detectability implies that the pair* (A, C) *is detectable, and if* $l = p$*, then strong detectability implies that the pair* (A, G) *is stabilizable (cf. [131, Theorem 1] for more details).*

Definition 4.1.3 (Lipschitz Vector Fields)**.** *A vector field* $f(\cdot) : \mathcal{D}_f \to \mathbb{R}^m$ *is globally* L_f*-Lipschitz continuous on* $\mathcal{D}_f \subseteq \mathbb{R}^n$*, if there exists* $L_f \in \mathbb{R}_{++}$*, such that*

$$\|f(x_1) - f(x_2)\| \leq L_f \|x_1 - x_2\|, \forall x_1, x_2 \in D_f.$$

Definition 4.1.4 (LPV Functions)**.** *A vector field* $f(\cdot) : \mathbb{R}^p \to \mathbb{R}^q$ *is Linear Parameter-Varying (LPV), if at each time step* k*,* $f(x_k)$ *can be decomposed into a convex combination of linear functions with* known *coefficients, i.e.,* $\forall k \geq 0, \exists N \in \mathbb{N}$ *such that* $\forall i \in \{1, \ldots, N\}$*, there exist known* $\lambda_{i,k} \in [0,1]$ *and* $A^i \in \mathbb{R}^{p \times q}$ *such that* $\sum_{i=1}^{N} \lambda_{i,k} = 1$ *and* $f(x_k) = \sum_{i=1}^{N} \lambda_{i,k} A^i x_k$*. Each linear function* $A^i x$ *is called a* constituent *function of the original nonlinear function.*

Definition 4.1.5 (δ-QC Vector Fields [4])**.** *A symmetric matrix* $M \in \mathbb{R}^{(n_q+n_f)\times(n_q+n_f)}$ *is an incremental multiplier matrix (*δMM*) for* $f(\cdot)$ *if the following incremental quadratic constraint (*δ*-QC) is satisfied for all* $q_1, q_2 \in \mathbb{R}^{n_q}$*:*

$$\begin{bmatrix} (\Delta f)^\top & (\Delta q)^\top \end{bmatrix} M \begin{bmatrix} (\Delta f)^\top & (\Delta q)^\top \end{bmatrix}^\top \geq 0,$$

where $\Delta q \triangleq q_2 - q_1$ *and* $\Delta f \triangleq f(q_2) - f(q_1)$*.*

Next, we introduce a new class of systems we call decremental quadratic constrained (DQC) that is at least as general as δ-QC and includes a broad range of nonlinearities.

Definition 4.1.6 (*DQC* Functions)**.** *A vector field* $f(\cdot) : \mathbb{R}^p \to \mathbb{R}^q$ *is* $(\mathcal{M}, \gamma)$*-Decremental Quadratic Constrained (*$(\mathcal{M}, \gamma)$*-DQC), if there exist symmetric matrix* $\mathcal{M} \in \mathbb{R}^{(p+q)\times(p+q)}$ *and* $\gamma \in \mathbb{R}_+$ *such that*

$$\begin{bmatrix} (\Delta f)^\top & (\Delta x)^\top \end{bmatrix} \mathcal{M} \begin{bmatrix} (\Delta f)^\top & (\Delta x)^\top \end{bmatrix}^\top \leq \gamma, \tag{4.2}$$

for all $x_1, x_2 \in \mathbb{R}^p$*, where* $\Delta x \triangleq x_2 - x_1$ *and* $\Delta f \triangleq f(x_2) - f(x_1)$*. We call* $\mathcal{M}$ *a decremental multiplier matrix for function* $f(\cdot)$*.*

First of all, we show that a vector field may satisfy DQC property with different pairs of $(\mathcal{M}, \gamma)$'s. For clarity, all proofs are provided in the Appendix.

Proposition 4.1.7. *Suppose* $f(\cdot)$ *is* $(\mathcal{M}, \gamma)$*-DQC. Then it is also* $(\kappa\mathcal{M}, \kappa\gamma)$*-DQC,* $(\nu\mathcal{M}, \gamma)$*-DQC,* $(\mathcal{M}, \rho)$*-DQC and* $(\mathcal{M}', \gamma)$*-DQC for every* $\kappa \geq 0$*,* $0 \leq \nu \leq 1$*,* $\rho \geq \gamma$ *and* $\mathcal{M}' \preceq \mathcal{M}$*.*

Moreover, we next show that the DQC property includes Lipschitz continuity and is at least as general as the incremental quadratic constrained (δ-QC) property (cf. Definition 4.1.5), which recently has received considerable attention in nonlinear system state and input estimation (e.g., in [4, 22, 23]). Consequently, the class of DQC functions is a generalization of several types of nonlinearities (cf. Corollary 4.1.10).

Proposition 4.1.8. *Every globally* L_f*-Lipschitz continuous function is* δ*-QC with multiplier matrix* $M = \begin{bmatrix} -I & 0 \\ 0 & L_f^2 \end{bmatrix}$.

Proposition 4.1.9. *Every nonlinearity which is* δ*-QC with multiplier matrix* M *is* $(-M, \gamma)$*-DQC for any* $\gamma \geq 0$*.*

Corollary 4.1.10. *Lipschitz nonlinearities, incrementally sector bounded nonlinearities and nonlinearities with matrix parameterizations, etc., which are* δ*-QC (cf. Figure 4.1 and [4, Sections 5.1–5.2]), are also DQC (the reader is referred to [4, 22, 23] for definitions, demonstrations and more detailed examples).*

Next, we provide some instances of nonlinear DQC vector fields, that to our best knowledge, have not been shown to be δ-QC.

Example 4.1.11. *Consider any monotonically increasing vector-filed* $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$*, which is not necessarily globally Lipschitz. By monotonically increasing, we mean that*

$\Delta f^\top \Delta x \geq 0$, for all $x_1, x_2 \in \mathcal{D}_f$, where Δf and Δx are defined in Definition 4.1.6. As simple examples, the reader can consider $g(x) = x^5$ with $\mathcal{D}_g = \mathbb{R}$ or $h(x) = \tan(x)$ with $\mathcal{D}_h = (-\frac{\pi}{2}, \frac{\pi}{2})$. It can be easily validated that such functions are $(\mathcal{M}, \gamma)$-DQC with $\mathcal{M} = \begin{bmatrix} 0_{n\times n} & -I_{n\times n} \\ -I_{n\times n} & 0_{n\times n} \end{bmatrix}$ and any $\gamma \geq 0$. Similarly, any monotonically decreasing vector field is $(-\mathcal{M}, \gamma)$-DQC.

Example 4.1.12. *Now, consider $f(x) = x^2$ with $\mathcal{D}_f = [-\overline{x}, \overline{x}] \in \mathbb{R}$, $\overline{x} \geq 0.5$, which is not a monotone function. Let $\mathcal{M}_0 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. It can be verified that $\begin{bmatrix} (\Delta f)^\top & (\Delta x)^\top \end{bmatrix} \mathcal{M}_0 \begin{bmatrix} (\Delta f)^\top & (\Delta x)^\top \end{bmatrix}^\top = \|\Delta f - \Delta x\|^2 = \|(\Delta x)^2 - \Delta x\|^2 \leq [2\overline{x}(2\overline{x} + 1)]^2 = 9$, for $x_1, x_2 \in \mathcal{D}_f$. Hence, $f(x) = x^2$ for all $x \in [-\overline{x}, \overline{x}] \in \mathbb{R}$ with $\overline{x} \geq 0.5$ is $(\mathcal{M}_0, 9)$-DQC.*

Furthermore, considering a specific structure for the decremental multiplier matrix $\mathcal{M}$, we introduce a new class of functions that is a subset of the DQC class.

Definition 4.1.13 (DQC* Functions)**.** *A vector field $f(\cdot)$ is a DQC* function, if it is $(\mathcal{M}, \gamma)$-DQC for some known $\mathcal{M} \in \mathbb{R}^{2n\times 2n}$ and $\gamma \geq 0$, and there exists a known $\mathcal{A} \in \mathbb{R}^{n\times n}$, such that $\begin{bmatrix} I_{n\times n} & -\mathcal{A} \\ -\mathcal{A}^\top & \mathcal{A}^\top \mathcal{A} \end{bmatrix} \preceq \mathcal{M}$.*

Now we present some results that establish the relationships between the aforementioned classes of nonlinearities.

Proposition 4.1.14. *Suppose $f(\cdot)$ is globally L_f-Lipschitz continuous and the state space, $\mathcal{X}$, is bounded, i.e., there exists $r \in \mathbb{R}_+$ such that for all $x \in \mathcal{X}$, $\|x\| \leq r$. Then, $f(\cdot)$ is a DQC* function with $\mathcal{A} = I_{n\times n}$, $\begin{bmatrix} I_{n\times n} & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{bmatrix} = \mathcal{M}$ and $\gamma = 4r^2 L_f^2$.*

Lemma 4.1.15. *Suppose vector field* $f(\cdot)$ *can be decomposed as the sum of an affine and a bounded nonlinear function* $g(.)$*, i.e.,* $f(x) = Ax + h + g(x)$*, where* $A \in \mathbb{R}^{n\times n}$*,* $h \in \mathbb{R}^n$ *and* $\|g(x)\| \leq r \in \mathbb{R}_+$ *for all* $x \in \mathcal{D}_g$*. Then,* $f(\cdot)$ *is a DQC* function with* $\mathcal{A} = A$*,* $\mathcal{M} = \begin{bmatrix} I_{n\times n} & -\mathcal{A} \\ -\mathcal{A}^\top & \mathcal{A}^\top\mathcal{A} \end{bmatrix}$ *and any* $\gamma \geq (2r)^2$*.*

Note that some DQC systems are also DQC*. The following Proposition 4.1.16 helps with finding such an $\mathcal{A}$ for some specific structures of $\mathcal{M}$.

Proposition 4.1.16. *Suppose* $f(\cdot) : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ *is a* $(\mathcal{M}, \gamma)$*-DQC vector field, with* $\mathcal{M} = \begin{bmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{12}^\top & \mathcal{M}_{22} \end{bmatrix}$*, where* $\mathcal{M}_{11}, \mathcal{M}_{12}, \mathcal{M}_{22} \in \mathbb{R}^{n\times n}$*,* $\mathcal{M}_{11} - I_{n\times n} \succeq 0$ *and* $\mathcal{M}_{22} - \mathcal{M}_{12}^\top\mathcal{M}_{12} \succeq 0$*. Then,* $f(\cdot)$ *is a DQC* function with* $\mathcal{A} = -\mathcal{M}_{12}$*.*

The reader can verify that such sufficient conditions in Proposition 4.1.16 hold for the function in Example 4.1.12.

Proposition 4.1.17. *Every LPV function* $f(\cdot)$ *with constituent matrices* $A^i, \forall i \in 1\ldots N$*, is* $\|A^m\|$*-globally Lipschitz continuous, where* $\|A^m\| = \max_{i\in 1\ldots N} \|A^i\|$*.*

Corollary 4.1.18. *As a direct corollary of Propositions 4.1.14 and 4.1.17, any bounded domain LPV function is a DQC* function.*

Figure 4.1 summarizes all the above results on the relationships between several classes of nonlinearities [b] . We end this section with restating a result from [126], that will be used frequently later in deriving some of our main results.

---

[b] Lipschitz, LPV, δ-QC, DQC and DQC* nonlinearities are defined in Definitions 4.1.3–4.1.13. Incrementally sector bounded nonlinearities can be characterized by four fixed matrices K_{11}, K_{12}, K_{21}, and K_{22}, and a set of matrices, $\mathcal{X}$. In particular, they satisfy $(K_{11}\Delta x + K_{12}\Delta f)^\top X(K_{21}\Delta x + K_{22}\Delta f) \geq 0$, $\forall X \in \mathcal{X}$. Matrix parametrized nonlinearities can be characterized by some known set of matrices, $\Im$. Specifically, for any Δx and corresponding Δf, there exists a $\Theta \in \Im$ such that $\Delta f = \Theta\Delta x$. The reader is referred to [4, Sections 5.1 and 5.2] for detailed discussions about these two classes of nonlinearities, which are omitted here for the sake of brevity.

Lemma 4.1.19. *[126, Lemma 2.2] Let D, S and F be real matrices of appropriate dimensions and $F^\top F \preceq I$. Then, for any scalar $\epsilon > 0$ and $x, y \in \mathbb{R}^n$,*

$$2x^\top DFSy \leq \epsilon^{-1} x^\top DD^\top x + \epsilon y^\top S^\top Sy.$$

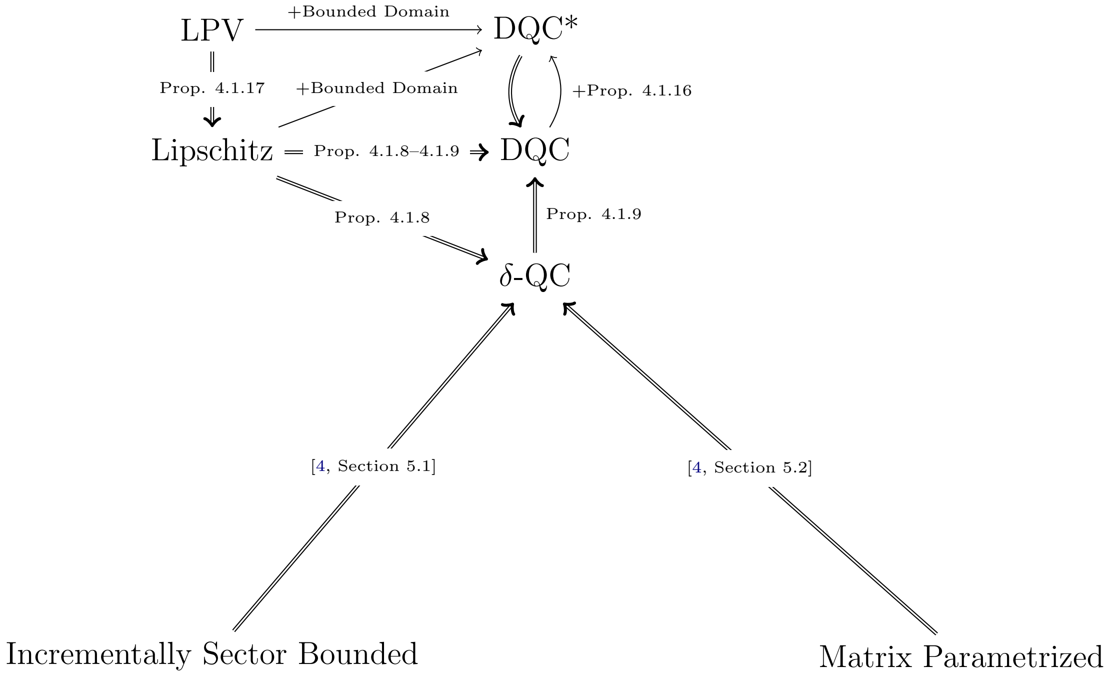


Figure 4.1: Relationships Between Different Classes of Nonlinearities: ⇒ Denotes Direct Implication, While → Denotes Implication with Addition Assumptions

4.2 Problem Statement

In this section, we describe the system, vector field and unknown input signal assumptions as well as formally state the observer design problem.

System Assumptions. Consider the nonlinear discrete-time bounded-error system

$$\begin{aligned} x_{k+1} &= f(x_k) + Bu_k + Gd_k + Ww_k, \\ y_k &= Cx_k + Du_k + Hd_k + v_k, \end{aligned} \tag{4.3}$$

where $x_k \in \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$, $u_k \in \mathbb{R}^m$ is a known input vector, $d_k \in \mathbb{R}^p$ is an unknown input vector, and $y_k \in \mathbb{R}^l$ is the measurement vector. The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^l$ are assumed to be bounded, with $\|w_k\| \leq \eta_w$ and $\|v_k\| \leq \eta_v$ (thus, they are ℓ_∞ sequences). We also assume an estimate $\hat{x}_0$ of the initial state x_0 is available, where $\|\hat{x}_0 - x_0\| \leq \delta_0^x$. The vector field $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and matrices B, C, D, G, W and H are known and of appropriate dimensions, where G and H are matrices that encode the *locations* through which the unknown input or attack signal can affect the system dynamics and measurements. Note that no assumption is made on H to be either the zero matrix (no direct feedthrough), or to have full column rank when there is direct feedthrough. Without loss of generality, we assume that $\text{rk}[G^\top \ H^\top] = p$, $n \geq l \geq 1$, $l \geq p \geq 0$ and $m \geq 0$.

Vector Field Assumptions. Here, we formally state the classes of nonlinear systems, related to the assumptions about the nonlinear vector field $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n = \begin{bmatrix} f_1^\top(.) & \dots & f_j^\top(.) & \dots & f_n^\top(.) \end{bmatrix}^\top$ $\forall j \in \{1, \dots, n\}$, that we consider in this paper.

Class 0. *Nonlinear systems without any additional assumptions.*

For this general case of Class 0 systems, we expect to derive conservative sufficient conditions for stability and optimality of the designed observers. However, to enable the computation of upper bounds for the estimation errors, we need some assumptions on the variations of the vector field in terms of state variations.

Class I. *Globally L_f-Lipschitz continuous systems.*

Class II. *DQC* systems, with some known $\mathcal{M} \in \mathbb{R}^{2n \times 2n}$, $\gamma \geq 0$, and $\mathcal{A} \in \mathbb{R}^{n \times n}$.*

Class III. *LPV systems with constituent matrices $A^i \in \mathbb{R}^{n \times n}, \forall i \in \{1, \dots, N\}$.*

For Class III of systems, the system dynamics is governed by an LPV system with known parameters at run-time. We call each tuple $(A^i, C, C, H), \forall i \in \{1 \dots N\}$, an LTI constituent of system (4.3).

Unknown Input (or Attack) Signal Assumptions. The unknown inputs d_k are not constrained to be a signal of any type (random or strategic) nor to follow any model, thus no prior 'useful' knowledge of the dynamics of d_k is available (independent of $\{d_\ell\}$ $\forall k \neq \ell$, $\{w_\ell\}$ and $\{v_\ell\}$ $\forall \ell$). We also do not assume that d_k is bounded or has known bounds and thus, d_k is suitable for representing adversarial attack signals.

The simultaneous input and state set-valued observer design problem is twofold and can be stated as follows:

Problem 4.2.1. *Given the nonlinear discrete-time bounded-error system with unknown inputs* (4.3)*,*

1) *Design stable observers that simultaneously find bounded sets of compatible states and unknown inputs for the four classes of nonlinear systems.*
2) *Among the observers that satisfy 1, find the optimal observer in the minimum $\mathcal{H}_\infty$-norm sense, i.e., with minimum average power amplification.*

4.3 More General Nonlinear System Model

Note that the proposed observer in this paper can also apply to more general nonlinear systems than (4.3) with some minor modifications; however, throughout the paper, we will mainly focus on the system model given by (4.3) for the sake of notation simplicity and understandability. In particular, we can generalize our framework to the following nonlinear discrete-time time-varying dynamical system:

$$\begin{aligned} x_{k+1} &= f_k(x_k) + B_k u_k + \hat{G} g_k(x_k, u_k, d_k^s) + W w_k, \\ y_k &= C x_k + D_k u_k + \hat{H} h_k(x_k, u_k, d_k^o) + v_k, \end{aligned} \tag{4.4}$$

where $f_k(.) : \mathbb{R}^n \to \mathbb{R}^n$ is a known *time-varying* nonlinear function, $d_k^s \in \mathbb{R}_s^p$ and $d_k^o \in \mathbb{R}_o^p$ can be interpreted as arbitrary (and *different*) unknown inputs effecting the state and observation equations through the known *time-varying nonlinear* vector fields $g_k(.,.) : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{p_s} \to \mathbb{R}^{n_{\hat{G}}}$ and $h_k(.,.) : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{p_o} \to \mathbb{R}^{n_{\hat{H}}}$, respectively and $\hat{G} \in \mathbb{R}^{n \times n_{\hat{G}}}$ and $\hat{H} \in \mathbb{R}^{l \times n_{\hat{H}}}$ are known matrices. Moreover, $B_k \in \mathbb{R}^{n \times m}$ and $D_k \in \mathbb{R}^{l \times m}$ are *time-varying* known matrices.

Courtesy of the fact that the unknown input signal in (4.3) can be completely arbitrary, we can lump the nonlinear functions with the unknown inputs in (4.4) into a newly defined unknown input signal to obtain an equivalent representation of the system (4.4) in the form of (4.3), where instead of vector field f, matrices G, H and unknown input d_k, we have f_k, $\check{G} \triangleq \begin{bmatrix} \hat{G} & 0_{n \times n_{\hat{H}}} \end{bmatrix}$, $\check{H} \triangleq \begin{bmatrix} 0_{l \times n_{\hat{G}}} & \hat{H} \end{bmatrix}$ and $\check{d}_k \triangleq \begin{bmatrix} g_k(x_k, u_k, d_k^s) \\ h_k(x_k, u_k, d_k^o) \end{bmatrix}$, correspondingly. Moreover, to deal with the time-varying f_k, a slight modification to the definition of the function increment can be considered, i.e., with $\Delta f_k \triangleq f_k(x_2) - f_{k-1}(x_1)$ for all x_1, x_2 instead of the time-invariant $\Delta f \triangleq f(x_2) - f(x_1)$ within the structural properties and assumptions in Section 4.1.1 [c] . Furthermore, time-varying B_k and D_k do not change the design process and results, since they cancel out during the procedure (cf. Section 4.5). Therefore, all our results still hold for the state estimates of the more general system (4.4), using the *new* matrices $\check{G}$ and $\check{H}$. As for the unknown inputs, where the proposed observer returns set-valued estimates for $\check{d}_k \triangleq \begin{bmatrix} g_k(x_k, u_k, d_k^s) \\ h_k(x_k, u_k, d_k^o) \end{bmatrix}$, we can apply any *pre-image* set computation techniques in the literature such as [87, 106, 24] to find set estimates for d_k^s and d_k^o using the set-valued estimate for x_k and the known u_k.

4.4 Fixed-Order Simultaneous Input and State Set-Valued Observer Framework

---

[c] For instance, a time-varying vector field $f_k(.)$ is Lipschitz if there exists $L_f \in \mathbb{R}_{++}$ such that for any time step k, $\|f_k(x_1) - f_{k-1}(x_2)\| \leq L_f \|x_1 - x_2\|$, $\forall x_1, x_2 \in D_{f_k}$ and similarly for the LPV, δ-QC, DQC and DQC* properties.

In this paper, we propose recursive *set-valued* observers that consist of three steps: (1) an *unknown input estimation* step that returns the set of compatible unknown inputs using the current measurement and the set of compatible states, (2) a *time update* step in which the compatible set of states is propagated based on the system dynamics, and (3) a *measurement update* step where the set of compatible states is updated according to the current measurement. Since the complexity of optimal observers increases with time, we will only focus on *fixed-order* recursive filters, similar to [17, 26, 131], and in particular, we consider *set-valued* estimates of the form:

$$\begin{aligned}\hat{D}_{k-1} &= \{d \in \mathbb{R}^p : \|d_{k-1} - \hat{d}_{k-1}\| \leq \delta^d_{k-1}\},\\ \hat{X}^\star_k &= \{x \in \mathbb{R}^n : \|x_k - \hat{x}^\star_{k|k}\| \leq \delta^{x,\star}_k\},\\ \hat{X}_k &= \{x \in \mathbb{R}^n : \|x_k - \hat{x}_{k|k}\| \leq \delta^x_k\},\end{aligned}$$

where $\hat{D}_{k-1}$, $\hat{X}^\star_k$ and $\hat{X}_k$ are the sets of compatible unknown inputs at time $k-1$, propagated, and updated states at time k, correspondingly. In other words, we restrict the estimation errors to balls of norm δ. In this setting, the observer design problem is equivalent to finding the centroids $\hat{d}_{k-1}$, $\hat{x}^\star_{k|k}$ and $\hat{x}_{k|k}$ as well as the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k of the sets $\hat{D}_{k-1}$, $\hat{X}^\star_k$ and $\hat{X}_k$, respectively. In addition, we limit our attention to observers for the centroids $\hat{d}_{k-1}$, $\hat{x}^\star_{k|k}$ and $\hat{x}_{k|k}$ that belong to the class of three-step recursive filters given in [47] and [135], with $\hat{x}_{0|0} = \hat{x}_0$.

4.4.1 System Transformation

To aid the observer design, we first carry out a transformation to decompose the unknown input signal d_k into two components $d_{1,k}$ and $d_{2,k}$, as well as to decouple the output equation into two components, $z_{1,k}$ and $z_{2,k}$, one with a full rank direct feedthrough matrix and the other without direct feedthrough, as follows:

$$\begin{aligned} x_{k+1} &= f(x_k) + Bu_k + G_1 d_{1,k} + G_2 d_{2,k} + W w_k, \\ z_{1,k} &= C_1 x_k + \Sigma d_{1,k} + D_1 u_k + v_{1,k}, \\ z_{2,k} &= C_2 x_k + D_2 u_k + v_{2,k}. \end{aligned} \tag{4.5}$$

For the sake of increasing readability and completeness, the reader is referred to Section 2.2.1 for details of this similarity transformation, where the transformed system matrices $G_1, G_2, C_1, C_2, D_1, D_2$ and noise signals $v_{1,k}, v_{2,k}$ are defined.

Remark 4.4.1. *It is important to note that $d_{2,k}$ cannot be estimated from y_k since it does not affect $z_{1,k}$ and $z_{2,k}$. Thus, in light of* (4.5)*, we can only obtain a (one-step) delayed estimate of $\hat{D}_{k-1}$. The reader may refer to [133] for a complete discussion on when a delay is absent or when we can expect further delays.*

4.4.2 Observer Structure

Using the above transformation, we propose the following three-step recursive observer structure to compute the state and input estimate sets:

Unknown Input Estimation (UIE):

$$\hat{d}_{1,k} = M_1(z_{1,k} - C_1 \hat{x}_{k|k} - D_1 u_k), \tag{4.6}$$

$$\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2 \hat{x}_{k|k-1} - D_2 u_k), \tag{4.7}$$

$$\hat{d}_{k-1} = V_1 \hat{d}_{1,k-1} + V_2 \hat{d}_{2,k-1}. \tag{4.8}$$

Time Update (TU):

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}) + Bu_{k-1} + G_1 \hat{d}_{1,k-1}, \tag{4.9}$$

$$\hat{x}^{\star}_{k|k} = \hat{x}_{k|k-1} + G_2 \hat{d}_{2,k-1}. \tag{4.10}$$

Measurement Update (MU):

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}^{\star}_{k|k} + L(y_k - C\hat{x}^{\star}_{k|k} - Du_k) \\ &= \hat{x}^{\star}_{k|k} + \tilde{L}(z_{2,k} - C_2\hat{x}^{\star}_{k|k} - D_2u_k), \end{aligned} \tag{4.11}$$

where $L \in \mathbb{R}^{n\times l}$, $\tilde{L} \triangleq LU_2 \in \mathbb{R}^{n\times(l-p_H)}$, $M_1 \in \mathbb{R}^{p_H\times p_H}$ and $M_2 \in \mathbb{R}^{(p-p_H)\times(l-p_H)}$ are observer gain matrices that are designed according to Lemma 4.5.1 and Theorem 4.5.8, to minimize the "volume" of the set of compatible states and unknown inputs, quantified by the radii δ^d_{k-1}, $\delta^{x,\star}_k$ and δ^x_k. Note also that we applied $L = LU_2U_2^\top = \tilde{L}U_2^\top$ from Lemma 4.5.1 into (4.11), where U_2 is defined in Section 2.2.1.

4.5 Observer Design and Analysis

In this section we derive LMI conditions for designing observers that are stable, i.e., the estimation errors are uniformly bounded (Section 4.5.1) and optimal in the $\mathcal{H}_\infty$ sense (Section 4.5.2). Moreover, we derive the resulting radii of the state and input estimates (Section 4.5.3). To do so, first, we will derive our observer error dynamics through the following Lemma 4.5.1. For conciseness, all proofs are provided in the Appendix.

Lemma 4.5.1. *Consider system* (4.3) *and the observer* (4.6)-(4.11)*. Suppose* $\text{rk}(C_2G_2) = p - p_H$*, where* C_2 *and* G_2 *are given in Section 2.2.1. Then, designing observer matrix gains as* $M_1 = \Sigma^{-1}$*,* $M_2 = (C_2G_2)^\dagger$*,* $LU_1 = 0$ *and* $L = LU_2U_2^\top = \tilde{L}U_2^\top$*, with* U_1 *and* U_2 *given in Section 2.2.1, yields* $M_1\Sigma = I$ *and* $M_2C_2G_2 = I$*, and leads to the following difference equation for the state estimation error dynamics (i.e., the dynamics of* $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$*):*

$$\tilde{x}_{k+1|k+1} = (I - \tilde{L}C_2)\Phi(\Delta f_k - \Psi\tilde{x}_{k|k}) + \mathcal{W}(\tilde{L})\overline{w}_k, \tag{4.12}$$

where

$$\begin{aligned}
\Delta f_k &\triangleq f(x_k) - f(\hat{x}_k), \Phi \triangleq I - G_2 M_2 C_2, \\
\overline{w}_k &\triangleq \begin{bmatrix} (\frac{1}{\sqrt{2}}) v_k^\top & w_k^\top & (\frac{1}{\sqrt{2}}) v_{k+1}^\top \end{bmatrix}^\top, \\
R &\triangleq \begin{bmatrix} -\sqrt{2}\Phi G_1 M_1 T_1 & -\Phi W & -\sqrt{2} G_2 M_2 T_2 \end{bmatrix}, \\
Q &\triangleq \begin{bmatrix} 0_{(l-p_H)\times l} & 0_{(l-p_H)\times n} & -\sqrt{2} T_2 \end{bmatrix}, \\
\Psi &\triangleq G_1 M_1 C_1, \quad \mathcal{W}(\tilde{L}) \triangleq (I - \tilde{L} C_2) R + \tilde{L} Q.
\end{aligned}$$

Note that $\overline{w}_k$ is chosen such that $\lim\limits_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \overline{w}_i^\top \overline{w}_i = \lim\limits_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} w_i^\top w_i + v_i^\top v_i$. The result in (4.12) shows that we successfully decoupled/canceled out d_k from the error dynamics, otherwise there would be a potentially unbounded and unknown term in the error dynamics.

4.5.1 Stable Observer Design

We first study the stability of the observer in the sense of Lyapunov. For the sake of clarity, we first formally define the considered notion of stability.

Definition 4.5.2. *[Lyapunov Stability] A simultaneous state and input set-valued observer is Lyapunov stable, if its estimation error norm sequences* $\{\|\tilde{x}_{k|k}\| \triangleq \|x_k - \hat{x}_{k|k}\|, \|\tilde{d}_{k-1|k-1}\| \triangleq \|d_{k-1} - \hat{d}_{k-1}\|\}_{k=1}^{\infty}$ *are uniformly bounded.*

Now, we are ready to provide our first set of main results on sufficient conditions for bounded-error stability of the observer (4.6)–(4.11), by supposing for the moment that there is no exogenous bounded noise w_k and v_k.

Theorem 4.5.3 (Observer Stability)**.** *Consider system* (4.3) *and the observer* (4.6)–(4.11)*. Suppose there is no bounded noise* w_k *and* v_k *and all the conditions in Lemma 4.5.1 hold. Then, the observer error dynamics is Lyapunov stable, if there exist* *matrices* $0 \prec P \in \mathbb{R}^{n\times n}$, $Y \in \mathbb{R}^{n\times(l-p_H)}$ *and* $0 \prec \Gamma \in \mathbb{R}^{(l-p_H)\times(l-p_H)}$, *such that the following LMIs hold:*

$$\Pi \triangleq \varrho \begin{bmatrix} I-\Gamma & 0 & 0 \\ 0 & \Gamma & Y^\top \\ 0 & Y & P \end{bmatrix} \succeq 0, \Upsilon^i \triangleq \begin{bmatrix} \Theta & \Lambda^i \\ {\Lambda^i}^\top & \Xi \end{bmatrix} \succeq 0,$$
$$\forall i \in \{1\dots N\}, \tag{4.13}$$

where $\varrho \in \{0,1\}$, Θ, Λ^i, Ξ *and* N *are defined for different cases as follows:*

0. *If* $f(\cdot)$ *is a Class 0 function, then* $\varrho \triangleq 1$, $N \triangleq 1$ *and*

$$\begin{aligned} \Theta &\triangleq \Phi^\top(F - C_2^\top C_2)\Phi, \\ \Lambda^i &\triangleq \Phi^\top(P - C_2^\top Y^\top - YC_2)\Phi\Psi, \\ \Xi &\triangleq P + \Psi^\top\Phi^\top F\Phi\Psi - \Phi^\top C_2^\top C_2\Phi. \end{aligned} \tag{4.14}$$

I. *If* $f(\cdot)$ *is a Class I function, then* $\varrho \triangleq 1$, $N \triangleq 1$ *and*

$$\begin{aligned} \Theta &\triangleq \Phi^\top(F - C_2^\top C_2)\Phi + I, \\ \Lambda^i &\triangleq \Phi^\top(P - C_2^\top Y^\top - YC_2)\Phi\Psi, \\ \Xi &\triangleq P + \Psi^\top\Phi^\top F\Phi\Psi - \Phi^\top C_2^\top C_2\Phi - L_f^2 I. \end{aligned} \tag{4.15}$$

II. *If* $f(\cdot)$ *is a Class II function, then* $\varrho \triangleq 1$, $N \triangleq 1$ *and*

$$\begin{aligned} \Theta &\triangleq \Phi^\top(F - C_2^\top C_2)\Phi, \\ \Lambda^i &\triangleq \Phi^\top(P - C_2^\top Y^\top - YC_2)\Phi(\Psi - \mathcal{A}), \\ \Xi &\triangleq P + (\Psi - \mathcal{A})^\top\Phi^\top F\Phi(\Psi - \mathcal{A}) - \Phi^\top C_2^\top C_2\Phi. \end{aligned} \tag{4.16}$$

III. *If* $f(\cdot)$ *is a Class III function, then* $\varrho \triangleq 0$, N *is the number of* constituent *LTI systems and* $\forall i \in \{1\dots N\}$*:*

$$\Theta \triangleq P, \Xi \triangleq P, \Lambda^i \triangleq (A^i - \Psi)^\top\Phi^\top(P - C_2^\top Y^\top), \tag{4.17}$$

with $F \triangleq YC_2 + C_2^\top Y^\top - P - C_2^\top \Gamma C_2$*. Moreover, the corresponding observer gain for all four classes can be obtained as* $\tilde{L} = P^{-1}Y$*.*

It is worth mentioning that for Class I functions, i.e., when the nonlinear vector field $f(\cdot)$ is globally Lipschitz continuous, stability of the observer can also be demonstrated using more succinct LMIs in the following Lemma 4.5.4.

Lemma 4.5.4 (Alternative LMIs (Class I))**.** *Consider system* (4.3) *and the observer* (4.6)–(4.11)*. Suppose all the conditions in Lemma 4.5.1 hold,* $f(\cdot)$ *is a Class I function and there is no bounded noise* w_k *and* v_k*. Then, the observer error dynamics is Lyapunov stable with the observer gain* $\tilde{L} = P^{-1}Y$*, if* $\exists\ 0 \prec P \in \mathbb{R}^{n\times n}$ *and* $Y \in \mathbb{R}^{n\times(l-p_H)}$ *such that*

$$\begin{bmatrix} I & (P-YC_2) & 0 \\ (P-YC_2) & P & 0 \\ 0 & 0 & \Delta \end{bmatrix} \succeq 0, \tag{4.18}$$

where $\Delta \triangleq P - 2L_f^2 \lambda_{\max}(\Phi^\top\Phi)I - 2\Psi^\top\Phi^\top\Phi\Psi$*, with* Φ *and* Ψ *defined in Lemma 4.5.1.*

Moreover, if $f(\cdot)$ is a Class III function, then we can provide necessary and sufficient conditions for the existence of stable observers. The necessary conditions are conveniently testable. They are also beneficial in the sense that if they are *not* satisfied, the designer knows *a priori* that there does not exist any $\mathcal{H}_\infty$-observer for such modified systems with unknown inputs/attacks. The conditions are formally derived in the following Lemma 4.5.5.

Lemma 4.5.5 (Existence of Stable Observers)**.** *Suppose* $f(\cdot)$ *is a Class III function and all the conditions in Lemma 4.5.1 hold. Then, there exists a stable observer for the system* (4.3)*, with any sequence* $\{\lambda_{i,k}\}_{k=0}^{\infty}$ *for all* $i \in \{1,2,\ldots,N\}$ *that satisfies* $0 \le \lambda_{i,k} \le 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$*, if* $(\overline{A}_k, C_2)$ *be uniformly detectable* [d] *for each* k*, and*

---

[d] The readers are referred to [10, Section 2] for the concise definition of uniform detectability. A spectral test can be found in [92].

only if all constituent LTI systems $(A^i, G, C, H), \forall i \in \{1 \dots N\}$, *are strongly detectable (cf. Definition 4.1.1), where* $\overline{A}_k \triangleq \Phi(\sum_{i=1}^{N} \lambda_{i,k} A^i - \Psi)$, *with* Φ *and* Ψ *defined in Lemma 4.5.1.*

Corollary 4.5.6. *There exists a stable simultaneous state and input set-valued observer for the LTI system* (4.1)*, through* (4.6)–(4.11)*, if and only if the tuple* (A, G, C, H) *is strongly detectable and only if* $\mathrm{rk}(C_2 G_2) = p - p_H$*. Moreover, the observer gain matrices can be designed as* $M_1 = \Sigma^{-1}$, $M_2 = (C_2 G_2)^\dagger$ *and* $L = \tilde{L} U_2^\top$ *and* $\tilde{L} = P^{-1} Y$*, where* $P \succ 0$ *and* Y *solve the following feasibility program with LMI constraints:*

$$\begin{aligned} &\text{Find } (P \succ 0, Y) \\ &s.t. \quad \begin{bmatrix} P & \Lambda \\ \Lambda^\top & P \end{bmatrix} \succeq 0, \end{aligned}$$

with $\Lambda \triangleq (A - \Psi)^\top \Phi^\top (P - C_2^\top Y^\top)$ *and* Φ *and* Ψ *defined in Lemma 4.5.1.*

4.5.2 $\mathcal{H}_\infty$ Observer Design

The goal of this section is to provide additional sufficient conditions to guarantee optimality of the observers in the $\mathcal{H}_\infty$ sense. We first define our considered notion of optimality via the following Definition 4.5.7.

Definition 4.5.7 ($\mathcal{H}_\infty$-Observer)**.** *Let* $T_{\tilde{x},w,v}$ *denote the transfer function matrix that maps the noise signals* $\vec{w}_k \triangleq \begin{bmatrix} w_k^\top & v_k^\top \end{bmatrix}^\top$ *to the updated state estimation error* $\tilde{x}_{k|k} \triangleq x_k - \hat{x}_{k|k}$*. For a given "noise attenuation level"* $\eta \in \mathbb{R}_+$*, the observer performance satisfies* $\mathcal{H}_\infty$ *norm bounded by* η*, if* $\|T_{\tilde{x},w,v}\|_\infty \leq \eta$*, i.e., the maximum average signal power amplification is upper-bounded by* η^2*:*

$$\frac{\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \tilde{x}_{i|i}^\top \tilde{x}_{i|i}}{\lim_{k\to\infty} \frac{1}{k+1} \sum_{i=0}^{k} \vec{w}_i^\top \vec{w}_i} \triangleq \|T_{\tilde{x},w,v}\|_\infty^2 \leq \eta^2. \tag{4.19}$$

Now we present our second set of main results, on designing stable and optimal observers in the minimum $\mathcal{H}_\infty$ sense.

Theorem 4.5.8 ($\mathcal{H}_\infty$-Observer Design)**.** *Consider system* (4.3)*, the observer* (4.6)–(4.11) *and given* $\eta > 0$*. Suppose all the conditions in Theorem 4.5.3 hold, consider* Φ*,* Ψ*,* Q *and* R *defined in Lemma 4.5.1 and let* $\Omega \triangleq C_2 R - Q$*. Then, with the gain* $\tilde{L} = P^{-1}Y$*, we obtain a stable observer with* $\mathcal{H}_\infty$ *norm bounded by* η*, if* (4.13) *holds and*

$$\mathcal{N} \triangleq \begin{bmatrix} \mathcal{N}_{11} & * & * & * \\ \mathcal{N}_{21}^i & \mathcal{N}_{22} & * & * \\ \mathcal{N}_{31} & \mathcal{N}_{32} & \mathcal{N}_{33} & * \\ \mathcal{N}_{41} & \mathcal{N}_{42} & \mathcal{N}_{43} & \mathcal{N}_{44} \end{bmatrix} \succeq 0, \forall i \in \{1 \dots N\}, \tag{4.20}$$

where $\mathcal{N}_{11}$*,* $\mathcal{N}_{21}^i$*,* $\mathcal{N}_{22}$*,* $\mathcal{N}_{31}$*,* $\mathcal{N}_{32}$*,* $\mathcal{N}_{33}$*,* $\mathcal{N}_{41}$*,* $\mathcal{N}_{42}$*,* $\mathcal{N}_{43}$*,* $\mathcal{N}_{44}$ *and* N *are defined for different cases as follows:*

0. *If* $f(\cdot)$ *is a Class 0 function, then* $N \triangleq 1$ *and*

$$\begin{aligned} &\mathcal{N}_{11} \triangleq \eta^2 I + R^\top Y \Omega + \Omega^\top Y^\top R - R^\top P R - \Omega^\top (\Gamma + 2I)\Omega \\ &\mathcal{N}_{21}^i \triangleq \Psi^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R), \\ &\mathcal{N}_{22} \triangleq -I - \Psi^\top \Phi^\top C_2^\top C_2 \Phi \Psi, \mathcal{N}_{44} \triangleq I, \\ &\mathcal{N}_{31} \triangleq Y\Omega + C_2^\top Y^\top R - PR, \mathcal{N}_{33} \triangleq -\Phi^\top C_2^\top C_2 \Phi, \end{aligned} \tag{4.21}$$

   and $\mathcal{N}_{32}$*,* $\mathcal{N}_{41}$*,* $\mathcal{N}_{42}$*,* $\mathcal{N}_{43}$ *are zero matrices with appropriate dimensions.*

I. *If* $f(\cdot)$ *is a Class I function, then* $N \triangleq 1$*,*

$$\begin{aligned} &\mathcal{N}_{11} \triangleq \eta^2 I + R^\top Y \Omega + \Omega^\top Y^\top R - R^\top P R - \Omega^\top (\Gamma + 2I)\Omega, \\ &\mathcal{N}_{21}^i \triangleq \Psi^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R), \\ &\mathcal{N}_{22} \triangleq -I - \Psi^\top \Phi^\top C_2^\top C_2 \Phi \Psi - L_f^2 \lambda_{max}(\Phi^\top C_2^\top C_2 \Phi) I, \\ &\mathcal{N}_{31} \triangleq Y\Omega + C_2^\top Y^\top R - PR, \mathcal{N}_{33} \triangleq 0, \mathcal{N}_{44} \triangleq I, \end{aligned} \tag{4.22}$$

and $\mathcal{N}_{32}$*,* $\mathcal{N}_{41}$*,* $\mathcal{N}_{42}$*,* $\mathcal{N}_{43}$ *are zero matrices with appropriate dimensions.*

II. If $f(\cdot)$ *is a Class II function, then* $N \triangleq 1$*,*

$$\begin{aligned}
&\mathcal{N}_{11} \triangleq \eta^2 I + R^\top Y \Omega + \Omega^\top Y^\top R - R^\top P R - \Omega^\top (\Gamma + 2I)\Omega,\\
&\mathcal{N}_{21}^i \triangleq -(\mathcal{A} - \Psi)^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R),\\
&\mathcal{N}_{22} \triangleq -I - (\mathcal{A} - \Psi)^\top \Phi^\top C_2^\top C_2 \Phi (\mathcal{A} - \Psi), \mathcal{N}_{44} \triangleq I,\\
&\mathcal{N}_{31} \triangleq Y\Omega + C_2^\top Y^\top R - PR, \mathcal{N}_{33} \triangleq -\Phi^\top C_2^\top C_2 \Phi,
\end{aligned} \tag{4.23}$$

and $\mathcal{N}_{32}$*,* $\mathcal{N}_{41}$*,* $\mathcal{N}_{42}$*,* $\mathcal{N}_{43}$ *are zero matrices with appropriate dimensions.*

III. If $f(\cdot)$ *is a Class III function, then* N *is the number of* constituent *LTI systems,*

$$\begin{aligned}
&\mathcal{N}_{11} \triangleq \mathcal{N}_{22} \triangleq P, \mathcal{N}_{32} \triangleq ((P - YC_2)R + YQ)^\top,\\
&\mathcal{N}_{21}^i \triangleq (P - YC_2)\Phi(A^i - \Psi), \mathcal{N}_{44} \triangleq \eta^2 I,\\
&\mathcal{N}_{41} \triangleq I, \mathcal{N}_{33} \triangleq \eta^2 I,
\end{aligned} \tag{4.24}$$

and $\mathcal{N}_{31}$*,* $\mathcal{N}_{42}$ *and* $\mathcal{N}_{43}$ *are zero matrices with appropriate dimensions.*

Finally, the minimum $\mathcal{H}_\infty$ *bound can be found by solving the following semi-definite program with LMI constraints:*

$$\begin{aligned}
(\eta^\star)^2 = &\min_{P \succ 0, \Gamma \succ 0, Y, \eta^2 > 0} \eta^2\\
&\quad s.t. \quad (4.13), (4.20) \text{ hold},
\end{aligned}$$

where η^2 *is a decision variable. Solving this Semi-Definite Program (SDP), we have* $\|T_{\tilde{x},w,v}\|_\infty \leq \eta^\star$*. This bound is obtained by applying the observer gain* $\tilde{L}^\star = P^{\star -1} Y^\star$*, where* $(P^\star, Y^\star, \Gamma^\star)$ *solves the above SDP.*

4.5.3 *Radii of Estimates and Convergence of Errors*

In this section, we are interested in computing closed form expressions for the estimation radii and sufficient conditions for their convergence, as well as their steady state values (if they exist). Notice that considering the general case of Class 0 functions, i.e., without imposing any additional assumption on $f(\cdot)$, to the best of our knowledge, there is no guarantee that any closed form expressions for the radii can be found using (4.12), since there is no means to relate the state error $\tilde{x}_{k|k}$ to the function increment Δf_k, whereas when $f(\cdot)$ belongs to either of the classes I, II or III, it is possible to relate function variations to the estimation errors and to find closed form expressions for the radii (cf. Theorem 4.5.9).

It is worth mentioning that for linear time-invariant systems, strong detectability of the system is a sufficient condition for the convergence of the radii δ^x_k and δ^d_{k-1} to steady state [131], but it is less clear for general nonlinear systems. Notice that if $f(\cdot)$ is a Class III function, i.e., in the LPV case, even strong detectability of all constituent LTI systems does not guarantee that the radii converge. The reason is that the convergence hinges on the stability of the product of *time-varying* matrices (cf. proof of Theorem 4.5.9), which is not guaranteed even if all the multiplicands are stable. In the following, we discuss some sufficient conditions for the convergence of the radii to steady state, where first we characterize the resulting radii δ^x_k and δ^d_{k-1} when using our proposed observer.

Theorem 4.5.9 (Radii of Estimates)**.** *Consider system* (4.3) *along with the observer* (4.6)–(4.11)*. Suppose the conditions of Theorem 4.5.8 hold. Let* $\Re \triangleq -(\Psi\Phi G_1 M_1 T_1 + \Psi G_2 M_2 T_2 + \tilde{L} T_2)$*,* $\alpha \triangleq \|V_2 M_2 C_2\|\eta_w + \big[\|(V_2 M_2 C_2 G_1 - V_1) M_1 T_1\| + \|V_2 M_2 T_2\|\big]\eta_v$ *and* $\tilde{\eta} \triangleq \|\Re\|\eta_v + \|\Psi\Phi W\|\eta_w$*, with* Φ *and* Ψ *defined in Lemma 4.5.1 and* T_1, T_2 *given in*

Section 2.2.1. Then, the radii δ_k^x *and* δ_{k-1}^d *can be obtained as:*

$$\begin{aligned}\delta_k^x \quad &= \min(\sqrt{\frac{\tilde{x}_{0|0}^\top P \tilde{x}_{0|0}}{\lambda_{\min}(P)}}, \delta_0^x \theta^k + \overline{\eta} \textstyle\sum_{i=1}^k \theta^{i-1}), \qquad (4.25)\\ \delta_{k-1}^d &= \beta \delta_{k-1}^x + \overline{\alpha}, \qquad (4.26)\end{aligned}$$

where P *is derived in Theorem 4.5.3 for different classes of systems by solving the LMIs in* (4.13) *and* θ, $\overline{\eta}$, β *and* $\overline{\alpha}$ *are defined for the different function classes as follows:*

I. If $f(\cdot)$ *is a Class I function, then*

$$\begin{aligned}&\theta \triangleq (L_f + \|\Psi\|)\|(I - \tilde{L}C_2)\Phi\|,\\ &\overline{\eta} \triangleq \tilde{\eta},\\ &\beta \triangleq \|V_1 M_1 C_1 - V_2 M_2 C_2 \Psi\| + L_f \|V_2 M_2 C_2\|,\\ &\overline{\alpha} \triangleq \alpha.\end{aligned} \qquad (4.27)$$

II. If $f(\cdot)$ *is a Class II function, then*

$$\begin{aligned}&\theta \triangleq \|(I - \tilde{L}C_2)\Phi(\mathcal{A} - \Psi)\|,\\ &\overline{\eta} \triangleq \tilde{\eta} + \|(I - \tilde{L}C_2)\Phi\|\gamma,\\ &\beta \triangleq \|V_1 M_1 C_1 + V_2 M_2 C_2 (\mathcal{A} - \Psi)\|,\\ &\overline{\alpha} \triangleq \alpha + \|V_2 M_2 C_2\|\gamma.\end{aligned} \qquad (4.28)$$

III. If $f(\cdot)$ *is a Class III function, then*

$$\begin{aligned}&\theta \triangleq \max\nolimits_{i \in \{1,2,\ldots,N\}} \|A_{e,i}\|,\\ &\overline{\eta} \triangleq \tilde{\eta},\\ &\beta \triangleq \max\nolimits_{i \in \{1,2,\ldots,N\}} \|V_1 M_1 C_1 + V_2 M_2 C_2 A_{e,i}\|,\\ &\overline{\alpha} \triangleq \alpha.\end{aligned} \qquad (4.29)$$

with $A_{e,i} \triangleq (I - \tilde{L}C_2)\Phi(A^i - \Psi)$, *for all* $i \in \{1 \dots N\}$ *and* V_1, V_2 *given in Section 2.2.1.*

Hence, the sequences of the error radii $\{\delta^x_k\}_{k=1}^\infty$ *and* $\{\delta^d_{k-1}\}_{k=1}^\infty$ *are uniformly bounded by* $\overline{\delta} \triangleq \sqrt{\frac{\tilde{x}_{0|0}^\top P \tilde{x}_{0|0}}{\lambda_{\min}(P)}}$ *and* $\beta\overline{\delta} + \overline{\eta}$, *respectively. Furthermore, they are convergent if* $\theta < 1$ *and if so, the steady state radii are given by:*

$$\lim_{k\to\infty} \delta^x_k = \min\left(\sqrt{\frac{\tilde{x}_{0|0}^\top P \tilde{x}_{0|0}}{\lambda_{\min}(P)}}, \frac{\overline{\eta}}{1-\theta}\right),$$
$$\lim_{k\to\infty} \delta^d_k = \beta \min\left(\sqrt{\frac{\tilde{x}_{0|0}^\top P \tilde{x}_{0|0}}{\lambda_{\min}(P)}}, \frac{\overline{\eta}}{1-\theta}\right) + \overline{\alpha}.$$

The resulting fixed-order set-valued observer is summarized in Algorithm 3.

Remark 4.5.10. *Note that according to* (4.25) *and* (4.26), *if the sufficient conditions in Theorem 4.5.9 hold, i.e., when the observer is stable, the sequences of radii,* $\{\delta^x_k, \delta^d_{k-1}\}_{k=1}^\infty$, *are uniformly bounded, regardless of the value of* θ. *Consequently, the sequences of errors,* $\{\tilde{x}_{k|k}, \tilde{d}_{k-1|k-1}\}_{k=1}^\infty$, *are uniformly bounded and do not diverge. Moreover, if* $\theta > 1$, *the sequences of radii may be non-convergent (albeit uniformly bounded), but the the sequences of errors may still converge.*

Corollary 4.5.11. *If* $f(\cdot)$ *is a Class III function and the conditions of Theorem 4.5.8 hold, then, the radii* δ^x_k *and* δ^d_{k-1}, *computed in* (4.25) *and* (4.26), *are convergent if* $\|A_{e,i}\| < 1$ *for all* $i \in \{1, 2, \dots, N\}$, *where* $A_{e,i} \triangleq (I - \tilde{L}C_2)\Phi(A^i - \Psi)$, *with* Φ *and* Ψ *defined in Lemma 4.5.1.*

Remark 4.5.12. *Alternatively, we can trade off between observer "optimality" (i.e., the noise attenuation level) and "convergence" of the error radii (i.e., the steady state values). We can find* η *(e.g., by a line search) that satisfies the following feasibility problem:*

$$\begin{aligned} &\text{Find } (P, Y, \Gamma) \\ &s.t. \quad (4.13), (4.20) \text{ hold,} \end{aligned}$$

and $\theta < 1$*, with* θ *defined in Theorem 4.5.9. Although the designed observer may not be optimum in the minimum* $\mathcal{H}_\infty$ *sense when using this method, we can guarantee the steady state convergence of the radii instead.*

Algorithm 3 Simultaneous Input and State Observer

```
Initialize:
Compute $M_1, M_2, \tilde{L}$ via Theorem 4.5.8 and $\theta, \overline{\eta}, \beta, \overline{\alpha}$ via Theorem 4.5.9; $\Phi = I - G_2 M_2 C_2$;
$\hat{x}_{0|0} = \hat{x}_0 = \text{centroid}(\hat{X}_0)$;
$\hat{d}_{1,0} = M_1(z_{1,0} - C_1 \hat{x}_{0|0} - D_1 u_0)$;
$\delta_0^x = \min_{\delta} \{\|x - \hat{x}_{0|0}\| \leq \delta, \forall x \in \hat{X}_0\}$;
for $k = 1$ to $K$ do
▷ Estimation of $d_{2,k-1}$ and $d_{k-1}$
    $\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}) + B u_{k-1} + G_1 \hat{d}_{1,k-1}$;
    $\hat{d}_{2,k-1} = M_2(z_{2,k} - C_2 \hat{x}_{k|k-1} - D_2 u_k)$;
    $\hat{d}_{k-1} = V_1 \hat{d}_{1,k-1} + V_2 \hat{d}_{2,k-1}$;
    $\delta_{k-1}^d = \beta(\delta_0^x \theta^{k-1} + \overline{\eta} \sum_{i=1}^{k-1} \theta^{i-1}) + \overline{\alpha}$;
    $\hat{D}_{k-1} = \{d \in \mathbb{R}^l : \|d - \hat{d}_{k-1}\| \leq \delta_{k-1}^d\}$;
▷ Time update
    $\hat{x}_{k|k}^\star = \hat{x}_{k|k-1} + G_2 \hat{d}_{2,k-1}$;
▷ Measurement update
    $\hat{x}_{k|k} = \hat{x}_{k|k}^\star + \tilde{L}(z_{2,k} - C_2 \hat{x}_{k|k}^\star - D_2 u_k)$;
    $\delta_k^x = \delta_0^x \theta^k + \overline{\eta} \sum_{i=1}^{k} \theta^{i-1}$;
    $\hat{X}_k = \{x \in \mathbb{R}^n : \|x - \hat{x}_{k|k}\| \leq \delta_k^x\}$;
▷ Estimation of $d_{1,k}$
    $\hat{d}_{1,k} = M_1(z_{1,k} - C_1 \hat{x}_{k|k} - D_1 u_k)$;
end for
```

4.6 Simulation Results and Comparison with Benchmark Observers

Two simulation examples are considered in this section to demonstrate the performance of the proposed observer. In the first example, where the dynamic system belongs to Classes I and II, we consider simultaneous input and state estimation problem and design observers for each class to study their performances. Our second example is a benchmark dynamical Lipschitz continuous (i.e., Class I) system, where we compare the results of our observer with two other existing observers in the literature, [23, 25]. We consider two different scenarios, one with a bounded unknown input, and the other with an unbounded unknown input. The results show that in the unbounded input scenario, when applying the observers in [23, 25], the estimation errors diverge, while as expected from our theoretical results, the estimation errors of our proposed observer converge to steady state values.

4.6.1 Single-Link Flexible-Joint Robotic System

We consider a single-link manipulator with flexible joints [3, 95], where the system has 4 states. We slightly modify the dynamical system described in [3], by ignoring the dynamics for the unknown inputs (different from the existing bounded disturbances) to make them *completely unknown* input signals. We also consider bounded-norm disturbances (instead of stochastic noise signals in [3, 95]). So, we have the dynamical system (4.3) with $n = 4$, $f(x) = Ax + \begin{bmatrix} 0 & 2.16T_s & 0 & -3.33T_s\sin(x_3) \end{bmatrix}^\top$, $p = m = 1$,

$$A = \begin{bmatrix} 1 & T_s & 0 & 0 \\ -48.6T_s & 1-1.25T_s & 48.6T_s & 0 \\ 0 & 0 & 1 & T_s \\ 19.5T_s & 0 & -19.5T_s & 1 \end{bmatrix},\ l = 2,\ B = 0_{4\times 1},\ G = T_s \begin{bmatrix} 5 & 5 & 2 & 1 \end{bmatrix}^\top,$$

$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $W = I$, $D = 0_{2\times 1}$, $T_s = 0.01$, $H = T_s \begin{bmatrix} 1.1 & 2 \end{bmatrix}^\top$ and $\eta_w = \eta_v = 0.1$. The unknown input signal is depicted in Figure 4.3. Vector field $f(\cdot)$ is a Class I function with $L_f = 3.33T_s\|diag\{0,0,0,1\}\| = 3.33T_s$ (cf. [3]), as well as a Class II function, with $\mathcal{A} = A$ and $\gamma = 0.56$ (cf. Lemma 4.1.15). Considering Theorem 4.5.8, cases I and II of the sufficient conditions are satisfied. Solving the corresponding SDPs returns

$$P^\star_{Lip} = \begin{bmatrix} 1.3347 & -1.4121 & -0.2308 & -0.1154 \\ -1.4121 & 2.4642 & -0.1435 & -0.0717 \\ -0.2308 & -0.1435 & 1.1048 & -0.0364 \\ -0.1154 & -0.0717 & -0.0364 & 1.1594 \end{bmatrix}, \Gamma^\star_{Lip} = 0.6391,$$

$$(Y^\star{}_{Lip})^\top = \begin{bmatrix} 0.3237 & 0.2261 & 0.0265 & 0.0132 \end{bmatrix}^\top, \qquad \eta^\star_{Lip} = 1.0560$$

and consequently, $(\tilde{L}^\star_{Lip})^\top = \begin{bmatrix} 1.1360 & 0.7694 & 0.3672 & 0.1836 \end{bmatrix}^\top$ for case I, and

$$P^\star_{DQC} = \begin{bmatrix} 2.8641 & -2.9917 & -1.0025 & -0.5012 \\ -2.9917 & 4.9445 & -0.7205 & -0.3602 \\ -1.0025 & -0.7205 & 4.5920 & -0.1786 \\ -0.5012 & -0.3602 & -0.1786 & 4.8599 \end{bmatrix}, \Gamma^\star_{DQC} = 1.1473,$$

$$(Y^\star{}_{DQC})^\top = \begin{bmatrix} -0.3234 & 1.0288 & 0.0453 & 0.0227 \end{bmatrix}^\top, \qquad \eta^\star_{Lip} = 0.9641,$$

and $(\tilde{L}^\star_{DQC})^\top = \begin{bmatrix} 0.8205 & 0.7619 & 0.3147 & 0.1573 \end{bmatrix}^\top$ for case II. We observe from Figures 4.2 and 4.3 that our proposed observer, i.e., Algorithm 1 is able to find set-valued estimates of the states and unknown inputs, for Lipschitz continuous (Class I) and DQC* (Class II) functions. The actual estimation errors are also within the predicted upper bounds (cf. Figure 4.4), which converge to steady state values as established in Theorem 2.3.5. Furthermore, Figures 4.2–4.4 show that for this specific example system, estimation errors and their radii are tighter when applying the obtained observer gains for Class I (i.e., Lipschitz) functions, when compared to applying the ones corresponding to the Class II (i.e, DQC*) functions.

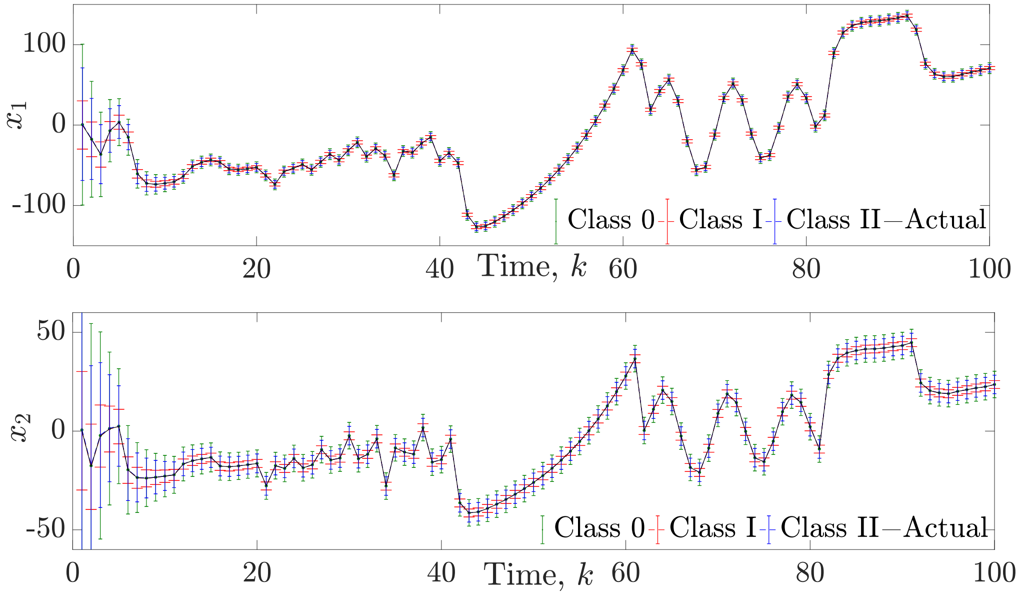


Figure 4.2: Actual States x_1,x_2, as Well as Their Class 0 Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case (0) in Theorem 4.5.8), Class I Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case I in Theorem 4.5.8) and Class II Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case II in Theorem 4.5.8)

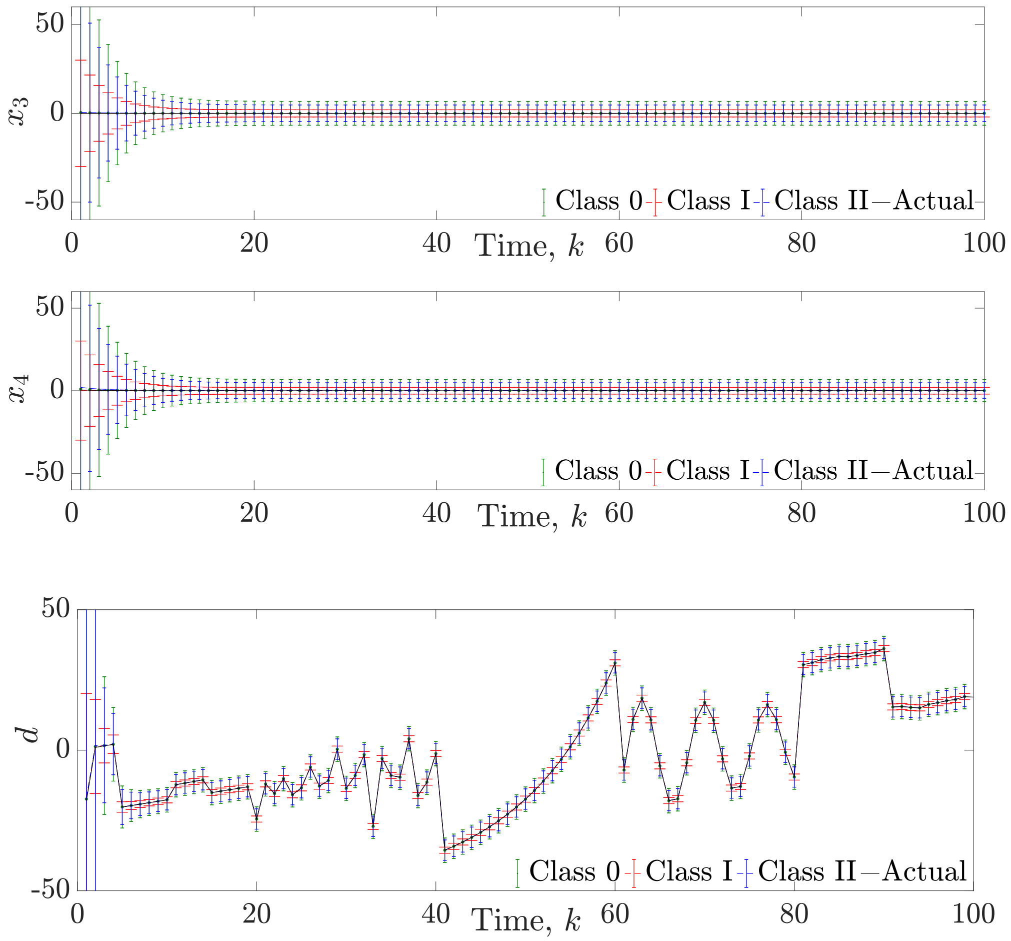


Figure 4.3: Actual States x_3,x_4 and Input d, as Well as Their Class 0 Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case (0) in Theorem 4.5.8), Class I Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case I in Theorem 4.5.8) and Class II Estimates (i.e., the Obtained Estimates by Applying the Corresponding Gains for Case II in Theorem 4.5.8)

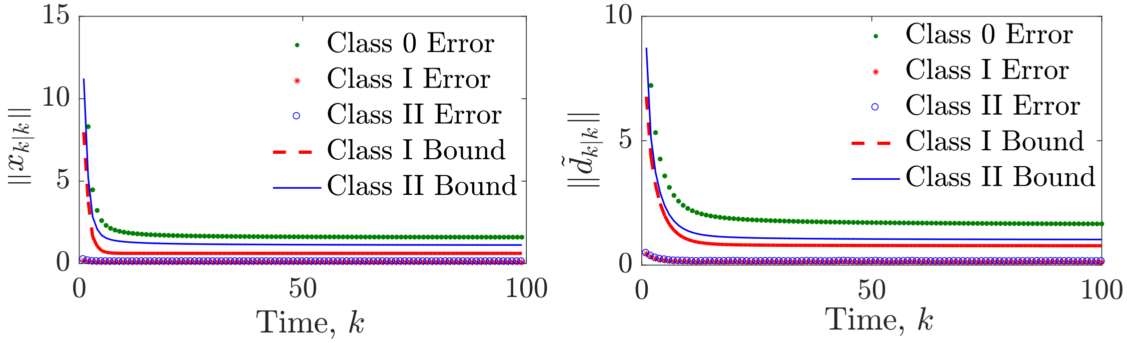


Figure 4.4: Estimation Errors and Their Upper Bounds for Class 0 (General Nonlinear), Class I (Lipschitz) and Class II (DQC*) Functions

4.6.2 Comparison with Benchmark Observers

In this section, we illustrate the effectiveness of our Simultaneous Input and State Set-Valued Observer (SISO), by comparing its performance with two benchmark observers in [23] and [25]. The designed estimator in [23] calculates both (point) state and unknown input estimates, while the observer in [25], only obtains (point) state estimates. For comparison, we apply all the three observers on a benchmark dynamical system in [23], which is in the form of (4.3) with $n = 2$, $m = l = p = 1$, $f(x) = \begin{bmatrix} -0.42x_1 + x_2 & -0.6x_1 - 1.25\tanh(x_1) \end{bmatrix}^\top$, $G = \begin{bmatrix} 1 & -0.65 \end{bmatrix}^\top$,$B = D = H = 0_{1\times 1}$, $C = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $W = I$, $\eta_w = 0.2$ and $\eta_v = 0.1$. The vector field $f(\cdot)$ is Lipschitz continuous (i.e., Class I) with $L_f = 1.1171$. We consider two scenarios for the unknown input. In the first, we consider a random signal with bounded norm, i.e., $\|d_k\| \leq 0.2$ for the unknown input d_k, while d_k in the second scenario is a time-varying signal that becomes unbounded when time increases. As is demonstrated in Figures 4.5 and 4.6, in the first scenario, i.e., with bounded unknown inputs, the set estimates of our approach (i.e., SISO estimates) converge to steady state values and the point estimates of the two benchmark approaches [23, 25] are within the predicted upper bounds and exhibit a convergent behavior for all 50 randomly chosen initial values (cf. Figure 4.6). In this scenario, the two benchmark approaches result in slightly better performance than SISO, since they benefit from the additional assumption of *bounded* input.

More interestingly, considering the second scenario, i.e., with unbounded unknown inputs, Figures 4.7 and 4.8 demonstrate that our set-valued estimates still converge, i.e., our observer remains stable for all 50 randomly chosen initial values, with $P^\star = \begin{bmatrix} 1.9543 & 1.2561 \\ 1.2561 & 5.1084 \end{bmatrix}$, $Y^\star = \begin{bmatrix} -0.1196 & 0.3887 \end{bmatrix}^\top$, $\tilde{L}^\star = \begin{bmatrix} -0.1307 & 0.1082 \end{bmatrix}$, $\Gamma^\star = 0.6360$ and $\eta^\star = 1.9093$, while the estimates of the two benchmark approaches exceed the boundaries of the compatible sets of states and inputs after some time steps of our approach and display a divergent behavior for all the initial values (cf. Figure 4.8).

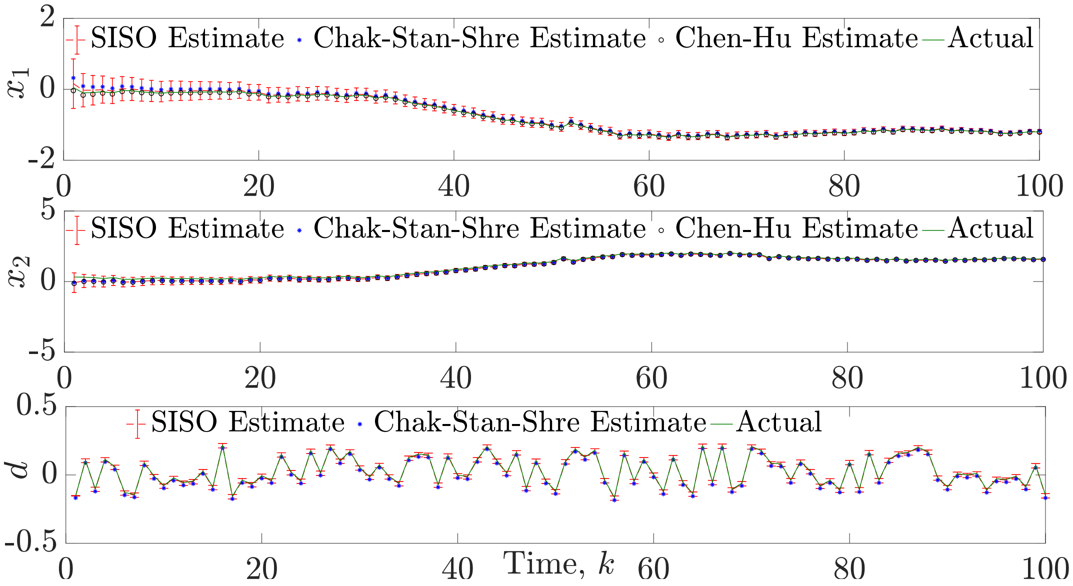


Figure 4.5: Actual States x_1, x_2, and Their Estimates, as Well as Unknown Input d and Its Estimates in the Bounded Unknown Input Scenario, Obtained by Applying the Observer in [25] (Chen-Hu Estimate), the Observer in [23] (Chak-Stan-Shre Estimate) and Our Designed observer (SISO Estimate)

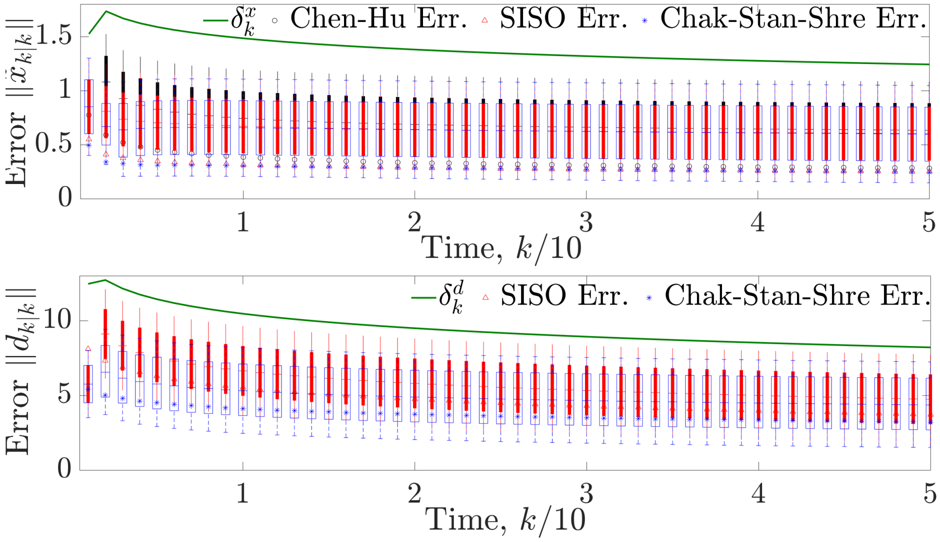


Figure 4.6: Estimation Errors in the Bounded Unknown Input Scenario for 50 Different Initial Values (Using Box Plots), Obtained by Applying the Observer in [25] (Chen-Hu Err.), the observer in [23] (Chak-Stan-Shre Err.) and Our Designed Observer (SISO Err.), as Well as the Computed Upper Bounds for the State and Input Errors (δ_k^x and δ_k^d)

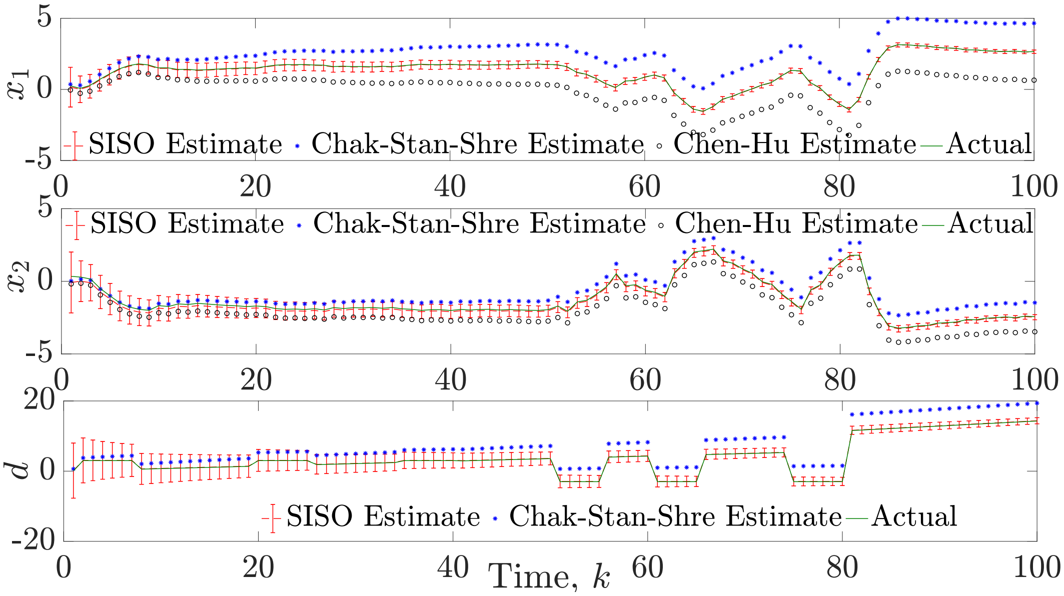


Figure 4.7: Actual States x_1, x_2, and Their Estimates, as Well as Unknown Input d and Its Estimates in the Unbounded Unknown Input Scenario, Obtained by Applying the Observer in [25] (Chen-Hu Estimate), the Observer in [23] (Chak-Stan-Shre Estimate) and Our Designed Observer (SISO Estimate)

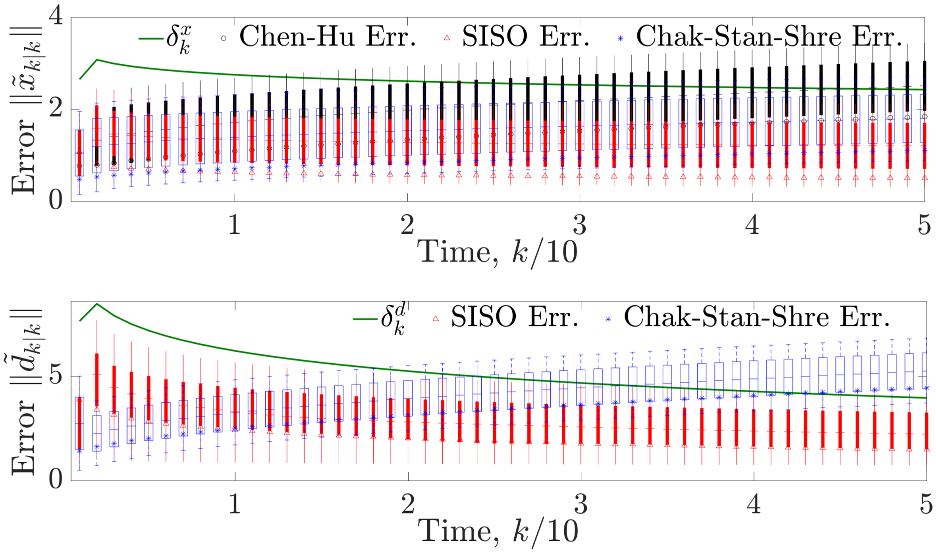


Figure 4.8: Estimation Errors in the Unbounded Unknown Input Scenario for 50 Different Initial Values (Using Box Plots), Obtained by Applying the Observer in [25] (Chen-Hu Err.), the Observer in [23] (Chak-Stan-Shre Err.) and Our Designed Observer (SISO Err.), as Well as the Computed Upper Bounds for the State and Input Errors (δ_k^x and δ_k^d)

4.7 Conclusion

We presented fixed-order set-valued $\mathcal{H}_\infty$-observers for nonlinear bounded-error discrete-time dynamic systems with unknown inputs. Sufficient Linear Matrix Inequalities for Lyapunov stability of the designed observer were derived for different classes of nonlinear systems, including general nonlinear systems, Lipschitz continuous systems, Decremental Quadratic Constrained systems and Linear Parameter-Varying systems. Moreover, we derived additional LMI conditions and corresponding tractable semi-definite programs for obtaining the minimum $\mathcal{H}_\infty$ norm for the transfer function that maps the noise signal to the state error of the stable observers.

In addition, we derived sufficient conditions for the convergence of the radii of the set-valued state and input estimates and derived their steady state values. Finally, using two illustrative examples, we demonstrated the effectiveness of our proposed design, as well as its advantages over two existing benchmark observers. For future work, we plan to generalize this framework to hybrid and switched nonlinear systems and consider other forms of CPS attacks.

Chapter 5

SIMULTANEOUS MODE, STATE AND INPUT SET-VALUED OBSERVERS FOR SWITCHED NONLINEAR SYSTEMS

In this chapter [a] , we study the problem of designing a simultaneous mode, input and state set-valued observer for a class of hidden mode switched nonlinear systems with bounded-norm noise and unknown input signals, where the hidden mode and unknown inputs can represent fault or attack models and exogenous fault/disturbance or adversarial signals, respectively. The proposed multiple-model design has three constituents: (i) a bank of mode-matched set-valued observers, (ii) a mode observer and (iii) a global fusion observer. The mode-matched observers recursively find the sets of compatible states and unknown inputs conditioned on the mode being the true mode, while the mode observer eliminates incompatible modes by leveraging a residual-based criterion. Then, the global fusion observer outputs the estimated sets of states and unknown inputs by taking the union of the mode-matched set-valued estimates over all compatible modes. Moreover, sufficient conditions to guarantee the elimination of all false modes (i.e., mode detectability) are provided and the effectiveness of our approach is demonstrated and compared with existing approaches using an illustrative example.

5.1 Problem Formulation

Consider a hidden mode switched nonlinear system with bounded-norm noise and unknown inputs (i.e., a hybrid system with nonlinear and noisy system dynamics in

---

[a] The content of this chapter is documented as a submitted and under review paper in [121].

each mode, where the mode and some inputs are not known/measured):

$$\begin{aligned} x_{k+1} &= f^q(x_k) + B^q u_k^q + G^q d_k^q + W^q w_k^q, \\ y_k &= C^q x_k + D^q u_k^q + H^q d_k^q + v_k^q, \end{aligned} \tag{5.1}$$

where $x_k \in \mathbb{R}^n$ is the continuous system state and $q \in \mathbb{Q} = \{1, 2, \ldots, Q\} \subset \mathbb{N}$ is the hidden discrete state or *mode*. For each $q \in \mathbb{Q}$, $y_k \in \mathbb{R}^l$ is the measurement output signal and $w_k^q \in \mathbb{R}^n$ and $v_k^q \in \mathbb{R}^l$ are external process and measurement disturbances with known ℓ_2-norm bounds, i.e., $\|w_k\|_2 \leq \eta_w$ and $\|v_k\|_2 \leq \eta_v$, respectively. Moreover, $u_k^q \in U_k \subset \mathbb{R}^m$ is the *known* input and $d_k^q \in \mathbb{R}^p$ the unknown input signal (representing, e.g., the input of other agents/robots or adversarially injected data signal). It is worth mentioning that no prior 'useful' knowledge or assumption of the dynamics of d_k^q is assumed. For each (fixed) mode q, the mapping $f^q(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$ and the matrices $B^q \in \mathbb{R}^{n \times m}$, $G^q \in \mathbb{R}^{n \times p}$, $C^q \in \mathbb{R}^{l \times n}$, $D^q \in \mathbb{R}^{l \times m}$ and $H^q \in \mathbb{R}^{l \times p}$ are the corresponding mode-dependent known state vector field and system matrices, respectively.

The above modeling framework can capture a very broad range of problems, including intention estimation, fault detection and resilient state estimation against sparse data injection and switching/mode attacks. Specifically, in the context of intention estimation or fault diagnosis, each mode represents an intent or fault model and the unknown inputs can model the inputs of other agents/robots or exogenous fault signals. On the other hand, with regard to resilient state estimation, the switching/mode attacks (e.g., attacks on circuit breakers) can be represented with a set of different $f^q(\cdot)$, B^q, C^q and D^q, while the unknown attack location of sparse data injection attacks can be modeled by a set of different G^q and H^q that represent the different hypotheses for which actuators and sensors are attacked or not attacked. Further, the attack signal magnitudes can be modeled as the unknown inputs in this scenario.

In addition, we assume the following:

Assumption 5.1.1. *There is only one "true" mode, i.e. the true mode* q^* *is constant over time.*

Assumption 5.1.2. *For each* $q \in \mathbb{Q}$*,* $f^q(\cdot)$ *is twice continuously differentiable and Lipschitz continuous on its domain with a known Lipschitz constant* $L_f^q > 0$*.*

Using the above modeling framework, the simultaneous state, unknown input and hidden mode estimation problem based on a multiple-model framework can be stated as follows:

Problem 5.1.3. *Given a hidden mode switched nonlinear discrete-time system with unknown inputs and bounded-norm noise in the form of* (5.1)*,*

(i) Design a bank of mode-matched observers, where each mode-matched observer, conditioned on the mode being true, optimally returns the set-valued estimates of compatible states and unknown inputs in the minimum $\mathcal{H}_\infty$*-norm sense, i.e., with minimum average power amplification.*

(ii) Find a threshold criterion to eliminate false modes and subsequently, develop a mode observer via elimination.

(iii) Derive sufficient conditions for the elimination of all false modes.

5.2 Proposed Observer Design

In this section, we propose a multiple-model approach for simultaneous mode, state and unknown input estimation for the system in (5.1), with the goal of recursively finding the sets of states $\hat{X}_k$, unknown inputs $\hat{D}_k$ and modes $\hat{\mathbb{Q}}_k$ that are compatible with observed outputs y_k.

5.2.1 Overview of Multiple-Model Approach

The multiple-model design approach consists of three steps: (i) designing a bank of mode-matched set-valued observers, (ii) developing a mode observer for eliminating incompatible modes using a residual-based threshold, and (iii) devising a global fusion observer that returns the desired set-valued mode, input and state estimates.

Mode-Matched Set-Valued Observer

First, based on the optimal fixed-order observer design in [118], we develop a bank of mode-matched observers, which includes $Q \in \mathbb{N}$ simultaneous state and input $\mathcal{H}_\infty$ set-valued observers, which can be briefly summarized as follows. For each mode-matched observer corresponding to mode q, following the approach in [118, Section 4], we consider set-valued fixed-order estimates in the form of ℓ_2-norm balls:

$$\hat{D}^q_{k-1} = \{d_{k-1} \in \mathbb{R}^p : \|d_{k-1} - \hat{d}^q_{k-1}\|_2 \leq \delta^{d,q}_{k-1}\}, \tag{5.2}$$

$$\hat{X}^q_k = \{x_k \in \mathbb{R}^n : \|x_k - \hat{x}^q_{k|k}\|_2 \leq \delta^{x,q}_k\}, \tag{5.3}$$

where their centroids $\hat{x}^q_{k|k}$ and $\hat{d}^q_{k-1}$ are obtained with the following three-step recursive observer that is optimal in $\mathcal{H}_\infty$-norm sense (cf. [118, Section 4.2] for more details):
Unknown Input Estimation:

$$\begin{aligned} \hat{d}^q_{1,k} &= M^q_1(z^q_{1,k} - C^q_1\hat{x}^q_{k|k} - D^q_1 u^q_k), \\ \hat{d}^q_{2,k-1} &= M^q_2(z^q_{2,k} - C^q_2\hat{x}^q_{k|k-1} - D^q_2 u^q_k), \\ \hat{d}^q_{k-1} &= V^q_1\hat{d}^q_{1,k-1} + V^q_2\hat{d}^q_{2,k-1}; \end{aligned} \tag{5.4}$$

Time Update:

$$\begin{aligned} \hat{x}^q_{k|k-1} &= f^q(\hat{x}^q_{k-1|k-1}) + B^q u^q_{k-1} + G^q_1\hat{d}^q_{1,k-1}, \\ \hat{x}^{\star,q}_{k|k} &= \hat{x}^q_{k|k-1} + G^q_2\hat{d}^q_{2,k-1}; \end{aligned} \tag{5.5}$$

Measurement Update:

$$\hat{x}^q_{k|k} = \hat{x}^{\star,q}_{k|k} + \tilde{L}^q(z^q_{2,k} - C^q_2\hat{x}^{\star,q}_{k|k} - D^q_2 u^q_k), \tag{5.6}$$

where C^q_1, C^q_2, D^q_1, D^q_2, G^q_1, G^q_2, V^q_1, V^q_2, $z^q_{1,k}$ and $z^q_{2,k}$ can be computed by applying a similarity transformation described in Section 4.4.1 and $\tilde{L}^q \in \mathbb{R}^{n\times(l-p_{H^q})}$, $M^q_1 \in \mathbb{R}^{p_{H^q}\times p_{H^q}}$ and $M^q_2 \in \mathbb{R}^{(p-p_{H^q})\times(l-p_{H^q})}$ are observer gain matrices that are chosen via the following Proposition 5.2.1. This proposition is a restatement of the results in [118] that is tailored to the setting considered in this paper, where the main idea is to minimize the "volume" of the set of compatible states and unknown inputs, quantified by the radii $\delta^{d,q}_{k-1}$ and $\delta^{x,q}_k$.

Proposition 5.2.1. *[118, Proposition 5.16, Lemma 5.1 & Theorem 5.13] Consider system* (5.1) *and a bank of Q mode-matched observers in the form of* (5.4)–(5.6)*. Suppose that $\forall q \in \mathbb{Q} \triangleq \{1,\dots,Q\}$, $\mathrm{rk}(C^q_2 G^q_2) = p - p_{H^q}$ and M^q_1, M^q_2 are chosen as $M^q_1 = (\Sigma^q)^{-1}$ and $M^q_2 = (C^q_2 G^q_2)^\dagger$, where Σ^q is obtained by applying singular value decomposition on H^q (cf. Section 4.4.1 for more details). Then, the following statements hold:*

(a) Given mode $q \in \mathbb{Q}$, the following difference equation governs the state estimation error dynamics (i.e., the dynamics of $\tilde{x}^q_{k|k} \triangleq x_k - \hat{x}^q_{k|k}$):

$$\tilde{x}^q_{k+1|k+1} = (I - \tilde{L}^q C^q_2)\Phi^q(\Delta f^q_k - \Psi^q \tilde{x}^q_{k|k}) + \mathcal{W}^q(\tilde{L}^q)\overline{w}^q_k, \tag{5.7}$$

where

$$\begin{aligned}
&\Delta f_k^q \triangleq f^q(x_k) - f^q(\hat{x}_k^q), \quad \Phi^q \triangleq I - G_2^q M_2^q C_2^q,\\
&\overline{w}_k^q \triangleq \begin{bmatrix} (\tfrac{1}{\sqrt{2}}) v_k^{q\top} & w_k^{q\top} & (\tfrac{1}{\sqrt{2}}) v_{k+1}^{q\top} \end{bmatrix}^\top,\\
&R^q \triangleq \begin{bmatrix} -\sqrt{2}\Phi^q G_1^q M_1^q T_1^q & -\Phi^q W^q & -\sqrt{2} G_2^q M_2^q T_2^q \end{bmatrix},\\
&Q^q \triangleq \begin{bmatrix} 0_{(l-p_{H^q})\times l} & 0_{(l-p_{H^q})\times n} & -\sqrt{2} T_2^q \end{bmatrix},\\
&\Psi^q \triangleq G_1^q M_1^q C_1^q, \quad \mathcal{W}^q(\tilde{L}^q) \triangleq (I - \tilde{L}^q C_2^q) R^q + \tilde{L}^q Q^q.
\end{aligned}$$

(b) Solving the following mixed-integer SDP for each mode q:

$$\begin{aligned}
(\rho_q^\star)^2 = &\min_{\{P\succ 0, \Gamma\succ 0, \tilde{\Gamma}\succeq 0, \breve{Q}\succeq 0, Y, \breve{Z}, \rho^2>0, 0\le\alpha\le 1, \varepsilon_1>0, \varepsilon_2>0, \kappa>0, \kappa_1>0, \kappa_2>0\}} \rho^2\\
\text{s.t.}\; &\begin{bmatrix} P & \tilde{Y}_1^q \\ \tilde{Y}_1^{q\top} & \tilde{\mathbf{M}}_1^q \end{bmatrix} \succeq 0, \begin{bmatrix} P & \tilde{Y}_2^q \\ \tilde{Y}_2^{q\top} & \tilde{\mathbf{M}}_2^q \end{bmatrix} \succeq 0, \begin{bmatrix} P & \tilde{Y}_1^q \\ \tilde{Y}_1^{q\top} & \tilde{\mathbf{M}}_3^q \end{bmatrix} \succeq 0,\\
&\begin{bmatrix} P & \tilde{Y}_2^q \\ \tilde{Y}_2^{q\top} & \breve{Z} \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \tilde{\Gamma} & \breve{Z} \\ \breve{Z}^\top & \Psi^{q\top}\breve{Q}\Psi^q \end{bmatrix} \succeq 0,\\
&\begin{bmatrix} I-\Gamma & 0 & 0 \\ 0 & P & Y \\ 0 & Y^\top & I \end{bmatrix} \succeq 0, \quad \begin{bmatrix} \mathcal{N}_{11}^q & * & * \\ \mathcal{N}_{21}^q & \mathcal{N}_{22}^q & * \\ \mathcal{N}_{31}^q & 0 & \mathcal{N}_{33}^q \end{bmatrix} \succeq 0,\\
&\kappa_1 I \preceq P \preceq \kappa_2 I, \;\wedge\; ((\kappa_1 \ge 1, \kappa_2 - \kappa_1 < 1) \;\vee\; (\kappa_2 \le 1, \kappa_1 > 0.5)),
\end{aligned}$$

we obtain an observer in the form of (5.4)–(5.6) *with the observer gain* $\tilde{L}^q = (P^q)^{-1} Y^q$, *where* (P^q, Y^q) *are solutions to the above mixed-integer SDP, that*

- *is quadratically stable, and*
- *guarantees that*

$$\theta^q \triangleq \|(I - \tilde{L}^q C_2^q)\Phi^q\|_2 < 1, \tag{5.8}$$

and consequently, the upper bound sequences for the radii $\{\delta_k^{x,q}, \delta_{k-1}^{d,q}\}_{k=1}^{\infty}$*, which are computed as:*

$$\begin{aligned} \delta_k^{x,q} &\triangleq \delta_0^x (\theta^q)^k + \overline{\eta}^q \tfrac{1-(\theta^q)^k}{1-\theta^q}, \\ \delta_{k-1}^d &\triangleq \beta^q \delta_{k-1}^{x,q} + \overline{\alpha}^q, \end{aligned} \tag{5.9}$$

are convergent to some steady state value $\delta_\infty^{x,q}, \delta_\infty^{d,q}$*, where*

$$\begin{aligned}
&\tilde{Y}_1^q \triangleq (P - YC_2^q)\Phi^q, \quad \tilde{Y}_2^q \triangleq -(P - YC_2^q)\Phi^q\Psi^q, \\
&\tilde{\mathbf{M}}_1 \triangleq -\kappa I - \breve{Q}, \quad \tilde{\mathbf{M}}_2 \triangleq -\kappa (L_f^q)^2 I + (1-\alpha)P - \tilde{\Gamma}, \quad \tilde{\mathbf{M}}_3 \triangleq \kappa I, \\
&\mathcal{N}_{21}^q \triangleq \Psi^{q\top}\Phi^{q\top}(PR^q - Y\Omega^q - C_2^{q\top} Y^\top R^q), \\
&\mathcal{N}_{11}^q \triangleq \rho^2 I + 2R^{q\top} Y\Omega^q - R^{q\top} PR^q - \Omega^{q\top}(\Gamma + (\varepsilon_1^{-1} + \varepsilon_2^{-1})I)\Omega^q, \\
&\mathcal{N}_{31}^q \triangleq \Phi^{q\top}(Y\Omega^q + C_2^{q\top} Y^\top R^q - PR^q), \\
&\mathcal{N}_{33}^q \triangleq -\varepsilon_2 \Phi^{q\top} C_2^{q\top} C_2^q \Phi^q + I, \\
&\mathcal{N}_{22}^q \triangleq -I + \alpha P - \varepsilon_1 \Psi^{q\top}\Phi^{q\top} C_2^{q\top} C_2^q \Phi^q \Psi^q - L_f^{q\,2} I, \\
&\delta_\infty^x \triangleq \begin{cases} \delta_{\infty,1}^{x,q}, & \textit{if } \theta_1^q < 1, \theta_2^q \geq 1, \\ \delta_{\infty,2}^{x,q}, & \textit{if } \theta_1^q \geq 1, \theta_2^q < 1, \\ \min(\delta_{\infty,1}^{x,q}, \delta_{\infty,2}^{q,x}), & \textit{if } \theta_1^q < 1, \theta_2^q < 1, \end{cases}, \\
&\delta_{\infty,1}^{x,q} \triangleq \rho_q^\star \sqrt{\frac{\eta_w^{q\,2} + \eta_v^{q2}}{\lambda_{\min}(P^q)(1-\theta_1^q)}}, \quad \delta_{\infty,2}^{x,q} \triangleq \frac{\overline{\eta}^q}{1-\theta_2^q}, \quad \delta_\infty^{d,q} \triangleq \beta^q \delta_\infty^{x,q} + \overline{\alpha}^q, \\
&\theta_1^q \triangleq \tfrac{|\lambda_{\max}(P^q) - 1|}{\lambda_{\min}(P^q)}, \qquad \theta_2{}^q \triangleq (L_f^q + \|\Psi^q\|_2)\|(I - \tilde{L}^q C_2^q)\Phi^q\|_2, \\
&\Omega^q \triangleq C_2^q R^q - Q^q, \qquad \overline{\eta}^q \triangleq \|\Re^q\|_2 \eta_v^q + \|\Psi^q\Phi^q W^q\|_2 \eta_w^q, \\
&\Re^q \triangleq -(\Psi^q\Phi^q G_1^q M_1^q T_1^q + \Psi^q G_2^q M_2^q T_2^q + \tilde{L}^q T_2^q), \\
&\beta^q \triangleq \|V_1^q M_1^q C_1^q - V_2^q M_2^q C_2^q \Psi^q\|_2 + L_f^q \|V_2^q M_2^q C_2^q\|_2, \\
&\overline{\alpha}^q \triangleq \|V_2^q M_2^q C_2^q\|_2 \eta_w^q + \big[\|(V_2^q M_2^q C_2^q G_1^q - V_1^q)M_1^q T_1^q\|_2 + \|V_2^q M_2^q T_2^q\|_2\big]\eta_v^q.
\end{aligned}$$

5.2.2 Mode Observer

To estimate the set of compatible modes, we consider an elimination approach that compares the ℓ_2-norm of *residual* signals against some thresholds. Specifically, we will eliminate a specific mode q, if $\|r^q_k\|_2 > \hat{\delta}^q_{r,k}$, where the residual signal r^q_k is defined as follows and the thresholds $\hat{\delta}^q_{r,k}$ will be derived in Section 5.2.3.

Definition 5.2.2 (Residuals)**.** *For each mode q at time step k, the residual signal is defined as:*

$$r^q_k \triangleq z^q_{2,k} - C^q_2 \hat{x}^{\star,q}_{k|k} - D^q_2 u^q_k.$$

Global Fusion Observer

Finally, combining the outputs of both components above, our proposed global fusion observer will provide mode, unknown input and state set-valued estimates at each time step k as:

$$\hat{\mathbb{Q}}_k = \{q \in \mathbb{Q} \,\Big|\, \|r^q_k\|_2 \leq \hat{\delta}^q_{r,k}\},$$
$$\hat{D}_{k-1} = \cup_{q \in \hat{\mathbb{Q}}_k} D^q_{k-1}, \; \hat{X}_k = \cup_{q \in \hat{\mathbb{Q}}_k} X^q_k.$$

The multiple-model approach is summarized in Algorithm 4.

5.2.3 Mode Elimination Approach

We leverage a relatively simple idea to develop a criterion for elimination of false modes, as follows. We rule out a particular mode as incompatible, if the ℓ_2-norm of its corresponding residual signal exceeds its upper bound conditioned on this mode being true. To do so, for each mode q, we first compute an upper bound ($\hat{\delta}^q_{r,k}$) for the ℓ_2-norm of its corresponding residual at time k, conditioned on q being the *true* mode. Then, comparing the ℓ_2-norm of residual signal in Definition 5.2.2 with $\hat{\delta}^q_{r,k}$, mode q

Algorithm 4 Simultaneous Mode, State and Input Estimation of Nonlinear Systems

```
$\hat{\mathbb{Q}}_0 = \mathbb{Q}$;
for $k = 1$ to $N$ do
    for $q \in \hat{\mathbb{Q}}_{k-1}$ do
      ▷ Mode-Matched State and Input Set-Valued Estimates
        Compute $T_2^q, M_1^q, M_2^q, \tilde{L}^q, \hat{x}^{\star,q}_{k|k}, \hat{X}^q_k, \hat{D}^q_{k-1}$ via Proposition 5.2.1;
        $z^q_{2,k} = T_2^q y_k$;
      ▷ Mode Observer via Elimination
        $\hat{\mathbb{Q}}_k = \hat{\mathbb{Q}}_{k-1}$;
        Compute $r^q_k$ via Definition 5.2.2 and $\hat{\delta}^q_{r,k}$ via Theorem 5.2.7;
        if $\|r^q_k\|_2 > \hat{\delta}^q_{r,k}$ then $\hat{\mathbb{Q}}_k = \hat{\mathbb{Q}}_k \backslash \{q\}$;
        end if
    end for
  ▷ State and Input Estimates
    $\hat{X}_k = \cup_{q \in \hat{\mathbb{Q}}_k} \hat{X}^q_k$;  $\hat{D}_k = \cup_{q \in \hat{\mathbb{Q}}_k} \hat{D}^q_k$;
end for
```

can be eliminated if the residual's ℓ_2-norm is strictly greater than the upper bound. The following proposition and theorem formalize this procedure.

Proposition 5.2.3. *Consider mode q at time step k, its residual signal r^q_k (as defined in Definition 5.2.2) and the unknown true mode q^*. Then,*

$$r^q_k = r^{q|*}_k + \Delta r^{q|q*}_k,$$

with

$$\begin{aligned} r^{q|*}_k &\triangleq z^{q*}_{2,k} - C^q_2 \hat{x}^{\star,q}_{k|k} - D^q_2 u^q_k = T^{q*}_2 y_k - C^q_2 \hat{x}^{\star,q}_{k|k} - D^q_2 u^q_k, \\ \Delta r^{q|q*}_k &\triangleq (T^q_2 - T^{q*}_2) y_k, \end{aligned}$$

where $r_k^{q|*}$ *is the true mode's residual signal (i.e.,* $q = q^*$*), and* $\Delta r_k^{q|q^*}$ *is the* residual error.

Theorem 5.2.4. *Consider mode* q *and its residual signal* r_k^q *at time step* k*. Assume that* $\delta_{r,k}^{q,*}$ *is any signal that satisfies* $\|r_k^{q|*}\|_2 \leq \delta_{r,k}^{q,*}$*, where* $r_k^{q|*}$ *is defined in Proposition 5.2.3. Then, mode* q *is not the true mode, i.e., can be eliminated at time* k*, if* $\|r_k^q\|_2 > \delta_{r,k}^{q,*}$*.*

By the above theorem, our approach guarantees that the true mode is never eliminated. However, Theorem 5.2.4 only provides a sufficient condition for mode elimination at each time step and the capability of our proposed mode observer to eliminate as many false modes as possible is dependent on the tightness of the upper bound, $\delta_{r,k}^{q,*}$.

5.2.4 Tractable Computation of Thresholds

To apply the sufficient condition in Theorem 5.2.4, we need a tractable approach to compute the upper bound $\delta_{r,k}^{q,*}$ that is finite-valued. This procedure is derived and described in the following.

Lemma 5.2.5. *Consider any mode* q *with the unknown true mode being* q^**. Then, at time step* k*, we have*

$$r_k^{q|*} = C_2^q \tilde{x}_{k|k}^{\star,q} + v_{2,k}^q = \mathbb{A}_k^q t_k, \tag{5.10}$$

where

$$
\begin{aligned}
&t_k \triangleq \begin{bmatrix}\tilde{x}_{0|0}^{\top} & v_0^{q\top} & \dots & v_k^{q\top} & w_0^{q\top} & \dots & w_{k-1}^{q\top} & \Delta f_0^{q\top} \dots \Delta f_{k-1}^{q\top}\end{bmatrix}^{\top} \in \mathbb{R}^{(n+l)(k+1)+nk},\\
&\mathbb{A}_k^q \triangleq [A_k^q \quad J_{k-1}^{q,1} \quad (J_{k-1}^{q,2} + J_{k-2}^{q,1}) \cdots (J_1^{q,2} + J_0^{q,1}) \quad J_0^{q,2} \quad J_{k-1}^{q,3} \dots J_0^{q,3} \quad F_{k-1}^q \dots F_0^q],\\
&A_k^q \triangleq (-1)^k((I - \tilde{L}^q C_2^q)\Phi^q \Psi^q)^k,\\
&J_i^q \triangleq \begin{cases} \mathcal{Y}_q, & \textit{if } i = 0,\\ -C_2^q \Phi^q G_1^q M_1^q C_1^q (I - \tilde{L}^q C_2^q)^{i-1} \mathcal{W}^q, & \textit{if } 1 \le i \le k-1,\end{cases}\\
&F_i^q \triangleq \begin{cases} C_2^q \Phi^q, & \textit{if } i = 0,\\ (-1)^i C_2^q \Phi^q G_1^q M_1^q C_1^q ((I - \tilde{L}^q C_2^q)\Psi^q)^{i-1} (I - \tilde{L}^q C_2^q)\Phi^q, & \textit{if } 1 \le i \le k-1,\end{cases}\\
&\mathcal{Y}_q \triangleq \begin{bmatrix} -\sqrt{2} C_2^q \Phi^q G_1^q M_1^q T_1^q & C_2^q \Phi^q W^q & \sqrt{2}(I - C_2^q G_2^q M_2^q) T_2^q \end{bmatrix},\\
&J_i^{q,1} \triangleq J_i^q(1:l), \; J_i^{q,2} \triangleq J_i^q(l+1:2l), \; J_i^{q,3} \triangleq J_i^q(2l+1:2l+n), i = 1, \dots, k-1.
\end{aligned}
$$

Lemma 5.2.6. *For each mode q at time step k, there exists a finite-valued upper bound $\delta_{r,k}^q < \infty$ for $\|r_k^{q|*}\|_2$.*

Clearly, $\delta_{r,k}^q$ in Lemma 5.2.6, if computable, is the *tightest* possible upper bound for the norm of the residual signal and using this as the threshold can eliminate the most possible number of false modes. However, note that although the existence proof of a finite-valued $\delta_{r,k}^q$ is straightforward, the optimization problem in Lemma 5.2.6 is NP-hard [18], since it is a *norm maximization* (not minimization) over the intersection of level sets of lower dimensional norm functions, i.e., it is a non-concave maximization over intersection of quadratic constraints. To tackle this complexity, through the following Theorem 5.2.7, we propose a tractable over-approximation/upper bound for $\delta_{r,k}^q$, which we call $\hat{\delta}_{r,k}^q$ and is used instead as the elimination threshold.

Theorem 5.2.7. *Consider mode q. At time step k, let*

$$
\begin{aligned}
\hat{\delta}^{q}_{r,k} &\triangleq \min\{\delta^{q,tri}_{r,k}, \delta^{q,inf}_{r,k}\}, \\
\delta^{q,tri}_{r,k} &\triangleq \sum_{i=0}^{k-2} L^{q}_{f}\|F^{q}_{i}\|_2 \overline{\delta}^{x,q}_{k-1-i} + \frac{1}{\sqrt{2}}\eta^{q}_{v}(\|J^{q,1}_{i}\|_2 + \|J^{q,3}_{i}\|_2) + \eta^{q}_{w}\|J^{q,2}_{i}\|_2 \\
&\quad + (\|A^{q}_{k}\|_2 + L^{q}_{f}\|F^{q}_{k-1}\|_2)\delta^{x}_{0} + \frac{1}{\sqrt{2}}\eta^{q}_{v}(\|J^{q,1}_{k-1}\|_2 + \|J^{q,3}_{k-1}\|_2) + \eta^{q}_{w}\|J^{q,2}_{k-1}\|_2, \qquad (5.11) \\
\delta^{q,inf}_{r,k} &\triangleq \|\mathbb{A}^{q}_{k} t^{\star}_{k}\|_2,
\end{aligned}
$$

where $t^{\star}_{k} \triangleq \arg\max_{t_k \in \mathcal{T}_k} \|\mathbb{A}^{q}_{k} t_k\|_2$ *and* $\mathcal{T}_k$ *is the set of all vertices of the following hypercube:*

$$
\mathcal{X}^{q}_{k} \triangleq \left\{ x \in \mathbb{R}^{(n+l)(k+1)+nk} \,\middle|\, |x(i)| \leq \begin{cases}
\delta^{x}_{0}, & 1 \leq i \leq n, \\
\eta^{q}_{v}, & n+1 \leq i \leq n + l(k+1), \\
\eta^{q}_{w}, & n + l(k+1) + 1 \leq i \leq (n+l)(k+1), \\
L^{q}_{f}\delta^{x}_{0}, & (n+l)(k+1) + 1 \leq i \leq (n+l)(k+1) + n, \\
\vdots & \\
L^{q}_{f}\overline{\delta}^{x,q}_{j}, & (n+l)(k+1) + jn + 1 \leq i \leq (n+l)(k+1) + n(j+1), \\
\vdots & \\
L^{q}_{f}\overline{\delta}^{x,q}_{k-1}, & (n+l)(k+1) + (k-1)n + 1 \leq i \leq (n+l)(k+1) + nk.
\end{cases} \right\}.
$$

Then, $\hat{\delta}^{q}_{r,k}$ *is an over-approximation for* $\delta^{q}_{r,k}$ *in Lemma 5.2.6, i.e.,* $\hat{\delta}^{q}_{r,k} \geq \delta^{q}_{r,k}$.

Theorem 5.2.7 enables us to obtain an upper bound for $\|r^{q|*}_{k}\|_2$, by enumerating the objective function in (A.62) (cf. Proof of Theorem 5.2.7 in Appendix) for all vertices of the hypercube $\mathcal{X}^{q}_{k}$ and choosing the largest value as $\delta^{q,inf}_{r,k}$. Moreover, we can easily calculate $\delta^{q,tri}_{r,k}$; then, the upper bound is chosen as the minimum of the two as $\hat{\delta}^{q}_{r,k}$.

Remark 5.2.8. *The reason for not only using* $\delta^{q,inf}_{r,k}$ *is two-fold. First, as time increases, the number of required enumerations for* $\delta^{q,inf}_{r,k}$ *(i.e., the cardinality of* $\mathcal{T}_k$*) can be shown to be* $|\mathcal{T}_k| = 2^{(n+l)(k+1)+kn}$*, which increases at an exponential rate. Second and more importantly, as will be shown later in Lemma 5.3.4,* $\delta^{q,inf}_{r,k}$ *goes to infinity as time increases, which renders it ineffective in the limit. On the other hand, Lemma 5.3.4 will show that* $\delta^{q,tri}_{r,k}$ *converges to some steady-state value, so it can always be used as an over-approximation for* $\delta^{q}_{r,k}$ *in the mode elimination process. Nonetheless, we chose to use the minimum of the two bounds, since our simulation results in Section 5.4 show that* $\delta^{q,inf}_{r,k}$ *is generally smaller than* $\delta^{q,tri}_{r,k}$ *in the initial time steps.*

Further, the following result that we will make use of later can be easily obtained as a corollary of Theorem 5.2.7.

Corollary 5.2.9. $t^\star_k$ *(defined in Theorem 5.2.7) has the following norm:*

$$\eta^t_k \triangleq \|t^\star_k\|_2 = \sqrt{n\Big((1+{L^q_f}^2){\delta^x_0}^2 + k{\eta^q_w}^2 + {L^q_f}^2\sum_{j=1}^{k-1}\overline{\delta}^{x,q2}_j\Big) + l(k+1){\eta^q_v}^2}.$$

5.3 Mode Detectability

In addition to the nice properties regarding the quadratic stability and boundedness of the mode-matched set-valued estimates of the state and unknown input obtained from [118], we are interested in guaranteeing the effectiveness of our mode elimination algorithm. Thus, in the following, we search for some sufficient conditions based on the properties/structures of the system dynamics and/or unknown input signals for guaranteeing that the application of Algorithm 1 can eliminate *all* false (i.e., not true) modes after some large enough number of time steps.

To achieve this, we first define the concept of *mode detectability*.

Definition 5.3.1 (Mode Detectability)**.** *System* (5.1) *is called* mode detectable *if there exists a natural number $K > 0$, such that for all time steps $k \geq K$, all false modes are eliminated.*

Moreover, we consider two different sets of assumptions that we will use for deriving our sufficient conditions for mode detectability.

Assumption 5.3.2. *There exist known $R_y, R_x \in \mathbb{R}$ such that $\forall k, y_k \in Y \triangleq \{y \in \mathbb{R}^l \big| \, \|y\|_2 \leq R_y\}$ and $x_k \in X \triangleq \{x \in \mathbb{R}^n \big| \, \|x\|_2 \leq R_x\}$, i.e., there exist known bounds for the whole observation/measurement and state spaces, respectively.*

Assumption 5.3.3. *The state space X is bounded and the unknown input signal has* unlimited energy*, i.e., $\lim\limits_{k\to\infty} \|d^{q*}_{0:k}\|_2 = \infty$, where $d^{q*}_{0:k} \triangleq \begin{bmatrix} d^{q*\top}_k & d^{q*\top}_{k-1} & \ldots d^{q*\top}_0 \end{bmatrix}^\top$.*

Note that the unlimited energy condition in Assumption 5.3.3 is not restrictive if $f(\cdot)$, B, C and D are mode-independent, since otherwise, the unknown input signal must vanish asymptotically, which means that we effectively have a non-switched system in the limit and the mode estimation would be trivial.

Next, in order to derive the desired sufficient conditions for mode-detectability in Theorem 5.3.7, we first present the following Lemmas 5.3.4–5.3.6.

Lemma 5.3.4. *For each mode q,*

$$\lim_{k\to\infty} \delta^{q,inf}_{r,k} = \infty. \tag{5.12}$$

$$\lim_{k\to\infty} \hat{\delta}^{q}_{r,k} = \lim_{k\to\infty} \delta^{q,tri}_{r,k} < \infty, \tag{5.13}$$

Lemma 5.3.5. *Suppose that Assumption 5.3.2 holds. Consider two different modes $q \neq q' \in Q$ and their corresponding upper bounds for their residuals' norms, $\delta^q_{r,k}$ and $\delta^{q'}_{r,k}$, at time step k. At least one of the two modes $q \neq q'$ will be eliminated if*

$$\|C^q_2 \hat{x}^{\star,q}_{k|k} - C^{q'}_2 \hat{x}^{\star,q'}_{k|k} + D^q_2 u^q_k - D^{q'}_2 u^{q'}_k\|_2 > \delta^q_{r,k} + \delta^{q'}_{r,k} + R^{q,q'}_z, \tag{5.14}$$

where $R^{q,q'}_z \triangleq R_y \|T^q_2 - T^{q'}_2\|_2$.

Lemma 5.3.6. *Consider any mode q with the unknown true mode being q^*. Suppose without loss of generality that $f^q(0) = 0$. Then, at time step k, we have*

$$r_k^q = \mathbb{A}_k^q t_k^q + \alpha_k^{q^*} + \epsilon_k^{q^*}, \tag{5.15}$$

with $\varepsilon_k^{q^*}$ being an error term that satisfies

$$\exists \xi_1, \dots, \xi_k \in X, \ s.t. \ \|\varepsilon_k^{q^*}\|_2 \leq \frac{1}{2} \sum_{i=1}^{k} \|J_{f,0}^{q^*}\|_2^{k-i} \|x_{i-1}\|_2^2 \|H_f^{q^*}(\xi_i)\|_2, \tag{5.16}$$

where

$$\begin{aligned}
\alpha_k^{q^*} &\triangleq (T_2^q - T_2^{q^*})(C_{f,k}^{q^*} x_0 + C_{d,k}^{q^*} d_{0:k}^{q^*} + C_{u,k}^{q^*} u_{0:k}^{q^*} + C_{\tilde{w},k}^{q^*} \tilde{w}_{0:k}^{q^*}) \\
C_{d,k}^{q^*} &\triangleq \begin{bmatrix} H^{q^*} & C^{q^*} G^{q*} & C^{q^*} J_{f,0}^{q^*} G^{q^*} & \dots & C^{q^*} (J_{f,0}^{q^*})^{k-1} G^{q^*} \end{bmatrix}, \\
C_{u,k}^{q^*} &\triangleq \begin{bmatrix} D^{q^*} & C^{q^*} B^{q*} & C^{q^*} J_{f,0}^{q^*} B^{q^*} & \dots & C^{q^*} (J_{f,0}^{q^*})^{k-1} B^{q^*} \end{bmatrix}, \\
C_{\tilde{w},k}^{q^*} &\triangleq \begin{bmatrix} I & C^{q^*} W^{q*} & C^{q^*} J_{f,0}^{q^*} W^{q^*} & \dots & C^{q^*} (J_{f,0}^{q^*})^{k-1} W^{q^*} \end{bmatrix}, \\
d_{0:k}^{q^*} &\triangleq \begin{bmatrix} d_k^{q^* \top} & \dots & d_0^{q^* \top} \end{bmatrix}^\top, u_{0:k}^{q^*} \triangleq \begin{bmatrix} u_k^{q* \top} & \dots & u_0^{q* \top} \end{bmatrix}^\top, C_{f,k}^{q^*} \triangleq C^{q^*} (J_{f,0}^{q^*})^k, \\
\tilde{w}_{0:k}^{q^*} &\triangleq \begin{bmatrix} v_k^{q^* \top} & w_{k-1}^{q^* \top} & \dots & w_0^{q^* \top} \end{bmatrix}^\top, \epsilon_k^{q^*} \triangleq (T_2^q - T_2^{q^*}) \varepsilon_k^{q^*},
\end{aligned}$$

and $J_{f,0}^{q^*}$ and $H_f^{q^*}(\xi)$ are the Jacobian and Hessian matrices of the vector field $f^{q^*}(\cdot)$ at 0 and ξ, respectively.

Theorem 5.3.7 (Sufficient Conditions for Mode Detectability)**.** *System (5.1) is mode detectable, i.e., by applying Algorithm 4, all false modes will be eliminated at some large enough time step K, if the assumptions in Proposition 5.2.1 and either of the following hold:*

i. Assumption 5.3.2 holds and $\forall q, q' \in Q$, $q \neq q'$,

$$\sigma_{min}(W^{q,q'}) > \frac{\overline{\delta}_r^{q,tri} + \overline{\delta}_r^{q',tri} + R_y'^{q,q'}}{\sqrt{R_x^2 + \eta_v^2}},$$

where $W^{q,q'} \triangleq \begin{bmatrix} (C_2^q - C_2^{q'}) & (T_2^q - T_2^{q'}) & -I & I & D_2^q & -D_2^{q'} \end{bmatrix}.$

ii. Assumption 5.3.3 holds and $T_2^q \neq T_2^{q'}$ *holds* $\forall q, q' \in Q, q \neq q'$*. Moreover,* $H_f^{q^*}(\cdot)$ *is bounded on* X *and* $\|J_{f,0}^{q^*}\|_2 < 1$.

5.4 Simulation Results

In this section, we evaluate the effectiveness of our Simultaneous Mode, Input, and State Set-Valued Observer (SMIS), by comparing its performance with the LMI-based $\mathcal{H}_\infty$-observer in [141] that obtains point state estimates. For comparison, we apply the two observers on a modified version of the discrete-time nonlinear switched system in [141], where we increase the number of modes (subsystems) to five, i.e., $Q = 5$. The considered system is in the form of (5.1), with the following parameters: $n = l = 2, m = p = 1$ and $\forall q = 1, \ldots, 5$:

$$B^q = D^q = 0_{2\times 1},\ f^q(x) = \tilde{A}^q \gamma(x) + \hat{A}^q x,$$

where $\gamma(x) \triangleq \frac{1}{2}\begin{bmatrix}\sin(x_1) & \sin(x_2)\end{bmatrix}^\top$. Moreover,

$$\hat{A}^1 = \begin{bmatrix} 0.3 & 0 \\ 0.4 & -0.7 \end{bmatrix}, \tilde{A}^1 = \begin{bmatrix} 0.8 & -0.4 \\ 0.4 & -0.8 \end{bmatrix}, C^1 = \begin{bmatrix} 0.8 & 0.1 \\ 0.8 & 0.1 \end{bmatrix}, H^1 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, G^1 = \begin{bmatrix} 0.4 \\ -0.1 \end{bmatrix},$$

$$\hat{A}^2 = \begin{bmatrix} -0.5 & 0 \\ 1 & -0.5 \end{bmatrix}, \tilde{A}^2 = \begin{bmatrix} 0.6 & -0.1 \\ 0.1 & -0.6 \end{bmatrix}, C^2 = \begin{bmatrix} 0.5 & -0.1 \\ 0.6 & -0.1 \end{bmatrix}, H^2 = \begin{bmatrix} 0.6 \\ -0.5 \end{bmatrix}, G^2 = \begin{bmatrix} -0.2 \\ 0.1 \end{bmatrix},$$

$$\hat{A}^3 = \begin{bmatrix} 0.6 & -0.2 \\ -0.4 & 0.7 \end{bmatrix}, \tilde{A}^3 = \begin{bmatrix} 0.4 & -0.8 \\ -0.2 & -0.4 \end{bmatrix}, C^3 = \begin{bmatrix} 0.2 & 0.7 \\ -0.8 & 0.2 \end{bmatrix}, H^3 = \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, G^3 = \begin{bmatrix} 0.5 \\ 0.2 \end{bmatrix},$$

$$\hat{A}^4 = \begin{bmatrix} -0.6 & -0.2 \\ 0.4 & 0.7 \end{bmatrix}, \tilde{A}^4 = \begin{bmatrix} -0.4 & 0.9 \\ 0.2 & -0.3 \end{bmatrix}, C^4 = \begin{bmatrix} 0.3 & -0.7 \\ 0.8 & -0.6 \end{bmatrix}, H^4 = \begin{bmatrix} -0.4 \\ 0.9 \end{bmatrix}, G^4 = \begin{bmatrix} 0.9 \\ 0.3 \end{bmatrix},$$

$$\hat{A}^5 = \begin{bmatrix} -0.2 & 0.9 \\ -0.1 & 0.3 \end{bmatrix}, \tilde{A}^5 = \begin{bmatrix} -0.8 & 0.1 \\ 0.3 & -0.7 \end{bmatrix}, C^5 = \begin{bmatrix} -0.3 & -0.1 \\ -0.8 & 1 \end{bmatrix}, H^5 = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}, G^5 = \begin{bmatrix} 0.6 \\ 0.1 \end{bmatrix}.$$

The initial state estimate and noise signals satisfy $\|x_0\|_2 \leq \delta_x = 0.5$, $\|w_k\|_2 \leq \eta_w = 0.02$ and $\|v_k\|_2 \leq \eta_w = 0.02$. Furthermore, we assume that $\hat{x}_{0|0} = \begin{bmatrix} 0.4 & 0.4 \end{bmatrix}^\top$.

We consider two scenarios for the unknown input. In the first (Scenario I), the unknown input is a random signal with bounded norm, i.e., $\|d_k\|_2 \leq 0.4$, while d_k in the second scenario (Scenario II) is a time-varying signal that becomes unbounded as time increases. As is demonstrated in Figure 5.1, in the first scenario, i.e., with bounded unknown inputs, the set estimates of our approach (i.e., SMIS estimates) converge to steady-state values and the point estimates of the approach in [141] are within the predicted upper bounds and exhibit convergent behavior. More interestingly, considering the second scenario, i.e., with unbounded unknown inputs, Figure 5.2 shows that our set-valued estimates still converge, i.e., our observer remains stable, while the estimates of the approach in [141] exceed the boundaries of the compatible sets of states and inputs of our approach after some time steps and display a divergent behavior (cf. Figure 5.2).

Further, Tables 5.1 and 5.2 show the matrix T_2^q for each mode q for Scenarios I and II, respectively. It can be verified that the second set of sufficient conditions in Theorem 5.3.7 holds, i.e., $T_2^q \neq T_2^{q'}$ for all $q \neq q'$, for both scenarios. Hence, we expect that all false modes are eliminated, i.e., exactly one (true) mode remains, after some large enough time in both scenarios, which is indeed what we observe in Figures 5.1 and 5.2, where the number of non-eliminated modes at each time step is shown.

Moreover, for each mode q, the signals $\|r_k^q\|_2$, $\|r_k^{q|*}\|_2$, $\delta_{r,k}^{q,tri}$ and $\delta_{r,k}^{q,inf}$ are depicted in Figures 5.3 and 5.4 for Scenarios I and II, respectively. In both scenarios, we observe that $\delta_{r,k}^{q,inf}$ is smaller than $\delta_{r,k}^{q,tri}$ up until approximately 40 time steps, after which $\delta_{r,k}^{q,tri}$ is smaller/tighter. This is one of the main reasons we considered the minimum of both as the threshold in our mode elimination algorithm (also see Remark 5.2.8). Furthermore, for all modes, $\delta_{r,k}^{q,tri}$ is eventually convergent while $\delta_{r,k}^{q,inf}$ diverges, as

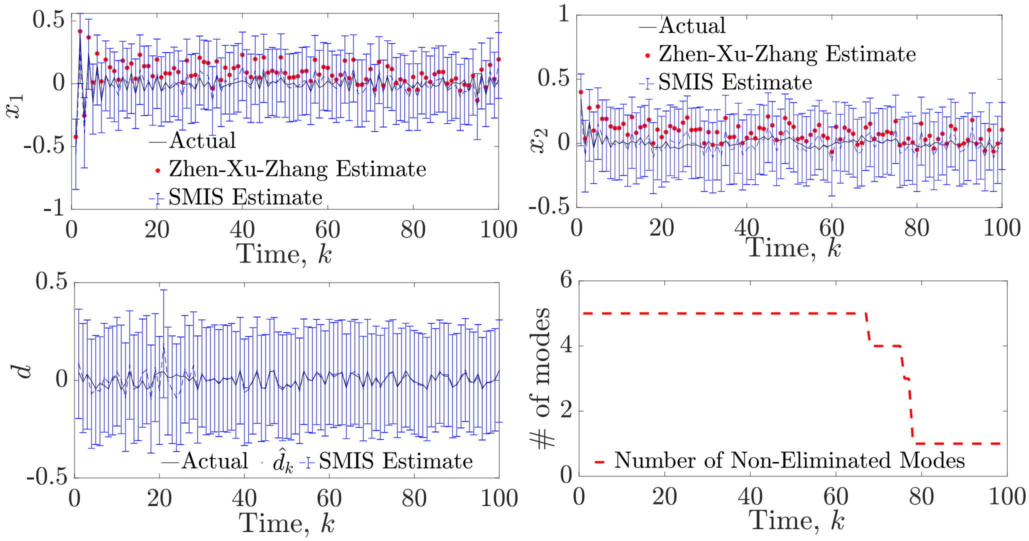


Figure 5.1: Actual states x_1, x_2, and their estimates, as well as the unknown input d and its estimates, and the number of non-eliminated modes at each time step in the bounded unknown input scenario (Scenario I), when applying the observer in [141] (Zhen-Xu-Zhang Estimate) and our proposed observer (SMIS Estimate)

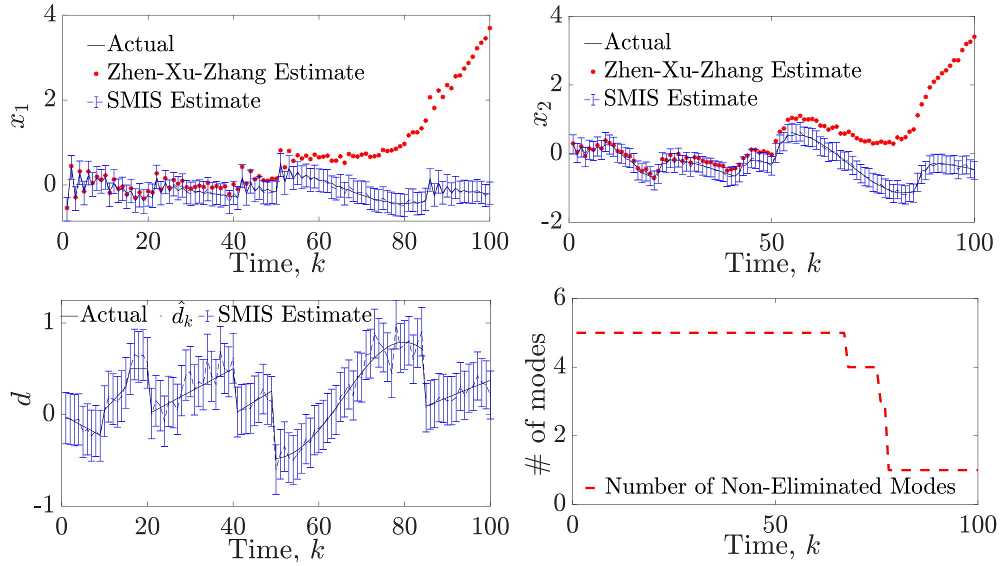


Figure 5.2: Actual States x_1, x_2, and Their Estimates, as Well as the Unknown Input d and its Estimates, and the Number of Non-eliminated Modes at Each Time Step in the Unbounded Unknown Input Scenario (Scenario II), When Applying the Observer in [141] (Zhen-Xu-Zhang Estimate) and Our Oroposed Observer (SMIS Estimate)

Table 5.1: Different Modes and Their T_2^q in Scenario I (i.e., with Bounded d_k)

Mode	T_2^q
$q = 1$	$[0.3629\ -0.2179\ ]^\top$
$q = 2$	$[0.1191\ 0.8715\ ]^\top$
$q = 3$	$[-0.6468\ 0.8390\ ]^\top$
$q = 4$	$[0.8103\ -0.6681\ ]^\top$
$q = 5$	$[0.2780\ -0.6793\ ]^\top$

Table 5.2: Different Modes and Their T_2^q in Scenario II (i.e., with Unbounded d_k)

Mode	T_2^q
$q = 1$	$[0.4730\ -0.3280\ ]^\top$
$q = 2$	$[0.2202\ 0.9826\ ]^\top$
$q = 3$	$[-0.7579\ 0.9401\ ]^\top$
$q = 4$	$[0.9214\ -0.7792\ ]^\top$
$q = 5$	$[0.3891\ -0.7804\ ]^\top$

proven in Lemma 5.3.4. So, after some large enough time, $\delta_{r,k}^{q,tri}$ can be used as our upper bound threshold, while $\delta_{r,k}^{q,inf}$ becomes ineffective.

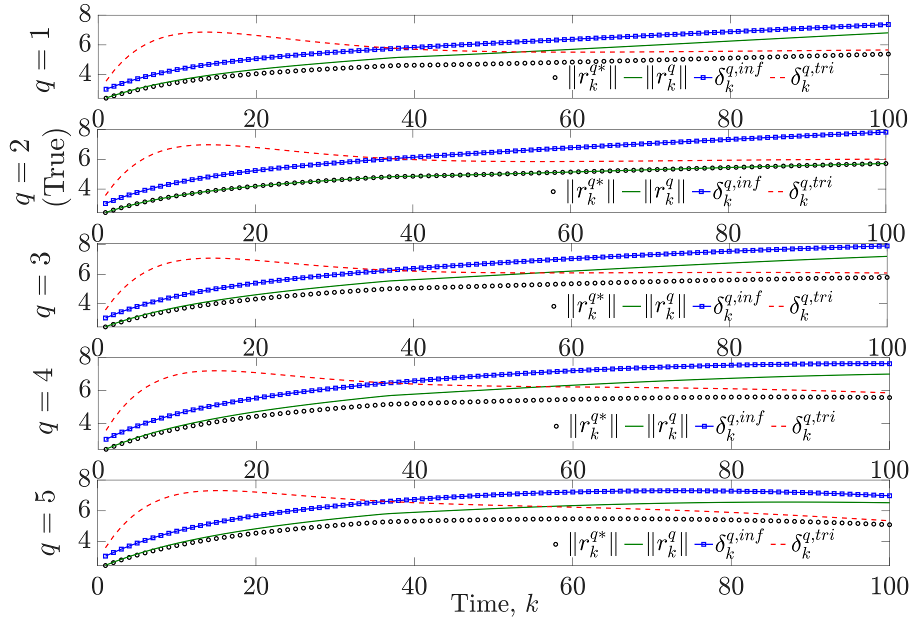


Figure 5.3: $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes in the Bounded Unknown Input Scenario (Scenario I)

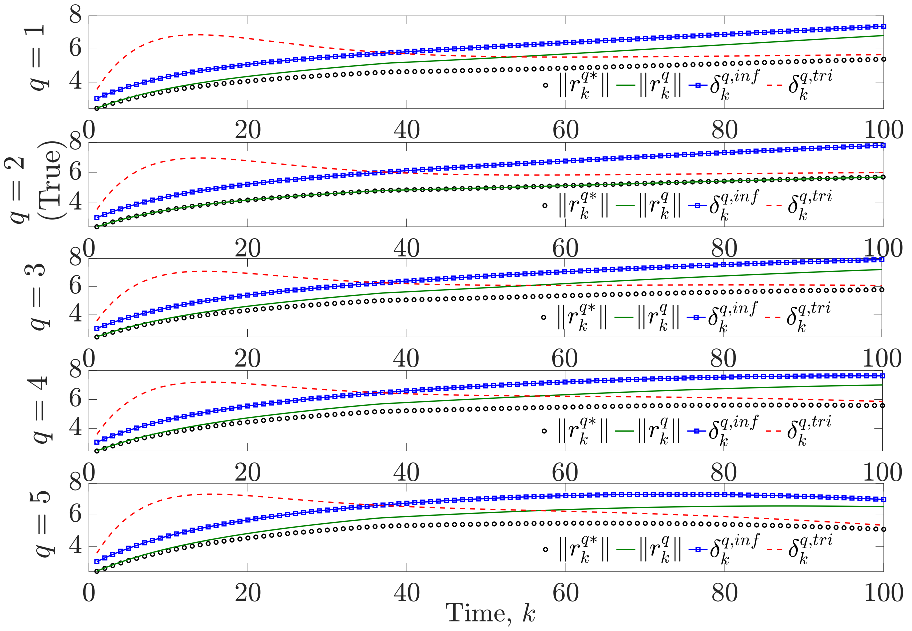


Figure 5.4: $\|r^q_{r,k}\|_2$,$\|r^{q|*}_{r,k}\|_2$ and Their Upper Bounds for Different Modes in the Unbounded Unknown Input Scenario (Scenario II)

5.5 Conclusion

This paper introduced a novel multiple-model approach for simultaneous mode, unknown input and state estimation for hidden mode switched nonlinear systems with bounded-norm noise and unknown inputs. The proposed approach consists of a bank of mode-matched state and unknown input observer that is optimal in the $\mathcal{H}_\infty$ sense and a mode observer, which eliminates inconsistent modes and their corresponding observers at each time step. The proposed mode elimination criterion is based on the use of a provably finite-valued upper bound for the norm of a residual signal as the threshold. Moreover, we provided a tractable approach to compute the threshold signal and proved the convergence of the upper bound/threshold signal as well as derived sufficient conditions for eventually eliminating all false modes when using our mode elimination algorithm. Finally, we demonstrated the effectiveness of our observer using an illustrative example, where we compared our approach with an existing $\mathcal{H}_\infty$ observer in the literature under two different scenarios.

Chapter 6

SIMULTANEOUS INPUT AND STATE INTERVAL OBSERVERS FOR NONLINEAR SYSTEMS

In this chapter [a] , we address the problem of designing simultaneous input and state interval observers for Lipschitz continuous nonlinear systems with unknown inputs and bounded noise signals. Benefiting from the existence of nonlinear decomposition functions and affine abstractions, our proposed observer recursively computes the maximal and minimal elements of the estimate intervals that are proven to contain the true states and unknown inputs, and leverages the output/measurement signals to shrink the intervals by eliminating estimates that are incompatible with the measurements. Moreover, we provide sufficient conditions for the existence and stability (i.e., uniform boundedness of the sequence of estimate interval widths) of the designed observer, and show that the input interval estimates are tight, given the state intervals and decomposition functions.

6.1 Preliminary Material

Definition 6.1.1 (Interval, Maximal and Minimal Elements, Interval Width)**.** *An (multi-dimensional) interval* $\mathcal{I} \subset \mathbb{R}^n$ *is the set of all real vectors* $x \in \mathbb{R}^n$ *that satisfies* $\underline{s} \leq x \leq \overline{s}$*, where* $\underline{s}$*,* $\overline{s}$ *and* $\|\overline{s} - \underline{s}\|$ *are called minimal vector, maximal vector and width of* $\mathcal{I}$*, respectively.*

Next, we will briefly restate our previous result in [108], tailoring it specifically for intervals to help with computing affine bounding functions for our vector fields.

---

[a] The content of this chapter is documented as a published paper in [117] and an accepted paper in [120].

Proposition 6.1.2. *[108, Affine Abstraction] Consider the vector field* $f(.) : \mathcal{B} \subset \mathbb{R}^n \to \mathbb{R}^m$*, where* $\mathcal{B}$ *is an interval with* $\overline{x}, \underline{x}, \mathcal{V}_{\mathcal{B}}$ *being its maximal, minimal and set of vertices, respectively. Suppose* $\overline{A}_{\mathcal{B}}, \underline{A}_{\mathcal{B}}, \overline{e}_{\mathcal{B}}, \underline{e}_{\mathcal{B}}, \theta_{\mathcal{B}}$ *is a solution of the following linear program (LP):*

$$\begin{aligned} &\min_{\theta, \overline{A}, \underline{A}, \overline{e}, \underline{e}} \theta \\ &s.t\ \underline{A}x_s + \underline{e} + \sigma \leq f(x_s) \leq \overline{A}x_s + \overline{e} - \sigma, \\ &(\overline{A} - \underline{A})x_s + \overline{e} - \underline{e} - 2\sigma \leq \theta \mathbf{1}_m,\ \forall x_s \in \mathcal{V}_{\mathcal{B}}, \end{aligned} \tag{6.1}$$

where $\mathbf{1}_m \in \mathbb{R}^m$ *is a vector of ones and* σ *can be computed via [108, Proposition 1] for different function classes. Then,* $\underline{A}x + \underline{e} \leq f(x) \leq \overline{A}x + \overline{e}, \forall x \in \mathcal{B}$*. We call* $\overline{A}, \underline{A}$ *upper and lower affine abstraction slopes of function* $f(.)$ *on* $\mathcal{B}$*.*

Corollary 6.1.3. *By taking the average of upper and lower affine abstractions and adding/subtracting half of the maximum distance, it is straightforward to parallelize the above upper and lower abstractions as* $Ax + (1/2)(\overline{e} + \underline{e} - \theta\mathbf{1}_m) \leq f(x) \leq Ax + (1/2)(\overline{e} + \underline{e} + \theta\mathbf{1}_m)$*, where* $A = (1/2)(\overline{A} + \underline{A})$*.*

Proposition 6.1.4. *[41, Lemma 1] Let* $A \in \mathbb{R}^{m\times n}$ *and* $\underline{x} \leq x \leq \overline{x} \in \mathbb{R}^n$*. Then,* $A^+\underline{x} - A^{++}\overline{x} \leq Ax \leq A^+\overline{x} - A^{++}\underline{x}$*. As a corollary, if* A *is non-negative,* $A\underline{x} \leq Ax \leq A\overline{x}$*.*

Lemma 6.1.5. *Suppose the assumptions in Proposition 6.1.4 hold. Then, the returned bounds for* Ax *is tight, in the sense that* $\sup_{\underline{x}\leq x\leq\overline{x}} Ax = A^+\overline{x} - A^{++}\underline{x}$ *and* $\inf_{\underline{x}\leq x\leq\overline{x}} Ax = A^+\underline{x} - A^{++}\overline{x}$*, where* sup *and* inf *are considered element-wise.*

Definition 6.1.6 (Lipschitz Continuity)**.** *function* $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^m$ *is* L_f*-Lipschitz continuous on* $\mathbb{R}^n$*, if* $\exists L_f \in \mathbb{R}_{++}$*, such that* $\|f(x_1) - f(x_2)\| \leq L_f\|x_1 - x_2\|, \forall x_1, x_2 \in \mathbb{R}^n$*.*

Definition 6.1.7 (Mixed-Monotone Mappings and Decomposition Functions)**.** *[128, Definition 4] A mapping $f : \mathcal{X} \subseteq \mathbb{R}^n \to \mathcal{T} \subseteq \mathbb{R}^m$ is mixed monotone if there exists a decomposition function $f_d : \mathcal{X} \times \mathcal{X} \to \mathcal{T}$ satisfying:*

1. $f_d(x, x) = f(x)$,
2. $x_1 \geq x_2 \Rightarrow f_d(x_1, y) \geq f_d(x_2, y)$ *and*
3. $y_1 \geq y_2 \Rightarrow f_d(x, y_1) \leq f_d(x, y_2)$.

Proposition 6.1.8. *[32, Theorem 1] Let $f : \mathcal{X} \subseteq \mathbb{R}^n \to \mathcal{T} \subseteq \mathbb{R}^m$ be a mixed monotone mapping with decomposition function $f_d : \mathcal{X} \times \mathcal{X} \to \mathcal{T}$ and $\underline{x} \leq x \leq \overline{x}$, where $\underline{x}, x, \overline{x} \in \mathcal{X}$. Then $f_d(\underline{x}, \overline{x}) \leq f(x) \leq f_d(\overline{x}, \underline{x})$.*

Due to non-uniqueness of the decomposition function of a function, a specific one is given in [128, Theorem 2]: If a vector field $q = \begin{bmatrix} h_1^\top & \dots & q_n^\top \end{bmatrix}^\top : X \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable and its partial derivatives are bounded with known bounds, i.e., $\frac{\partial q_i}{\partial x_j} \in (a^q_{i,j}, b^q_{i,j}), \forall x \in X \in \mathbb{R}^n$, where $a^q_{i,j}, b^q_{i,j} \in \overline{\mathbb{R}}$, then h is mixed monotone with a decomposition function $q_d = \begin{bmatrix} q_{d1}^\top & \dots & q_{di}^\top & \dots q_{dn}^\top \end{bmatrix}^\top$, where $q_{di}(x, y) = q_i(z) + (\alpha^q_i - \beta^q_i)^\top (x - y), \forall i \in \{1, \dots, n\}$, and $z, \alpha^q_i, \beta^h_i \in \mathbb{R}^n$ can be computed in terms of $x, y, a^q_{i,j}, b^q_{i,j}$ as given in [128, (10)–(13)]. Consequently, for $x = [x_1 \dots x_j \dots x_n]^\top$, $y = [y_1 \dots y_j \dots y_n]^\top$, we have

$$q_d(x, y) = q(z) + C^q (x - y), \tag{6.2}$$

where $C^q \triangleq \begin{bmatrix} [\alpha^q_1 - \beta^q_1] & \dots & [\alpha^q_i - \beta^q_i] & \dots [\alpha^q_m - \beta^q_m] \end{bmatrix}^\top \in \mathbb{R}^{m \times n}$, with α^q_i, β^q_i given in [128, (10)–(13)], $z = [z_1 \dots z_j \dots z_m]^\top$ and $z_j = x_j$ or y_j (dependent on the case, cf. [128, Theorem 1 and (10)–(13)] for details). Moreover, if exact values of $a_{i,j}, b_{i,j}$ are unknown, their approximations can be obtained using Proposition 6.1.2 with the slopes set to 0.

Corollary 6.1.9. *As a direct implication of Propositions 6.1.2–6.1.8, for any Lipschitz mixed-monotone vector-field $q(.) : \mathbb{R}^n \to \mathbb{R}^m$, with a decomposition function $q_d(.,.)$, we can find upper and lower vectors $\overline{q}, \underline{q}$ such that $\underline{q} \leq q(x) \leq \overline{q}, \forall x \in [\underline{x}, \overline{x}]$, and*

$$\underline{q} = \max(q_d(\underline{x}, \overline{x}), \underline{\hat{q}}), \quad \overline{q} = \min(q_d(\overline{x}, \underline{x}), \hat{\overline{q}}),$$

$$\underline{\hat{q}} = (\underline{A}^q)^+ \underline{x} - (\underline{A}^q)^{++} \overline{x} + \underline{e}^q, \hat{\overline{q}} = (\overline{A}^q)^+ \overline{x} - (\overline{A}^q)^{++} \underline{x} + \overline{e}^q,$$

where $(\overline{A}^q, \underline{A}^q, \overline{e}^q, \underline{e}^q)$ is a solution of (6.1) *for the function q.*

Finally, we derive a Lipschitz-like property for the bounding functions in Corollary 6.1.9, which will be used later for determining observer stability.

Lemma 6.1.10. *Let $q(.) : [\underline{x}, \overline{x}] \subset \mathbb{R}^n \to \mathbb{R}^m$ be the Lipschitz mixed-monotone vector-field in Corollary 6.1.9, with its decomposition function $q_d(.,.)$ constructed using* (6.2)*. Then, $\|\overline{q} - \underline{q}\| \leq \|q_d(\overline{x}, \underline{x}) - q_d(\underline{x}, \overline{x})\| \leq L_{q_d} \|\overline{x} - \underline{x}\|$, where $L_{q_d} \triangleq L_q + 2\|C_q\|$, with C_q given in* (6.2)*.*

6.2 Problem Formulation

System Assumptions. Consider the nonlinear discrete-time system with unknown inputs and bounded noise

$$\begin{aligned} x_{k+1} &= f(x_k) + Bu_k + Gd_k + w_k, \\ y_k &= g(x_k) + Du_k + Hd_k + v_k, \end{aligned} \tag{6.3}$$

where at time $k \in \mathbb{N}$, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $d_k \in \mathbb{R}^p$ and $y_k \in \mathbb{R}^l$ are the state vector, a known input vector, an unknown input vector, and the measurement vector, correspondingly. The process and measurement noise signals $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^l$ are assumed to be bounded, with $\underline{w} \leq w_k \leq \overline{w}$, $\underline{v} \leq v_k \leq \overline{v}$, and the known lower and upper bounds, $\underline{w}$, $\overline{w}$ and $\underline{v}$, $\overline{v}$, respectively. We also assume that lower and upper bounds for the initial state, $\underline{x}_0$ and $\overline{x}_0$, are available, i.e., $\underline{x}_0 \leq x_0 \leq \overline{x}_0$. The vector fields $f(\cdot) : \mathbb{R}^n \to \mathbb{R}^n$, $g(\cdot) : \mathbb{R}^n \to \mathbb{R}^l$ and matrices B, D, G and H are known and of appropriate dimensions, where G and H encoding the *locations* through which the unknown input (or attack) signal can affect the system dynamics and measurements. Note that no assumption is made on H to be either the zero matrix (no direct feedthrough), or to have full column rank when there is direct feedthrough (in contrast to [117]). Moreover, we assume the following, which is satisfied for a broad range of nonlinear functions [129]:

Assumption 6.2.1. *Vector fields $f(\cdot)$ and $g(\cdot)$ are mixed-monotone with decomposition functions $f_d(\cdot,\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ and $g_d(\cdot,\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}^l$ and L_f-Lipschitz and L_g-Lipschitz continuous, respectively.*

Unknown Input (or Attack) Signal Assumptions. The unknown inputs d_k are not constrained to follow any model nor to be a signal of any type (random or strategic), hence no prior 'useful' knowledge of the dynamics of d_k is available (independent of $\{d_\ell\}$ $\forall k \neq \ell$, $\{w_\ell\}$ and $\{v_\ell\}$ $\forall \ell$). We also do not assume that d_k is bounded or has known bounds and thus, d_k is suitable for representing adversarial attack signals.

Next, we briefly introduce a similar system transformation as in [131], which will be used later in our observer structure.

System Transformation. Let $p_H \triangleq \text{rk}(H)$. Similar to [131], by applying singular value decomposition, we have $H = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1{}^\top \\ V_2{}^\top \end{bmatrix}$ with $V_1 \in \mathbb{R}^{p\times p_H}$, $V_2 \in \mathbb{R}^{p\times(p-p_H)}$, $\Sigma \in \mathbb{R}^{p_H\times p_H}$ (a diagonal matrix of full rank), $U_1 \in \mathbb{R}^{l\times p_H}$ and $U_2 \in \mathbb{R}^{l\times(l-p_H)}$. Then, since $V \triangleq \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ is unitary:

$$d_k = V_1 d_{1,k} + V_2 d_{2,k},\ d_{1,k} = V_1{}^\top d_k,\ d_{2,k} = V_2{}^\top d_k. \tag{6.4}$$

Finally, by defining $T_1 \triangleq U_1^\top, T_2 \triangleq U_2^\top$, the output equation can be decoupled as:

$$z_{1,k} = g_1(x_k) + D_1 u_k + v_{1,k} + \Sigma d_{1,k}, \tag{6.5}$$

$$z_{2,k} = g_2(x_k) + D_2 u_k + v_{2,k}, \tag{6.6}$$

$$g_1(x,k) \triangleq T_1 g(x_k), g_2(x_k) \triangleq T_2 g(x_k). \tag{6.7}$$

The observer design problem can be stated as follows:

Problem 6.2.2. *Given a nonlinear discrete-time system with unknown inputs and bounded noise* (6.3)*, design a stable observer that simultaneously finds bounded intervals of compatible states and unknown inputs.*

6.3 General Simultaneous Input and State Interval Observers (GSISIO)

6.3.1 Interval Observer Design

We consider a recursive three-step interval-valued observer design, composed of a *state propagation* (SP) step, which propagates the previous time state estimates through the state equation to find propagated intervals, a *measurement update* (MU) step, which iteratively updates the state intervals using the observation, and an *unknown input estimation* (UIE) step, which computes the input intervals using state intervals and observation. We design the observer in the following form:

$$\begin{aligned}
\textit{State Propagation: } \mathcal{I}_k^{x^p} &= \mathcal{F}_x^p(\mathcal{I}_{k-1}^x, y_{k-1}, u_{k-1}), \\
\textit{Measurement Update: } \quad \mathcal{I}_k^x &= \mathcal{F}_x(\mathcal{I}_k^{x^p}, y_k, u_k), \\
\textit{Unknown Input Estimation: } \mathcal{I}_{k-1}^d &= \mathcal{F}_d(\mathcal{I}_k^x, y_{k-1}, u_{k-1}),
\end{aligned}$$

where $\mathcal{F}_x^p$, $\mathcal{F}_x$ and $\mathcal{F}_d$ are to-be-designed interval mappings, while $\mathcal{I}_k^{x^p}$, $\mathcal{I}_k^x$ and $\mathcal{I}_{k-1}^d$ are intervals of compatible propagated states, updated states and unknown inputs at time steps k, k and $k-1$, respectively. Note that we are constrained with obtaining a one-step delayed estimate of $\mathcal{I}^d_{k-1}$, because in contrast with [117], the matrix H is not necessarily full-rank, and hence d_k cannot be estimated from the current measurement, y_k. However, in Lemma 6.3.6 and Remark 6.3.7, we will discuss a way of obtaining the current estimate of *a component of* the input signal, i.e., $d_{1,k}$ in (6.5).

Considering the computational complexity of optimal observers [82], as well as nice properties of interval sets [42], we consider set estimates of the form:

$$\begin{aligned}\mathcal{I}^{x^p}_k &= \{x \in \mathbb{R}^n : \underline{x}^p_k \leq x \leq \overline{x}^p_k\},\\ \mathcal{I}^x_k &= \{x \in \mathbb{R}^n : \underline{x}_k \leq x \leq \overline{x}_k\},\\ \mathcal{I}^d_{k-1} &= \{d \in \mathbb{R}^p : \underline{d}_{k-1} \leq d \leq \overline{d}_{k-1}\},\end{aligned}$$

i.e., we restrict the estimation errors to be closed intervals. In this case, the observer design problem boils down to finding $\underline{x}^p_k$, $\overline{x}^p_k$, $\underline{x}_k$, $\overline{x}_k$, $\underline{d}_{k-1}$ and $\overline{d}_{k-1}$. Our interval observer can be defined at each time step $k \geq 1$ as follows (with known $\underline{x}_0$ and $\overline{x}_0$ such that $\underline{x}_0 \leq x_0 \leq \overline{x}_0$):

State Propagation (SP):

$$\begin{aligned}\begin{bmatrix}\overline{x}^{p\top}_k & \underline{x}^{p\top}_k\end{bmatrix}^\top = M_f \begin{bmatrix}\overline{f}^\top_k & \underline{f}^\top_k\end{bmatrix}^\top + M_g \begin{bmatrix}\overline{g}^\top_k & \underline{g}^\top_k\end{bmatrix}^\top + \omega^p + \\ M_v \begin{bmatrix}\overline{v}^\top & \underline{v}^\top\end{bmatrix}^\top + M_w \begin{bmatrix}\overline{w}^\top & \underline{w}^\top\end{bmatrix}^\top + M_y y_{k-1} + M_u u_{k-1};\end{aligned} \tag{6.8}$$

Measurement Update (MU):

$$\overline{x}_k = \lim_{i\to\infty} \overline{x}^{*,i}_k, \quad \underline{x}_k = \lim_{i\to\infty} \underline{x}^{*,i}_k; \tag{6.9}$$

Unknown Input Estimation (UIE):

$$\overline{d}_{k-1} = N_{11}\overline{h}_k + N_{12}\underline{h}_k, \quad \underline{d}_{k-1} = N_{21}\overline{h}_k + N_{22}\underline{h}_k, \tag{6.10}$$

where $\forall q \in \{f, g\}$, $\overline{q}_k$ and $\underline{q}_k$ are upper and lower vector values for the function $q(.)$ on the interval $[\underline{x}_{k-1}, \overline{x}_{k-1}]$, which can be recursively computed using Corollary 6.1.9. Moreover,

$$\overline{h}_k = \begin{bmatrix}\overline{x}_k^\top & y_{k-1}^\top\end{bmatrix}^\top - \begin{bmatrix}\underline{f}_k^\top & \underline{g}_k^\top\end{bmatrix}^\top - \begin{bmatrix}B^\top & D^\top\end{bmatrix}^\top u_{k-1} - \begin{bmatrix}\underline{w}^\top & \underline{v}^\top\end{bmatrix}^\top, \tag{6.11}$$

$$\underline{h}_k = \begin{bmatrix}\underline{x}_k^\top & y_{k-1}^\top\end{bmatrix}^\top - \begin{bmatrix}\overline{f}_k^\top & \overline{g}_k^\top\end{bmatrix}^\top - \begin{bmatrix}B^\top & D^\top\end{bmatrix}^\top u_{k-1} - \begin{bmatrix}\overline{w}^\top & \overline{v}^\top\end{bmatrix}^\top. \tag{6.12}$$

Furthermore, $\{\overline{x}_k^{*,i}, \underline{x}_k^{*,i}\}_{i=0}^{\infty}$ are the sequences of *updated state framers*, iteratively computed in the following form

$$\underline{x}_k^{*,0} = \underline{x}_k^p, \quad \overline{x}_k^{*,0} = \overline{x}_k^p, \quad \forall i \in \{1 \dots \infty\} : \tag{6.13}$$

$$\underline{x}_k^{*,i} = \max(\underline{x}_k^{*,i-1}, \underline{x}_k^{u,i}), \overline{x}_k^{*,i} = \min(\overline{x}_k^{*,i-1}, \overline{x}_k^{u,i}), \tag{6.14}$$

where

$$\begin{aligned}
&\underline{x}_k^{u,i} = (A_{i,k}^{\dagger})^{+}\underline{\alpha}_k^i - (A_{i_k}^{\dagger})^{++}\overline{\alpha}_k^i - \omega_{i,k}^u,\\
&\overline{x}_k^{u,i} = (A_{i,k}^{\dagger})^{+}\overline{\alpha}_k^i - (A_{i_k}^{\dagger})^{++}\underline{\alpha}_k^i + \omega_{i,k}^u,\\
&\underline{\alpha}_k^i = \max_{j \in \{1 \dots 3\}} \{\underline{\alpha}_k^{i,j}\}, \overline{\alpha}_k^i = \min_{j \in \{1 \dots 3\}} \{\overline{\alpha}_k^{i,j}\}, \underline{\alpha}_k^{i,1} = \underline{t}_k - \overline{c}_k^i, \overline{\alpha}_k^{i,1} = \overline{t}_k - \underline{c}_k^i,\\
&\underline{\alpha}_k^{i,2} = A_{i,k}^{+}\underline{x}_k^{*,i-1} - A_{i,k}^{++}\overline{x}_k^{*,i-1}, \overline{\alpha}_k^{i,2} = A_{i,k}^{+}\overline{x}_k^{*,i-1} - A_{i,k}^{++}\underline{x}_k^{*,i-1},\\
&\underline{\alpha}_k^{i,3} = g_{2,d}(\underline{x}_{k-1}^{*.i}, \overline{x}_{k-1}^{*.i}) - \overline{c}_k^i, \overline{\alpha}_k^{i,3} = g_{2,d}(\overline{x}_{k-1}^{*.i}, \underline{x}_{k-1}^{*.i}) - \underline{c}_k^i,
\end{aligned}$$

$$\overline{t}_k = z_{2,k} - D_2 u_k - \underline{v}_2, \underline{t}_k = z_{2,k} - D_2 u_k - \overline{v}_2, \tag{6.15}$$

$$\overline{c}_k^i \triangleq (1/2)(\overline{e}_k^i + \underline{e}_k^i + \theta_k^i), \underline{c}_k^i \triangleq (1/2)(\overline{e}_k^i + \underline{e}_k^i - \theta_k^i). \tag{6.16}$$

Finally, ω_k^p, M_s, N_{nm}, $\forall s \in \{f, g, u, w, v, y\}, n, m \in \{1, 2\}$, $\omega_{i,k}^u, A_{i,k}, \overline{e}_k^i, \underline{e}_k^i, \theta_k^i$, $\forall i \in \{1 \dots \infty\}$ and $g_{2d}(.,.)$ are to-be-designed observer parameters, matrix gains (with appropriate dimensions) and bounding function, at time k and iteration i with the purpose of achieving desirable observer properties.

Note that the measurement update step is iterative (see proof of Theorem 6.3.4 for a more detailed explanation) because the tightness of the upper and lower bounding functions for the observation function g_2 (cf. Propositions 6.1.2 and 6.1.8) is dependent on the *a priori* interval $\mathcal{B}$. Thus, starting from the compatible intervals from the

Algorithm 5 Generalized Simultaneous Input and Sate Interval Observer (GSISIO)

```
Initialize: maximal($\mathcal{I}^x_0$) = $\overline{x}_0$; minimal($\mathcal{I}^x_0$) = $\underline{x}_0$;
    ▷ Observer Gains Computation
    Compute $M_s, N_{ij}, \forall s \in \{f, g, u, v, w\}, i, j \in \{1, 2\}$ via Theorem 6.3.4;
for $k = 1$ to $\overline{K}$ do
    ▷ Estimation of $x_k$
      Compute $\overline{x}^p_k, \underline{x}^p_k$ via (6.8); Compute $\{\overline{x}^{*,i}, \underline{x}^{*,i}\}_{i=0}^{\infty}$ via (6.13),(6.14);
      $(\overline{x}_k, \underline{x}_k) = (\overline{x}^{*,\infty}_k, \underline{x}^{*,\infty}_k)$; $\mathcal{I}^x_k = \{x \in \mathbb{R}^n : \underline{x}_k \leq x \leq \overline{x}_k\}$;
      Compute $\delta^x_k$ through Lemma 6.3.10;
    ▷ Estimation of $d_{k-1}$
      Compute $\overline{d}_{k-1}, \underline{d}_{k-1}, \delta^d_{k-1}$ via (6.10)–(6.12) and Lemma 6.3.10;
      $\mathcal{I}^d_{k-1} = \{d \in \mathbb{R}^p : \underline{d}_{k-1} \leq d \leq \overline{d}_{k-1}\}$;
end for
```

propagation step, if we obtain tighter updated intervals, they can be used as the new $\mathcal{B}$ to obtain better bounding functions for g_2, which in turn may lead to even tighter updated intervals. This process can be repeated and results in a sequence of monotonically tighter updated intervals, where its limit (that exists by the monotone convergence theorem) is chosen as the final interval estimate at time k. Algorithm 5 summarizes GSISIO.

6.3.2 Observer Design

The objective of this section is to design observer gains such that the GSISIO returns *correct* and *tight* intervals. We first define these properties through the following definitions.

Definition 6.3.1 (Correctness (Framer Property [77])). *Given an initial interval $\underline{x}_0 \leq x_0 \leq \underline{x}_0$, the GSISIO observer returns correct interval estimates, if the true* *states and unknown inputs of the system* (6.3) *are within the estimated intervals* (6.8)–(6.10) *for all times. If the observer is correct, we call* $\{\overline{x}^p_k, \underline{x}^p_k\}_{k=0}^{\infty}$, $\{\overline{x}_k, \underline{x}_k\}_{k=0}^{\infty}$ *and* $\{\overline{d}_{k-1}, \underline{d}_{k-1}\}_{k=1}^{\infty}$ *the propagated state, updated state and input framers, respectively.*

Definition 6.3.2 (Tightness of Input Estimates)**.** *The input interval estimates, i.e.,* $\{\mathcal{I}^d_{k-1}(\mathcal{I}^x_k, y_{k-1}, u_{k-1})\}_{k=1}^{\infty}$, *are tight, if at each time step k, given the state estimate* $\mathcal{I}^x_k$, *the input framers* $\overline{d}_{k-1}, \underline{d}_{k-1}$, *coincide with supremum and infimum values of the set of compatible inputs.*

We begin by using the result in Lemma 6.1.5 to conclude the correctness and tightness of the input estimates, assuming that the state estimates are given. To increase readability, all proofs will be provided in the appendix.

Lemma 6.3.3 (Correctness and Tightness of Input Estimates)**.** *Consider the system* (6.3) *along with the GSISIO in* (6.8)–(6.10), *let* $J \triangleq (\begin{bmatrix} G^\top & H^\top \end{bmatrix}^\top)^\dagger$ *and suppose that Assumption 6.2.1 holds,* $N_{11} = N_{22} = J^+$, *and* $N_{12} = N_{21} = -J^{++}$. *Then, given any pair of state framer sequences* $\{\overline{x}_k, \underline{x}_k\}_{i=0}^{\infty}$, *the input interval estimates given in* (6.10), *are correct and tight.*

Next, we state our first main result on the existence of the GSISIO and correctness of the state estimates.

Theorem 6.3.4 (Existence of Correct Framers)**.** *Consider the system* (6.3), *the transformed output equations* (6.5)-(6.7) *and the GSISIO introduced in* (6.8)-(6.10). *Suppose all the assumptions in Lemma 6.3.3 hold and there exists a pair of slope matrices* $(\overline{A}, \underline{A})$, *which construct affine upper and lower abstractions for the vector field* $g_2(.)$ *on the entire state space (cf. Proposition 6.1.2). Suppose that the observer gains are chosen as follows. Then, at each time step k, the GSISO returns finite and correct framers, i.e., finite correct interval estimates for the system* (6.3), *if*

$$r^\top((\mathbb{A}_1 + \mathbb{A}_2)r + \tilde{r}) = 0, \tag{6.17}$$

with $\mathbb{A}_1 \triangleq A^{\dagger +}A^{+} + A^{\dagger ++}A^{++}$*,* $\mathbb{A}_2 \triangleq A^{\dagger +}A^{++} + A^{\dagger ++}A^{+}$*,* $A = (1/2)(\overline{A} + \underline{A})$*,* $\tilde{r} \triangleq rowsupp(I - A^{\dagger}A)$*,* $r \triangleq rowsupp(I - A_x^{\dagger}A_x)_{(1:n)}$ *and* A_x *given below:*

$$\begin{aligned}
&\forall s \in \{f, g, u, w, v, y\} : M_s = A_x^{\dagger}A_s, A_u \triangleq \begin{bmatrix} F^{\top} & F^{\top} \end{bmatrix}^{\top}, A_w = A_f, \\
&A_x \triangleq \begin{bmatrix} I - K_1 & L_1 \\ L_1 & I - K_1 \end{bmatrix}, A_f \triangleq \begin{bmatrix} I + L_1 & -K_1 \\ -K_1 & I + L_1 \end{bmatrix}, A_g \triangleq \begin{bmatrix} L_2 & -K_2 \\ -K_2 & L_2 \end{bmatrix}, \\
&A_v = A_g, L \triangleq G^{++}J^{+} + G^{+}J^{++}, K \triangleq G^{++}J^{++} + G^{+}J^{+}, \\
&K_1 \triangleq K_{(1:n)}, K_2 \triangleq K_{(n+1:n+l)}, L_1 \triangleq L_{(1:n)}, L_2 \triangleq L_{(n+1:n+l)}, \\
&F \triangleq (I + L_1 - K_1)B + (L_2 - K_2)D, A_{i,k} = \frac{1}{2}(\overline{A}_{i,k} + \underline{A}_{i,k}).
\end{aligned}$$

Further, $\omega^p = \mu[r^{\top} - r^{\top}]^{\top}$*,* $g_{2d}(.,.)$ *is a decomposition function of* $g_2(.)$ *and* μ *is a very large positive real number (infinity), while* $\omega^u_{i,k} = \mu\ rowsupp(I - A^{\dagger}_{i,k}A_{i,k})$*, where* $\{\overline{A}_{i,k}, \underline{A}_{i,k}, \overline{e}^i_k, \underline{e}^i_k, \theta^i_k\}$ *is a solution of the LP* (6.1) *for the corresponding vector field* $g_2(x)$ *on the interval* $\mathcal{B}^{*,i}_k = [\underline{x}^{*,i-1}_k, \overline{x}^{*,i-1}_k]$ *with the following extra constraints:*

$$(\overline{A}_{i,k} - \overline{A})x^i_{s,k} + \overline{e}^i_k - \overline{e} \leq 0 \leq (\underline{A}_{i,k} - \underline{A})x^i_{s,k} + \underline{e}^i_k - \underline{e}, \tag{6.18}$$

for all $x^i_{s,k} \in \mathcal{V}_{\mathcal{B}^{*,i}_k}$ *at time* k *and at iteration* $i \in \{1 \ldots \infty\}$*.*

Corollary 6.3.5. *In the case that only the state propagation step is considered, the existence conditions boil down to* $\mathrm{rk}(I - K_1 - L_1) = \mathrm{rk}(I - K_1 + L_1) = n$*.*

Note that we can only obtain a one-step delayed estimate of d_k in (6.10), since we can find an estimate for $d_{1,k}$ at current time k, but not $d_{2,k}$. We formalize this as follows.

Lemma 6.3.6. *Suppose all the assumptions in Theorem 6.3.4 hold. Then, at time step* k*,* $\underline{d}_{1,k} \leq d_{1,k} \leq \overline{d}_{1,k}$*, where* $\overline{d}_{1,k} = \Sigma^{-1}(z_{1,k} - T_1Du_k) + \overline{\ell}_k$*,* $\underline{d}_{1,k} = \Sigma^{-1}(z_{1,k} - T_1Du_k) + \underline{\ell}_k$*,* *with*

$$\overline{\ell}_k \triangleq (\Sigma^{-1}T_1)^{++}(g(\overline{x}_k,\underline{x}_k)+\overline{v}) - (\Sigma^{-1}T)^{+}(g(\underline{x}_k,\overline{x}_k)+\underline{v}),$$

$$\underline{\ell}_k \triangleq (\Sigma^{-1}T_1)^{++}(g(\underline{x}_k,\overline{x}_k)+\underline{v}) - (\Sigma^{-1}T_1)^{+}(g(\overline{x}_k,\underline{x}_k)+\overline{v}).$$

(cf. (6.4)–(6.7)*). Moreover, no current estimate of* $d_{2,k}$ *can be computed.*

Remark 6.3.7. *The result in Lemma 6.3.6 is particularly helpful in the special case when the feedthrough matrix has full rank. In this case,* $d_k = d_{1,k}$ *and hence,* d_k *can be estimated at current time* k*. Thus, this can be considered as an alternative approach to the one in [117] for the full-rank* H *case.*

6.3.3 Uniform Boundedness of Estimates (Observer Stability)

In this section, we derive several sufficient conditions for the stability of GSISIO via Theorem 6.3.8.

Theorem 6.3.8 (Observer Stability)**.** *Consider the system* (6.3) *and the GSISIO* (6.8)–(6.10)*. Suppose all the assumptions in Theorem 6.3.4 hold, the decomposition functions* f_d, g_d *are constructed using* (6.2) *and* $\overline{A}, \underline{A}$ *are the upper and lower affine abstraction slopes for* $g_2(x)$ *on the entire state space. Then, the observer is stable, in the sense that interval width sequences* $\{\|\Delta^d_{k-1}\| \triangleq \|\overline{d}_{k-1} - \underline{d}_{k-1}\|, \|\Delta^x_k\| \triangleq \|\overline{x}_k - \underline{x}_k\|\}_{k=1}^{\infty}$ *are uniformly bounded, and consequently, interval input and state estimation errors* $\{\|\tilde{d}_{k-1}\| \triangleq \max(\|d_{k-1} - \underline{d}_{k-1}\|, \|\overline{d}_{k-1} - d_{k-1}\|), \|\tilde{x}_k\| \triangleq \max(\|x_k - \underline{x}_k\|, \|\overline{x}_k - x_k\|)\}_{k=1}^{\infty}$ *are also uniformly bounded, if either one of the following conditions hold:*

(i) $\hat{\mathcal{L}} \triangleq \min_{\mathbf{D}\in\mathbb{D}^*} L_{f_d}\|\hat{T}_f\| + L_{g_d}\|\hat{T}_g\| \leq 1$,

(ii) $\min_{\mathbf{D}\in\mathbb{D}^*} \lambda_{\max}(\hat{\mathcal{T}}) \leq 0$,

(iii) $\exists P \succ 0, \Gamma \succeq 0, \mathbf{D} \in \mathbb{D}^*$ *such that* $\mathcal{P}_{\mathbf{D}} \preceq 0$,

where $\hat{\mathbf{D}} \triangleq (\mathbf{D}+(I-\mathbf{D})(\mathbb{A}_1+\mathbb{A}_2))$, $\mathbb{D}^* = \{\mathbf{D}^* \in \mathbb{D} \,\big|\, \mathbf{D}^*_{jj} = r'(j)$ *if* $r(j) \neq r'(j), \forall j \in \{1\ldots n\}\}$, $\hat{\mathcal{T}} \triangleq \begin{bmatrix} Q & 0 & 0 & 0 & 0 \\ * & \hat{T}_g^\top \hat{T}_g & \hat{T}_g^\top \hat{T}_f & \hat{T}_g^\top \hat{T}_f & \hat{T}_g^\top \hat{T}_g \\ * & * & \hat{T}_f^\top \hat{T}_f & \hat{T}_f^\top \hat{T}_f & \hat{T}_f^\top \hat{T}_g \\ * & * & * & 0 & \hat{T}_f^\top \hat{T}_g \\ * & * & * & * & 0 \end{bmatrix}$, $\mathcal{P}_{\mathbf{D}} \triangleq \begin{bmatrix} P+\Gamma-I & 0 & P \\ 0 & \mathcal{L}_{\mathbf{D}}^2 I - P & 0 \\ P & 0 & P \end{bmatrix}$, $\hat{T}_f \triangleq \hat{\mathbf{D}} T_f \triangleq \hat{\mathbf{D}}(I-K_1-L_1)^\dagger(I-K_1+L_1)$, $\hat{T}_g \triangleq \hat{\mathbf{D}} T_g \triangleq \hat{\mathbf{D}}(I-K_1-L_1)^\dagger(K_2+L_2)$, $Q \triangleq \lambda_{\max}(\hat{T}_f^\top \hat{T}_f)L_{f_d}^2 + \lambda_{\max}(\hat{T}_g^\top \hat{T}_g)L_{g_d}^2 - 1$, $\mathcal{L}_{\mathbf{D}} \triangleq L_{f_d}\|\hat{T}_f\| + L_{g_d}\|\hat{T}_g\|$, $J, \mathbb{A}_1, \mathbb{A}_2, r, L_{f_d}, L_{g_d}$ *are given in Lemmas 6.1.10–6.3.3 and Theorem 6.3.4,* $\mathbb{D} \in \mathbb{R}^{n\times n}$ *is the set of all diagonal matrices whose diagonal elements are* 0 *or* 1 *and* $\lambda_{\max}(\mathcal{A}^\top \mathcal{A})$ *is the maximum eigenvalue of* $\mathcal{A}^\top \mathcal{A}$.

Remark 6.3.9. *The optimization and feasibility problems in* (i)-(iii) *are all (mixed-)integer programs with finitely countable feasible sets (*$|\mathbb{D}^*| \leq 2^n$*), which can be easily solved by enumerating all possible solutions and comparing the values.*

Finally, we will provide upper bounds for the interval widths and compute their steady-state values, if they exist.

Lemma 6.3.10 (Upper Bounds of the Interval Widths and their Convergence)**.** *Consider the system* (6.3) *and the GSISIO observer* (6.8)–(6.10)*. Suppose all assumptions in Theorem 6.3.4 hold and Condition* (i) *in Theorem 6.3.8 holds with strict inequality. Then, the interval width sequences* $\{\|\Delta^x_k\|, \|\Delta^d_{k-1}\|\}_{k=1}^{\infty}$ *are uniformly upper bounded by the convergent sequences* $\{\delta^x_k, \delta^d_{k-1}\}_{k=1}^{\infty}$*, as follows:*

$$\|\Delta^x_k\| \leq \delta^x_k = \hat{\mathcal{L}}^k \delta^x_0 + \|\tilde{\mathbf{D}}\Delta z\| \left(\frac{1-\hat{\mathcal{L}}^k}{1-\hat{\mathcal{L}}}\right) \xrightarrow{k\to\infty} \frac{\|\tilde{\mathbf{D}}\Delta z\|}{1-\hat{\mathcal{L}}},$$

$$\|\Delta^d_{k-1}\| \leq \delta^d_{k-1} = \mathcal{G}(\delta^x(k)) \xrightarrow{k\to\infty} \overline{\delta}^d = \mathcal{G}(\overline{\delta}^x),$$

where $\tilde{\mathbf{D}}$ *is a solution to* $\min\limits_{\mathbf{D}\in\mathbb{D}^{**}}\|\mathbf{D}\Delta z\|$, $\mathbb{D}^{**}$ *is the solution set of the optimization problem in* (i), $\mathcal{G}(x)\triangleq((1+L_{f_d})\|\hat{J}_1\|+L_{g_d}\|\hat{J}_2\|)x+\|\hat{J}_1\Delta w+\hat{J}_2\Delta v\|$, $\Delta z = T_f\Delta w + T_g\Delta v$, $\Delta w \triangleq \overline{w}-\underline{w}$, $\Delta v \triangleq \overline{v}-\underline{v}$, $\hat{J}\triangleq\begin{bmatrix}\hat{J}_1 & \hat{J}_2\end{bmatrix}\triangleq J^{+}+J^{++}$ *and* L_{f_d}, L_{g_d}, T_f, T_g *are given in Lemma 6.1.10 and Theorem 6.3.8. On the other hand, if Condition* (ii) *or* (iii) *in Theorem 6.3.8 hold, then the interval widths* $\|\Delta^x_k\|$ *and* $\|\Delta^d_k\|$ *are uniformly bounded by* $\min\{\|\Delta^x_0\|,\Delta^P_0\}$ *and* $\min\{\mathcal{G}(\|\Delta^x_0\|),\mathcal{G}((\Delta^P_0)\}$, *respectively, with* $\Delta^P_0 \triangleq \min\limits_{P\in\mathbb{P}}\sqrt{\frac{(\Delta^x_0)^\top P\Delta^x_0}{\lambda_{\min}(P)}}$, *where* $\mathbb{P}$ *is the set of all* P *that solve the LMI in Condition* (iii).

6.4 Simulation Results

We consider a slightly modified version of a nonlinear system in [37], without the uncertain matrices, with the inclusion of unknown inputs, and with the following parameters (cf. (6.3)): $n=l=p=2$, $m=1$, $f(x_k)=\begin{bmatrix}f_1(x_k) & f_2(x_k)\end{bmatrix}^\top$, $g(x_k)=\begin{bmatrix}g_1(x_k) & g_2(x_k)\end{bmatrix}^\top$, $B=D=0_{2\times 1}$, $G=\begin{bmatrix}0 & -0.1\\ 0.2 & -0.2\end{bmatrix}$, $H=\begin{bmatrix}-0.1 & 0.3\\ 0.25 & -0.75\end{bmatrix}$, $\overline{v}=-\underline{v}=\overline{w}=-\underline{w}=\begin{bmatrix}0.2 & 0.2\end{bmatrix}^\top$, $\overline{x}_0=\begin{bmatrix}2 & 1.1\end{bmatrix}^\top$, $\underline{x}_0=\begin{bmatrix}-1.1 & -2\end{bmatrix}^\top$ with

$$\begin{aligned}
f_1(x_k) &= 0.6x_{1,k}-0.12x_{2,k}+1.1\sin(0.3x_{2,k}-.2x_{1,k}),\\
f_2(x_k) &= -0.2x_{1,k}-0.14x_{2,k},\\
g_1(x_k) &= 0.2x_{1,k}+0.65x_{2,k}+0.8\sin(0.3x_{1,k}+0.2x_{2,k}),\\
g_2(x_k) &= \sin(x_{1,k}),
\end{aligned}$$

while the unknown input signals are depicted in Figure 6.1.

Note that $\text{rk}(H)=1<2=p$, thus the feedthrough matrix is not full rank and hence, the approach in [117] is not applicable. Moreover, applying [108, Theorem 1], we can compute finite-valued upper and lower bounds for partial derivatives of $f(\cdot)$

and $g(\cdot)$ as:

$$\begin{bmatrix} a^f_{11} & a^f_{12} \\ a^f_{21} & a^f_{22} \end{bmatrix} = \begin{bmatrix} 0.38 & -0.52 \\ -0.2-\epsilon & -0.14-\epsilon \end{bmatrix}, \begin{bmatrix} b^f_{11} & b^f_{12} \\ b^f_{21} & b^f_{22} \end{bmatrix} = \begin{bmatrix} 0.82 & 0.21 \\ -0.2+\epsilon & -0.14+\epsilon \end{bmatrix},$$
$$\begin{bmatrix} a^g_{11} & a^g_{12} \\ a^g_{21} & a^g_{22} \end{bmatrix} = \begin{bmatrix} -0.04 & 0.49 \\ -1 & -\epsilon \end{bmatrix}, \qquad \begin{bmatrix} b^g_{11} & b^g_{12} \\ b^g_{21} & b^g_{22} \end{bmatrix} = \begin{bmatrix} 0.44 & 0.81 \\ 1 & \epsilon \end{bmatrix},$$

where ϵ is a very small positive value, ensuring that the partial derivatives are in open intervals (cf. [128, Theorem 1]). Moreover, $L_f = 0.35$ and $L_g = 0.74$ and Assumption 6.2.1 holds by [128, Theorem 1]). Furthermore, computing $K = \begin{bmatrix} K_1 & K_2 \end{bmatrix} = \begin{bmatrix} 0.0267 & 0 & 0.0666 & 0.1061 \\ 0.4177 & 2.1203 & 1.0817 & 2.0209 \end{bmatrix}$ and $L = \begin{bmatrix} L_1 & L_2 \end{bmatrix} = \begin{bmatrix} 0 & 0.1017 & 0 & 0 \\ 0.5194 & 1.1814 & 1.2787 & 1.9302 \end{bmatrix}$, we obtain $\text{rk}(I - K_1 - L_1) = \text{rk}(I - K_1 + L_1) = 2$. Therefore, by Corollary 6.3.5 and Theorem 6.3.4, the existence of correct framers is guaranteed, i.e., the true states and unknown inputs are within the estimate intervals. This, can be verified from Figure 6.1 that depicts interval estimates as well as the true states and unknown inputs.

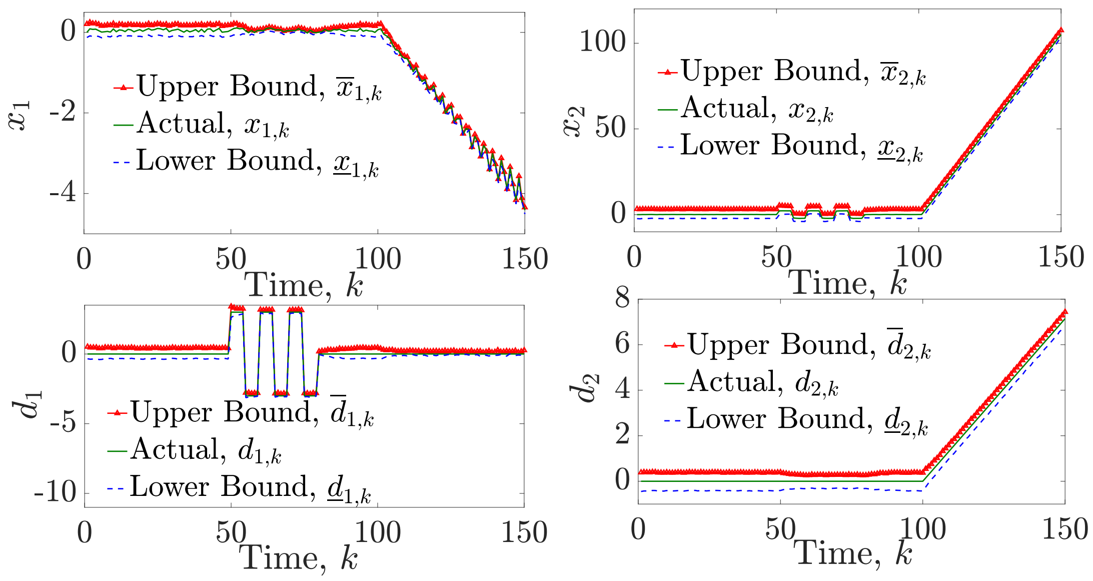


Figure 6.1: Actual states and inputs, $x_{1,k}$, $x_{2,k}$, $d_{1,k}$, $d_{2,k}$, as well as their estimated maximal and minimal values

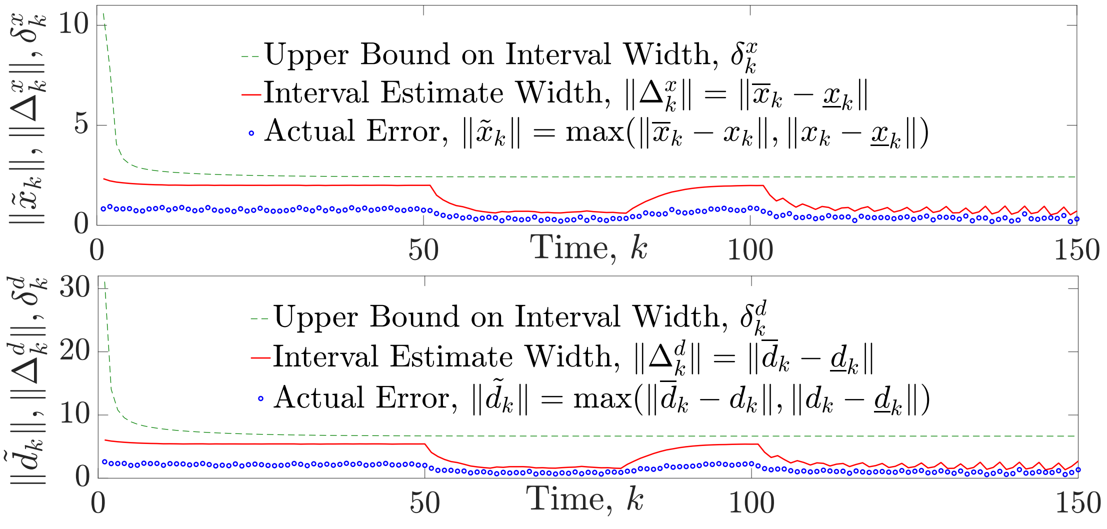


Figure 6.2: Estimation Errors, Estimate Interval Widths and Their Upper Bounds for the Interval-valued Estimates of States, $\|\tilde{x}_{k|k}\|$, $\|\Delta^x_k\|$, δ^x_k, and Unknown Inputs, $\|\tilde{d}_k\|$, $\|\Delta^d_k\|$, δ^d_k

In addition, from [128, (10)–(13)]), we obtain $C_f = \begin{bmatrix} 0.251 & 0 \\ 0.0029 & 0.201 \end{bmatrix}$, $C_g = \begin{bmatrix} 0 & 0.225 \\ -.374 & -.045 \end{bmatrix}$ using (6.2), which implies that $L_{f_d} = 0.852$ and $L_{g_d} = 1.19$ by Lemma 6.1.10. Consequently, $\hat{\mathcal{L}} = 0.643$ is the smallest one that satisfies Condition (i) in Theorem 6.3.8 with $\mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. So, we expect to obtain uniformly bounded estimate errors with convergent upper bounds. This is shown in Figure 7.3, where at each step, the actual error is less than or equal to the interval width, which in turn is less than or equal to the predicted upper bound for the interval width and the upper bounds converge to some steady-state values. Note that, despite our best efforts, we were unable to find interval-valued observers in the literature that simultaneously return both state and unknown input estimates for comparison with our results.

6.5 Conclusion

In this section, a simultaneous input and state interval-valued observer for bounded-error mixed monotone Lipschitz nonlinear systems with unknown inputs was proposed. We derived sufficient conditions for the existence of our observer, proved that the observer recursively outputs the correct state and unknown input framers and proved the tightness of the input interval estimates, given the state intervals and a specific pair of decomposition functions. Further, several conditions for the stability of the observer, i.e., the uniform boundedness of the interval widths were derived. Finally, we demonstrated the effectiveness of the proposed approach with an example. For future work, we seek to find tighter decomposition (bounding) functions and to provide necessary conditions for the existence and stability of the observer.

Chapter 7

INTERVAL OBSERVERS FOR SIMULTANEOUS STATE AND MODEL ESTIMATION OF PARTIALLY KNOWN NONLINEAR SYSTEMS

This chapter [a] addresses the problem of designing interval observers for *partially unknown* nonlinear systems with bounded noise signals that simultaneously estimate the system states and learn a model of the unknown dynamics. Leveraging affine abstraction methods and nonlinear decomposition functions, as well as a data-driven function over-approximation/abstraction approach to over-estimate the unknown dynamic model, our proposed observer recursively computes the maximal and minimal elements of the interval estimates that are proven to frame the true augmented states. Then, using observed output/measurement signals, the observer iteratively shrinks the intervals by eliminating estimates that are not compatible with the measurements. Moreover, given new interval estimates, the observer updates the over-approximation model of the unknown dynamics. Finally, we provide sufficient conditions for uniform boundedness of the sequence of interval estimate widths, i.e., for the stability of the designed observer.

7.1 Problem Formulation

System Assumptions. Consider a partially unknown nonlinear discrete-time system with bounded noise

$$\begin{aligned} x_{k+1} &= f(x_k, d_k, u_k, w_k), \\ y_k &= g(x_k, d_k, u_k, v_k), \end{aligned} \tag{7.1}$$

---

[a] The content of this chapter is documented as a published paper in [113] and an accepted paper in [122].

where $x_k \in \mathcal{X} \subset \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$, $u_k \in \mathcal{U} \subset \mathbb{R}^m$ is a known input vector, $y_k \in \mathbb{R}^l$ is the measurement vector and $d_k \in \mathcal{D} \subset \mathbb{R}^p$ is an unknown (dynamic) input vector whose dynamics is governed by an *unknown* [b] *vector field* $h(\cdot)$:

$$d_{k+1} = h(x_k, d_k, u_k, w_k). \tag{7.2}$$

Moreover, we refer to $z_k \triangleq \begin{bmatrix} x_k^\top & d_k^\top \end{bmatrix}^\top$ as the augmented state. The process noise $w_k \in \mathbb{R}^{n_w}$ and the measurement noise $v_k \in \mathbb{R}^l$ are assumed to be bounded, with $\underline{w} \le w_k \le \overline{w}$ and $\underline{v} \le v_k \le \overline{v}$, where $\underline{w}$, $\overline{w}$ and $\underline{v}$, $\overline{v}$ are the known lower and upper bounds of the process and measurement noise signals, respectively. We also assume that lower and upper bounds, $\underline{z}_0$ and $\overline{z}_0$, for the initial augmented state $z_0 \triangleq \begin{bmatrix} x_0^\top & d_0^\top \end{bmatrix}^\top$ are available, i.e., $\underline{z}_0 \le z_0 \le \overline{z}_0$.

The vector fields $f(\cdot) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{n_w} \to \mathbb{R}^n$ and $g(\cdot) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}^l$ are known, while the vector field $h(\cdot) = \begin{bmatrix} h_1^\top(\cdot) \dots h_p^\top(\cdot) \end{bmatrix}^\top : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{n_w} \to \mathbb{R}^p$ is *unknown*, but each of its arguments $h_j(\cdot) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{n_w} \to \mathbb{R}$, $\forall j \in \{1 \dots p\}$ is known to be Lipschitz continuous. For simplicity and without loss of generality, we assume that the Lipschitz constant L_j^h is known; otherwise, we can estimate the Lipschitz constants with any desired precision using the approach in [59, Equation (12) and Proposition 3]. Moreover, we assume the following:

Assumption 7.1.1. *The vector field* $f(\cdot)$ *is mixed-monotone with decomposition function* $f_d(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{n_w} \times \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^{n_w} \to \mathbb{R}^n$.

Assumption 7.1.2. *The entire space* $\mathbb{X} \triangleq \mathcal{Z} \times \mathcal{U}$ *is bounded, where* $\mathcal{Z} \triangleq \mathcal{X} \times \mathcal{D}$ *and* $\mathcal{U}$ *are the spaces of the augmented states* $z_k \triangleq \begin{bmatrix} x_k^\top & d_k^\top \end{bmatrix}^\top$ *and the known inputs* u_k, $\forall k \in \{0 \dots \infty\}$, *respectively.*

---

[b] Note that if the vector field $h(\cdot)$ is partially known (i.e., consists of the sum of a known component $\hat{h}(\cdot)$ and an unknown component $\tilde{h}(\cdot)$), we can simply consider $d_{k+1} - \hat{h}(\cdot)$ as the output data for the model learning procedure to learn a model of the (completely) unknown function $\tilde{h}(\cdot)$.

Note that Assumption 7.1.1 is satisfied for a broad range of nonlinear functions [129], while Assumption 7.1.2 is reasonable for most practical systems.

The observer design problem can be stated as follows:

Problem 7.1.3. *Given a partially known nonlinear discrete-time system* (7.1) *with bounded noise signals and unknown dynamics* (7.2)*, design a stable observer that simultaneously finds bounded intervals of compatible augmented states and learns an unknown dynamics model for* (7.2)*.*

7.2 State and Model Interval Observers (SMIO)

Recursive Interval Observer In this section, we introduce a three-step recursive interval observer that combines model-based estimation and data-driven model learning approaches. The observer structure is composed of a State Propagation (SP), a Measurement Update (MU) step and a Model Learning (ML) step. In the state propagation step, the interval estimate for the augmented states (consisting of the state and the unknown input) is propagated for one time step through the nonlinear state equation and the estimated model of the unknown dynamics function obtained in previous time step. In the update step, compatible intervals of the augmented states are iteratively updated given new measurements and the nonlinear observation function, and finally, the model learning step estimates the upper and lower *framer* functions (abstractions) for the unknown dynamics function. More formally, the three observer steps have the following form (with $z_k \triangleq [x_k^\top \ d_k^\top]^\top$, $z_k^p \triangleq [x_k^{p\top} \ d_k^{p\top}]^\top$):

$$\begin{aligned}
&\textit{SP:} \quad \mathcal{I}_k^{z^p} = \mathcal{F}^p(\mathcal{I}_{k-1}^z, y_{k-1}, u_{k-1}, \overline{h}_{k-1}(.), \underline{h}_{k-1}(.)),\\
&\textit{MU:} \quad \mathcal{I}_k^z = \mathcal{F}^u(\mathcal{I}_k^{z^p}, y_k, u_k),\\
&\textit{ML:} \quad [\underline{h}_k^\top(.) \ \overline{h}_k^\top(.)]^\top = \mathcal{F}^l(\{\mathcal{I}_{k-t}^z, u_{k-t}\}_{t=0}^k),
\end{aligned}$$

with $\mathcal{F}^p$ and $\mathcal{F}^u$ being to-be-designed interval-valued mappings and $\mathcal{F}^l$ a to-be-constructed function over-approximation procedure (abstraction model), while $\mathcal{I}_k^{z^p}$ and $\mathcal{I}_k^z$ are the intervals of compatible propagated and estimated augmented states, respectively, and $\{\overline{h}_k(\cdot), \underline{h}_k(\cdot)\}$ is a *data-driven abstraction/over-approximation model* for the unknown function $h(\cdot)$, at time step k, i.e.,

$$\forall \zeta_k \in \mathcal{D}_h : \underline{h}_k(\zeta_k) \leq h(\zeta_k) \leq \overline{h}_k(\zeta_k),$$

where $\mathcal{D}_h$ is the domain of $h(\cdot)$ and $\zeta_k \triangleq [z_k^\top \; u_k^\top \; w_k^\top]^\top$.

To leverage the properties of intervals [42] while avoiding the computational complexity of optimal observers [82], we consider the following form of interval estimates in the propagation and update steps:

$$\mathcal{I}_k^{z^p} = \{z \in \mathbb{R}^{n+p} : \underline{z}_k^p \leq z \leq \overline{z}_k^p\},$$
$$\mathcal{I}_k^{z} = \{z \in \mathbb{R}^{n+p} : \underline{z}_k \leq z \leq \overline{z}_k\},$$

where the estimation boils down to finding the maximal and minimal values of $\mathcal{I}_k^{z^p}$ and $\mathcal{I}_k^z$, i.e., $\overline{z}_k^p, \underline{z}_k^p, \overline{z}_k, \underline{z}_k$. Further, at the model learning step, given the sequence of interval estimates up to the current time, we plan to leverage the data-driven function abstraction/over-approximation approach developed in our previous work [59] to update and refine the learned/estimated model of the unknown dynamics function $h(\cdot)$ at the current time step.

Specifically, our interval observer at each time step $k \geq 1$ is given as follows (with the augmented state $z_k \triangleq \begin{bmatrix} x_k^\top & d_k^\top \end{bmatrix}^\top$, $\zeta_k \triangleq \begin{bmatrix} z_k^\top & u_k^\top & w_k^\top \end{bmatrix}^\top$ and known $\underline{x}_0$ and $\overline{x}_0$ such that $\underline{x}_0 \leq x_0 \leq \overline{x}_0$):

State Propagation (SP):

$$\begin{bmatrix} \overline{x}_k^p \\ \underline{x}_k^p \end{bmatrix} = \begin{bmatrix} \min(f_d(\overline{z}_{k-1}, u_{k-1}, \overline{w}, \underline{z}_{k-1}, u_{k-1}, \underline{w}), \overline{x}_k^{a,p}) \\ \max(f_d(\underline{z}_{k-1}, u_{k-1}, \underline{w}, \overline{z}_{k-1}, u_{k-1}, \overline{w}), \underline{x}_k^{a,p}) \end{bmatrix}, \tag{7.3a}$$

$$\begin{bmatrix} \overline{d}_k^p \\ \underline{d}_k^p \end{bmatrix} = \mathbb{A}_k^h \begin{bmatrix} \overline{z}_{k-1}^p \\ \underline{z}_{k-1}^p \end{bmatrix} + \mathbb{B}_k^h u_{k-1} + \mathbb{W}_k^h \begin{bmatrix} \overline{w} \\ \underline{w} \end{bmatrix} + \tilde{e}_k^h, \tag{7.3b}$$

$$\overline{z}_k^p = \begin{bmatrix} \overline{x}_k^{p^\top} & \overline{d}_k^{p^\top} \end{bmatrix}^\top, \underline{z}_k^p = \begin{bmatrix} \underline{x}_k^{p^\top} & \underline{d}_k^{p^\top} \end{bmatrix}^\top, \tag{7.3c}$$

Measurement Update (MU):

$$\begin{bmatrix} \overline{z}_k & \underline{z}_k \end{bmatrix} = \lim_{i \to \infty} \begin{bmatrix} \overline{z}_{i,k}^u & \underline{z}_{i,k}^u \end{bmatrix}, \tag{7.4a}$$

$$\begin{bmatrix} \overline{x}_k & \underline{x}_k \\ \overline{d}_k & \underline{d}_k \end{bmatrix} = \begin{bmatrix} \overline{z}_{k,(1:n)} & \underline{z}_{k,(1:n)} \\ \overline{z}_{k,(n+1:n+p)} & \underline{z}_{k,(n+1:n+p)} \end{bmatrix}, \tag{7.4b}$$

Model Learning (ML):

$$\overline{h}_{k,j}(\zeta_k) = \min_{t \in \{0,\ldots,T-1\}} (\overline{d}_{k-t,j} + L_j^h \|\zeta_k - \tilde{\zeta}_{k-t}\|) + \varepsilon_{k-t}^j, \tag{7.5a}$$

$$\underline{h}_{k,j}(\zeta_k) = \max_{t \in \{0,\ldots,T-1\}} (\underline{d}_{k-t,j} - L_j^h \|\zeta_k - \tilde{\zeta}_{k-t}\|) + \varepsilon_{k-t}^j, \tag{7.5b}$$

where $j \in \{1 \ldots p\}$, $\{\tilde{\zeta}_{k-t} = \frac{1}{2}(\overline{\zeta}_{k-t} + \underline{\zeta}_{k-t})\}_{t=0}^k$ and $\{\overline{d}_{k-t}, \underline{d}_{k-t}\}_{t=0}^k$ are the *augmented* input-output data set. At each time step k, the augmented data set constructed from the estimated framers gathered from the initial to the current time step, is used in the model learning step to recursively derive over-approximations of the unknown function $h(\cdot)$, i.e., $\{\overline{h}_k(.), \underline{h}_k(.)\}$ by applying [59, Theorem 1]. In addition,

$$\begin{bmatrix} \overline{x}_k^{a,p} \\ \underline{x}_k^{a,p} \end{bmatrix} = \mathbb{A}_k^f \begin{bmatrix} \overline{z}_{k-1}^p \\ \underline{z}_{k-1}^p \end{bmatrix} + \mathbb{B}_k^f u_{k-1} + \mathbb{W}_k^f \begin{bmatrix} \overline{w} \\ \underline{w} \end{bmatrix} + \tilde{e}_k^f. \tag{7.6}$$

Moreover, the *sequences of updated framers* $\{\overline{z}_{i,k}^u, \underline{z}_{i,k}^u\}_{i=1}^\infty$ are iteratively computed as follows:

$$\begin{bmatrix} \overline{z}^u_{0,k} & \underline{z}^u_{0,k} \end{bmatrix} = \begin{bmatrix} \overline{z}^p_k & \underline{z}^p_k \end{bmatrix}, \quad \forall i \in \{1 \ldots \infty\} : \tag{7.7}$$

$$\begin{bmatrix} \overline{z}^u_{i,k} \\ \underline{z}^u_{i,k} \end{bmatrix} = \begin{bmatrix} \min(A^{g\dagger+}_{i,k} \overline{\alpha}_{i,k} - A^{g\dagger++}_{i,k} \underline{\alpha}_{i,k} + \omega_{i,k}, \overline{z}^u_{i-1,k}) \\ \max(A^{g\dagger+}_{i,k} \underline{\alpha}_{i,k} - A^{g\dagger++}_{i,k} \overline{\alpha}_{i,k} - \omega_{i,k}, \underline{z}^u_{i-1,k}) \end{bmatrix}, \tag{7.8}$$

where

$$\begin{bmatrix} \overline{t}_{i,k} \\ \underline{t}_{i,k} \end{bmatrix} = \begin{bmatrix} y_k - B^g_{i,k} u_k \\ y_k - B^g_{i,k} u_k \end{bmatrix} + \begin{bmatrix} W^{g++}_{i,k} & -W^{g+}_{i,k} \\ -W^{g+}_{i,k} & W^{g++}_{i,k} \end{bmatrix} \begin{bmatrix} \overline{v} \\ \underline{v} \end{bmatrix} - \begin{bmatrix} \underline{e}^g_{i,k} \\ \overline{e}^g_{i,k} \end{bmatrix}, \tag{7.9}$$

$$\begin{bmatrix} \overline{\alpha}_{i,k} \\ \underline{\alpha}_{i,k} \end{bmatrix} = \begin{bmatrix} \min(\overline{t}_{i,k}, A^{g+}_{i,k} \overline{z}^u_{i-1,k} - A^{g++}_{i,k} \underline{z}^u_{i-1,k}) \\ \max(\underline{t}_{i,k}, A^{g+}_{i,k} \underline{z}^u_{i-1,k} - A^{g++}_{i,k} \overline{z}^u_{i-1,k}) \end{bmatrix}. \tag{7.10}$$

Furthermore, $\omega_{i,k}$, $A^g_{i,k}$, $B^g_{i,k}$, $W^g_{i,k}$, $\overline{e}^g_{i,k}$, $\underline{e}^g_{i,k}$, $\mathbb{B}^q_k$, $\mathbb{J}^q_k$, $\tilde{e}^q_k$, ε^j_{k-t}, $\forall q \in \{f, h\}, \mathbb{J} \in \{\mathbb{A}, \mathbb{W}\}$, $i \in \{1, \ldots, \infty\}$, $j \in \{1, \ldots, p\}$, are to-be-designed observer parameters, matrix gains of appropriate dimensions at time k and iteration i (given in Theorem 7.2.2), while $f_d(., ., ., .)$ is the bounding function (based on (6.2)), with the purpose of achieving desirable observer properties. Algorithm 6 summarizes the SMIO observer.

Note that since the tightness of the upper and lower bounding functions for the observation function g (cf. Propositions 6.1.8 and 6.1.2) depends on the *a priori* interval $\mathcal{B}$, the measurement update step is done iteratively (see proof of Theorem 7.2.6 for more explanation). Hence, if tighter updated intervals are obtained starting from the compatible intervals from the propagation step, we can use them as the new $\mathcal{B}$ to obtain better abstraction/bounding functions for g, which in turn may lead to even tighter updated intervals. Repeating this process results in a sequence of monotonically tighter updated intervals, that is convergent by the monotone convergence theorem, and its limit is chosen as the final interval estimate at time k.

Further, building upon our previous result in [59, Theorem 1], in the model learning step with the history of obtained compatible intervals up to the current

Algorithm 6 State and Model Interval Observer (SMIO)

```
Initialize: maximal($\mathcal{I}^z_0$) = $\overline{z}_0$; minimal($\mathcal{I}^z_0$) = $\underline{z}_0$;
    ▷ Observer Gains Computation
    $\forall q \in \{f,h\}, \mathbb{J} \in \{\mathbb{A},\mathbb{W}\}$, $i \in \{1\dots\infty\}$, $j \in \{1\dots p\}$ compute $\omega_{i,k}$, $A^g_{i,k}$, $B^g_{i,k}$, $W^g_{i,k}$, $\overline{e}^g_{i,k}$,
    $\underline{e}^g_{i,k}$, $\mathbb{B}^q_k$, $\mathbb{J}^q_k$, $\tilde{e}^q_k$, $\varepsilon^j_{k-t}$ via Theorem 7.2.2 and (7.14)–(7.15) ;
for $k = 1$ to $\overline{K}$ do
    ▷ Augmented State Estimation
      Compute $\overline{z}^p_k, \underline{z}^p_k$ via (7.3a)–(7.3c) and $\{\overline{z}^u_{i,k}, \underline{z}^u_{i,k}\}^{\infty}_{i=0}$ via (7.7)–(7.10);
      $(\overline{z}_k, \underline{z}_k) = (\overline{z}^u_{\infty,k}, \underline{z}^u_{\infty,k})$; $\mathcal{I}^z_k=\{z \in \mathbb{R}^n : \underline{z}_k\leq z\leq \overline{z}_k\}$;
      Compute $\delta^z_k$ through Lemma 6.3.10;
    ▷ Model Estimation
      Compute $\overline{h}_k(\cdot), \underline{h}_k(\cdot)$ via (7.5a)–(7.5b);
end for
```

time, $\{[\underline{z}_s, \overline{z}_s]\}^k_{s=0}$ as the noisy input data and the compatible interval of unknown inputs, $[\underline{d}_k, \overline{d}_k]$, as the noisy output data, we recursively construct a sequence of *abstraction/over-approximation models* $\{\overline{h}_k(\cdot), \underline{h}_k(\cdot)\}^{\infty}_{k=1}$ for the unknown input function $h(\cdot)$, that by construction satisfy (7.16), i.e., our model estimation is correct (i.e., is guaranteed to frame/bracket the true function) and becomes more precise with time (cf. Lemma 7.2.3).

7.2.1 Correctness of the Observer

The objective of this section is to design the SMIO observer gains such that the *framer property* [77] holds, i.e., we desire to guarantee that the observer returns correct interval estimates, in the sense that starting from the initial interval $\underline{z}_0 \leq z_0 \leq \underline{z}_0$, the true augmented states of the dynamic system (7.1) are guaranteed to be within the estimated intervals, given by (7.3a)-(7.5b). If the observer is correct, we call

$\{\overline{z}_k, \underline{z}_k\}_{k=0}^{\infty}$ an *augmented state framer sequence* for system (7.1).

Before deriving our main first result on correctness of the observer, we state a modified version of our previous result in [108, Theorem 1], in a unified manner that enables us to derive parallel global and local affine bounding functions for our known $f(\cdot), g(\cdot)$ and unknown $h(\cdot)$ vector fields. For clarity, all proofs will be provided in the appendices.

Proposition 7.2.1 (Parallel Affine Abstractions)**.** *Let the entire space be defined as* $\mathbb{X}$ *and suppose that Assumption 7.1.2 holds. Consider the vector fields* $\overline{q}(.), \underline{q}(.) : \mathbb{X} \subset \mathbb{R}^{n'} \to \mathbb{R}^{m'}$*, where* $\forall \zeta \in \mathbb{X}, \underline{q}(\zeta) \leq \overline{q}(\zeta)$*, along with the following Linear Program (LP):*

$$\min_{\theta^q_{\mathcal{B}}, A^q_{\mathcal{B}}, \overline{e}^q_{\mathcal{B}}, \underline{e}^q_{\mathcal{B}}} \theta^q_{\mathcal{B}} \tag{7.11a}$$

$$\begin{aligned} s.t \quad & A^q_{\mathcal{B}}\zeta_s + \underline{e}^q_{\mathcal{B}} + \sigma^q \leq \underline{q}(\zeta_s) \leq \overline{q}(\zeta_s) \leq A^q_{\mathcal{B}}\zeta_s + \overline{e}^q_{\mathcal{B}} - \sigma^q, \\ & \overline{e}^q_{\mathcal{B}} - \underline{e}^q_{\mathcal{B}} - 2\sigma^q \leq \theta^q \mathbf{1}_{m'}, \\ & \underline{e}^q - \underline{e}^q_{\mathcal{B}} \leq (A^q_{\mathcal{B}} - \mathbb{A}^q)\zeta_s \leq \overline{e}^q - \overline{e}^q_{\mathcal{B}}, \forall \zeta_s \in \mathcal{V}_{\mathcal{B}}, \end{aligned} \tag{7.11b}$$

where $\mathcal{B}$ *is an interval with* $\overline{\zeta}, \underline{\zeta}$ *and* $\mathcal{V}_{\mathcal{B}}$ *being its maximal, minimal and set of vertices, respectively,* $\mathbf{1}_m \in \mathbb{R}^m$ *is a vector of ones,* σ^q *is given in [108, Proposition 1 and (8)] for different classes of vector fields and* $(\mathbb{A}^q, \overline{e}^q, \underline{e}^q)$ *are the global parallel affine abstraction matrices for the pair of functions* $\overline{q}(.), \underline{q}(.)$ *on the entire space* $\mathbb{X}$*, i.e.,*

$$\mathbb{A}^q\zeta + \underline{e}^q \leq \underline{q}(\zeta) \leq \overline{q}(\zeta) \leq \mathbb{A}^q\zeta + \overline{e}^q, \forall \zeta \in \mathbb{X}. \tag{7.12}$$

Using the above proposition, we first solve (7.11a) on the entire space $\mathbb{X}$, i.e., with $\mathcal{B} = \mathbb{X}$ (where the constraint (7.11b) is trivially satisfied and is thus redundant) and obtain a tuple of $(\theta^q, \mathbb{A}^q, \overline{e}^q, \underline{e}^q)$ that satisfies (7.12), i.e., we construct a global affine abstraction model for the pair of functions $\overline{q}(.), \underline{q}(.)$ on the entire space $\mathbb{X}$.

Next, given the (global) tuple $(\mathbb{A}^q, \overline{e}^q, \underline{e}^q)$ computed as described above, we solve (6.1) on $\mathcal{B}$ subject to (7.11b) to obtain a tuple of local parallel affine abstraction matrices for the pair of functions $\{\underline{q}(\cdot), \overline{q}(\cdot)\}$ on the interval $\mathcal{B}$, satisfying the following: $\forall \zeta \in \mathcal{B}$,

$$\mathbb{A}^q\zeta+\underline{e}^q \leq A^q_{\mathcal{B}}\zeta+\underline{e}^q_{\mathcal{B}} \leq \underline{q}(\zeta) \leq \overline{q}(\zeta) \leq A^q_{\mathcal{B}}\zeta+\overline{e}^q_{\mathcal{B}} \leq \mathbb{A}^q\zeta+\overline{e}^q. \tag{7.13}$$

Now, equipped with all the required tools, we state our first main result on the framer property of the SMIO observer.

Theorem 7.2.2 (Correctness of the Observer)**.** *Consider the system* (7.1) *with its augmented state defined as* $z \triangleq \begin{bmatrix} x^\top & d^\top \end{bmatrix}^\top$, *along with the SMIO observer in* (7.3a)–(7.5b). *Suppose that Assumptions 7.1.1 and 7.1.2 hold,* $f_d(\cdot)$ *is a decomposition function of* $f(\cdot)$ *and observer gains and parameters are designed as follows.* $\forall \mathbb{J} \in \{\mathbb{A}, \mathbb{W}\}, q \in \{f, h\}, J \in \{A, W\}, i \in \{1 \dots \infty\}$*:*

$$\mathbb{J}^q_k = \begin{bmatrix} J^{q+}_k & -J^{q++}_k \\ -J^{q++}_k & J^{q+}_k \end{bmatrix}, \mathbb{B}^q_k = \begin{bmatrix} B^{q\top}_k & B^{q\top}_k \end{bmatrix}^\top, \tilde{e}^q_k = \begin{bmatrix} \overline{e}^{q\top}_k & \underline{e}^{q\top}_k \end{bmatrix}^\top, \tag{7.14}$$

$$\omega_{i,k} = \kappa \mathrm{rowsupp}(I - A^{g\dagger}_{i,k} A^g_{i,k}), \varepsilon^j_{k-t} = 2L^h_j \|\overline{\zeta}_{k-t} - \underline{\zeta}_{k-t}\|. \tag{7.15}$$

In addition, $(A^q_k, B^q_k, W^q_k, \overline{e}^q_k, \underline{e}^q_k)$ *for* $q \in \{f, h\}$ *and* $(A^g_{i,k}, B^g_{i,k}, W^g_{i,k}, \overline{e}^g_{i,k}, \underline{e}^g_{i,k})$ *are solutions to the problem* (7.11a) *for the corresponding functions* $\{\underline{g}(\cdot) = \overline{g}(\cdot) = g(\cdot)\}$, $\{\underline{f}(\cdot) = \overline{f}(\cdot) = f(\cdot)\}$ *and* $\{\overline{h}_k(\cdot), \underline{h}_k(\cdot)\}$, *on the intervals*

$$[\begin{bmatrix} \underline{z}^{u\top}_{i-1,k} & u^\top_{k-1} & \underline{v}^\top \end{bmatrix}^\top, \begin{bmatrix} \overline{z}^{u\top}_{i-1,k} & u^\top_{k-1} & \overline{v}^\top \end{bmatrix}^\top]$$

for g *and*

$$[\begin{bmatrix} \underline{z}^\top_{k-1} & u^\top_{k-1} & \underline{w}^\top \end{bmatrix}, \begin{bmatrix} \overline{z}^\top_{k-1} & u^\top_{k-1} & \overline{w}^\top \end{bmatrix}^\top]$$

for $f, \overline{h}_k, \underline{h}_k$, *respectively, at time* k *and iteration* i, *while* κ *is a very large positive real number (infinity).*

Then, the SMIO observer estimates are correct, i.e., the sequences of intervals $\{\overline{z}_k, \underline{z}_k\}_{k=0}^{\infty}$ *are framers of the augmented state sequence of system* (7.1) *that satisfy* $\underline{z}_k \leq z_k \leq \overline{z}_k$ *for all* k*.*

Next, we show that given correct interval estimates, the abstraction model of the unknown dynamics function becomes tighter (i.e., more precise) over time, so our model estimate of the unknown dynamics becomes more accurate over time.

Lemma 7.2.3. *Consider the system* (7.1) *and the SMIO observer in* (7.3a)–(7.5b) *and suppose that all the assumptions in Theorem 7.2.2 hold. Then, the following holds:*

$$\begin{aligned} &\underline{h}_0(\zeta_0) \leq \cdots \leq \underline{h}_k(\zeta_k) \leq \cdots \leq \lim_{k\to\infty} \underline{h}_k(\zeta_k) \leq h(\zeta_k) \\ &h(\zeta_k) \leq \lim_{k\to\infty} \overline{h}_k(\zeta_k) \leq \cdots \leq \overline{h}_k(\zeta_k) \leq \cdots \leq \overline{h}_0(\zeta_0), \end{aligned} \tag{7.16}$$

i.e, the unknown input model estimations/abstractions are correct and become more precise or tighter with time.

7.2.2 Observer Stability

In this section, we investigate the stability of the designed observer in the following sense:

Definition 7.2.4 (Stability)**.** *The observer SMIO* (7.3a)-(7.5b) *is stable, if the sequence of interval widths* $\{\|\Delta^z_{k-1}\| \triangleq \|\overline{z}_{k-1} - \underline{z}_{k-1}\|\}_{k=1}^{\infty}$ *is uniformly bounded, and consequently, the sequence of estimation errors* $\{\|\tilde{z}_{k-1}\| \triangleq \max(\|z_{k-1} - \underline{z}_{k-1}\|, \|\overline{z}_{k-1} - z_{k-1}\|)$ *is also uniformly bounded.*

Next, we derive a property for the decomposition function given in (6.2), which will be helpful in deriving sufficient conditions for the observer stability.

Lemma 7.2.5. *Let* $q(\zeta) : \mathbb{X} \subset \mathbb{R}^n \to \mathbb{R}^m$ *be a mixed-monotone vector-field with a corresponding decomposition function* $q_d(.,.)$ *constructed using* (6.2)*. Suppose that*

Assumption 7.1.2 holds and let $(\mathbb{A}^q, \overline{e}^q, \underline{e}^q)$ *be the parallel affine abstraction matrices for function* $q(\cdot)$ *on its entire domain* $\mathbb{X}$ *(can be computed via Proposition 7.2.1). Consider any ordered pair* $\underline{\zeta} \leq \overline{\zeta} \in \mathbb{X}$*. Then,* $\Delta q_\zeta \leq (|\mathbb{A}^q| + 2C^q)\Delta\zeta + \Delta e^q$*, with* $\Delta q_\zeta \triangleq q_d(\overline{\zeta}, \underline{\zeta}) - q_d(\underline{\zeta}, \overline{\zeta})$*,* $\Delta\zeta \triangleq \overline{\zeta} - \underline{\zeta}$ *and* C^q *given in* (6.2)*.*

We are now ready to state our next main result on the SMIO observer stability in the following theorem.

Theorem 7.2.6 (Observer Stability)**.** *Consider the system* (7.1) *along with the SMIO observer in* (7.3a)–(7.5b)*. Let* $\mathbb{D}_m$ *be the set of all diagonal matrices in* $\mathbb{R}^{m\times m}$ *with their diagonal arguments being* 0 *or* 1*. Suppose that all the assumptions in Theorem 7.2.2 hold and the decomposition function* f_d *is constructed using* (6.2)*. Then, the observer is stable if there exist* $D_1 \in \mathbb{D}_{n+p}, D_2 \in \mathbb{D}_l, D_3 \in \mathbb{D}_n$ *that satisfy* $D_{1,i,i} = 0$ *if* $r(i) = 1$*, i.e., if there exist*

$$(D_1, D_2, D_3) \in \mathbb{D}^* \triangleq \{(D_1, D_2, D_3) \in \mathbb{D}_{n+p} \times \mathbb{D}_l \times \mathbb{D}_n \big| D_{1,ii} r(i) = 0\}$$

such that

$$\mathcal{L}^*(D_1, D_2, D_3) \triangleq \|\mathcal{A}^g(D_1, D_2)\mathcal{A}^{f,h}(D_3)\| \leq 1, \tag{7.17}$$

with

$$\begin{aligned}\mathcal{A}^g(D_1, D_2) &\triangleq (I - D_1) + D_1|A^{g\dagger}|(I - D_2)|A^g|,\\ \mathcal{A}^{f,h}(D_3) &\triangleq \begin{bmatrix}(|A^f| + 2(I - D_3)C^f_z)^\top & |A^h|^\top\end{bmatrix}^\top,\end{aligned}$$

$\{A^q \triangleq \mathbb{A}^q_{(1:n+p)}\}_{q\in\{f,g,h\}}$*,* $\mathbb{A}^q$ *given in Proposition 7.1.2,* $r \triangleq \mathrm{rowsupp}(I - A^{g\dagger}A^g)$*, and* $C^f \triangleq \begin{bmatrix}C^f_z & C^f_u & C^f_w\end{bmatrix}$ *from* (6.2)*.*

Remark 7.2.7. *The sufficient condition in Theorem 7.2.6 has a finitely countable feasible set* ($|\mathbb{D}^*| \leq 2^{2n+p+l}$)*; hence, the condition can be easily checked by enumerating all possible cases and checking the satisfaction of* (7.17)*.*

Finally, we conclude this section by providing upper bounds for the interval widths and compute their steady-state values, if they exist.

Lemma 7.2.8 (Upper Bounds of the Interval Widths and their Convergence)**.** *Consider the system* (7.1) *and the observer* (7.3a)–(7.5b) *and suppose all the assumptions in Theorem 7.2.6 hold. Then, the sequence of* $\{\Delta^z_k \triangleq \overline{z}_k - \underline{z}_k\}_{k=0}^{\infty}$ *is uniformly upper bounded by a convergent sequence as:*

$$\Delta^z_k \leq \overline{\mathcal{A}}^k \Delta^z_0 + \sum_{j=0}^{k-1} \overline{\mathcal{A}}^j \overline{\Delta} \xrightarrow{k\to\infty} e^{\overline{\mathcal{A}}}\overline{\Delta}, \tag{7.18}$$

where

$$\begin{aligned}
&\overline{\mathcal{A}} = \mathcal{A}(D_1^*, D_2^*, D_3^*) \triangleq \mathcal{A}^g(D_1^*, D_2^*)\mathcal{A}^{f,h}(D_3^*),\\
&\overline{\Delta} = \Delta^g(D_1^*, D_2^*) + \mathcal{A}^g(D_1^*, D_2^*)\Delta^{f,h}(D_3^*),\\
&\mathcal{A}^g(D_1, D_2) \triangleq D_1|A^{g\dagger}|D_2|A^g| + (I - D_1),\\
&\mathcal{A}^{f,h}(D_3) \triangleq \begin{bmatrix}(|A^f| + 2(I - D_3)C^f_z)^\top & |A^h|^\top\end{bmatrix}^\top,\\
&\Delta^g(D_1, D_2) \triangleq D_1|A^{g\dagger}|D_2(|W^g|\Delta v + \Delta e^g),\ \Delta^{f,h}(D_3) \triangleq\\
&\begin{bmatrix}((|W^f|+2(I-D_3)C^f_w)\Delta w+\Delta^f_e)^\top & (|W^h|\Delta w+\Delta^h_e)^\top\end{bmatrix}^\top,
\end{aligned}$$

and (D_1^*, D_2^*, D_3^*) *is a solution of the following problem:*

$$\begin{aligned}
&\min_{D_1,D_2,D_3} \|e^{\mathcal{A}(D_1,D_2,D_3)}(\Delta^g(D_1, D_2)+\mathcal{A}^g(D_1, D_2)\Delta^{f,h}(D_3))\|\\
&s.t. (D_1, D_2, D_3) \in \{(D_1, D_2, D_3) \in \mathbb{D}^* | \mathcal{L}^*(D_1, D_2, D_3) < 1\}.
\end{aligned}$$

Consequently, the sequence of interval widths $\{\|\Delta^z_k\|\}_{k=1}^{\infty}$ *is uniformly upper bounded by a convergent sequence as:*

$$\|\Delta^z_k\| \leq \delta^z_k \triangleq \|\overline{\mathcal{A}}^k \Delta^z_0 + \sum_{j=0}^{k-1} \overline{\mathcal{A}}^j \overline{\Delta}\| \xrightarrow{k\to\infty} \|e^{\overline{\mathcal{A}}}\overline{\Delta}\|. \tag{7.19}$$

7.3 Simulation Results

We consider a slightly modified version of the continuous-time predator-prey system in [94]:

$$\begin{aligned} \dot{x}_1 &= -x_1x_2 - x_2 + u + d + w_1, \\ \dot{x}_2 &= x_1x_2 + x_1 + w_2, \\ \dot{d} &= 0.1(\cos(x_1) - \sin(x_2)) + w_d, \end{aligned}$$

where the (unknown input) dynamics $\dot{d}$ is an unknown function, and the output equations are given by:

$$y_1 = x_1 + v_1, y_2 = x_2 + v_2, y_3 = \sin(d) + v_3,$$

We use the forward Euler method to discretize the system and the system can be described in the form (7.1)–(7.2) with the following parameters: $n = l = p = 2$, $m = 1$, $f(.) = \begin{bmatrix} f_1(.) & f_2(.) \end{bmatrix}^\top$, $g(.) = \begin{bmatrix} g_1(.) & g_2(.) & g_3(.) \end{bmatrix}^\top$, $u_k = 0$, $w_k = [w_{1,k}\ w_{2,k}\ w_{d,k}]^\top$, $v_k = [v_{1,k}\ v_{2,k}\ v_{3,k}]^\top$, $\overline{v} = -\underline{v} = \overline{w} = -\underline{w} = \begin{bmatrix} 0.1 & 0.1 & 0.1 \end{bmatrix}^\top$, $\overline{x}_0 = \begin{bmatrix} 0 & 0.6 \end{bmatrix}^\top$, $\underline{x}_0 = \begin{bmatrix} -0.35 & -0.1 \end{bmatrix}^\top$, where

$$\begin{aligned} f_1(\cdot) &= x_{1,k} + \delta_t(-x_{1,k}x_{2,k} - x_{2,k} + u_k + d_k + w_{1,k}), \\ f_2(\cdot) &= x_{2,k} + \delta_t(x_{1,k}x_{2,k} + x_{1,k} + w_{2,k}), \\ h(\cdot) &= d_k + \delta_t(0.1(\cos(x_{1,k}) - \sin(x_{2,k})) + w_{d,k}) \\ g_1(\cdot) &= x_{1,k} + v_{1,k}, \\ g_2(\cdot) &= x_{2,k} + v_{2,k}, \\ g_3(\cdot) &= \sin(d_k) + v_{3,k}, \end{aligned}$$

with sampling time $\delta_t = 0.01s$. Moreover, using Proposition 7.2.1 with abstraction slopes set to zero, we can obtain finite-valued upper and lower bounds (horizontal abstractions) for the partial derivatives of $f(\cdot)$ as:

$$\begin{bmatrix} a_{11}^f & a_{12}^f & a_{12}^f \\ a_{21}^f & a_{22}^f & a_{23}^f \end{bmatrix} = \begin{bmatrix} 0.994 & -0.01 & 1-\epsilon \\ 0.009 & 0.9965 & -\epsilon \end{bmatrix},$$

$$\begin{bmatrix} b_{11}^f & b_{12}^f & b_{13}^f \\ b_{21}^f & b_{22}^f & b_{23}^f \end{bmatrix} = \begin{bmatrix} 1.006 & -0.0065 & 1+\epsilon \\ 0.016 & 1 & \epsilon \end{bmatrix},$$

where ϵ is a very small positive value, ensuring that the partial derivatives are in open intervals (cf. [128, Theorem 1]). Therefore, Assumption 7.1.1 holds by [128, Theorem 1]). Hence, we expect that the true states and unknown inputs are within the interval estimates by Theorem 7.2.2, i.e., the interval estimates are correct. This can be observed from Figures 7.1 and 7.2, where the true states and unknown inputs as well as interval estimates are depicted.

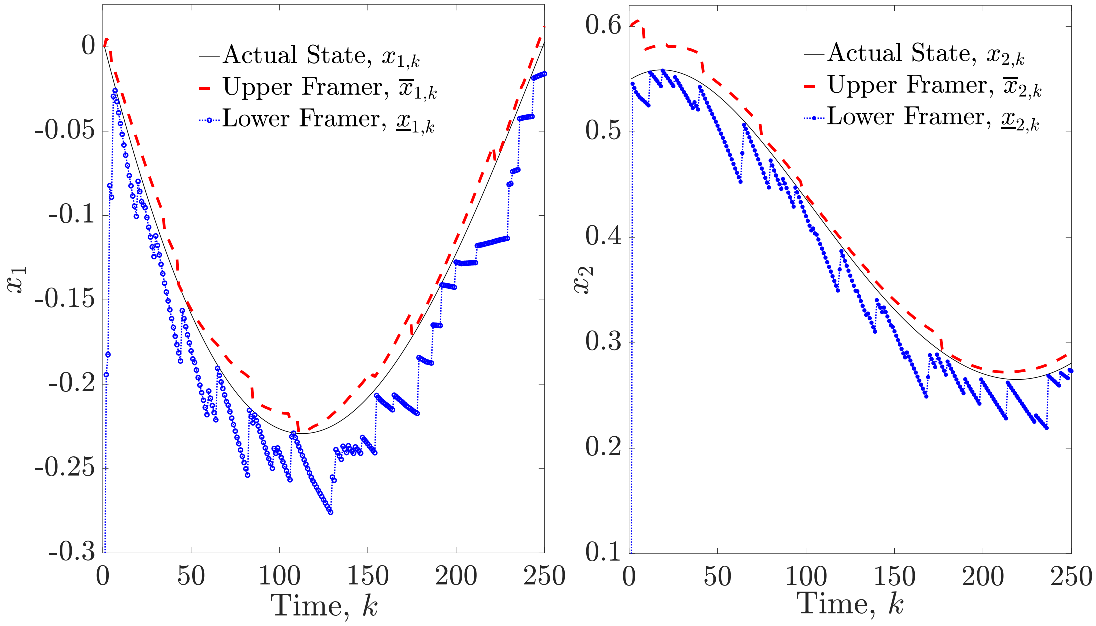


Figure 7.1: Actual States, as Well as Their Estimated Maximal and Minimal Values

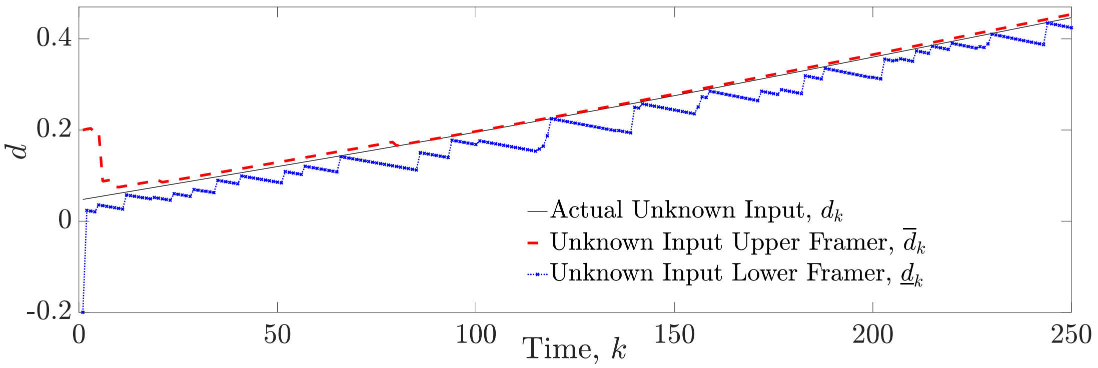


Figure 7.2: Actual Unknown Input, d_k, as Well as Its Estimated (Learned) Maximal and Minimal Values

Furthermore, solving the optimization problem in Proposition 7.2.1 for the global abstraction matrices, we obtained

$$A^f = \begin{bmatrix} 0.6975 & -0.0083 & 0.01 \\ 0.0125 & 0.9982 & 0 \end{bmatrix},$$

$$A^g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.995 \end{bmatrix}, \quad A^h = \begin{bmatrix} 0 & -0.0015 & .6 \end{bmatrix},$$

and from [128, (10)–(13)]), we obtained $C_f = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ when using (6.2). Consequently, (7.17) is satisfied and so, the sufficient condition in Theorem 7.2.6 holds. Moreover, as can be seen in Figure 7.3, we obtain uniformly bounded and convergent interval estimate errors when applying our observer design procedure, where at each time step, the actual error sequence is upper bounded by the interval widths, which converge to steady-state values. Further, Figure 7.4 shows the framer intervals of the learned/estimated unknown dynamics model (depicted by the "kinky" red and blue meshes) that frame the actual unknown dynamics function $h(\cdot)$, as well as the global abstraction that is computed via Proposition 7.2.1 at the initial step.

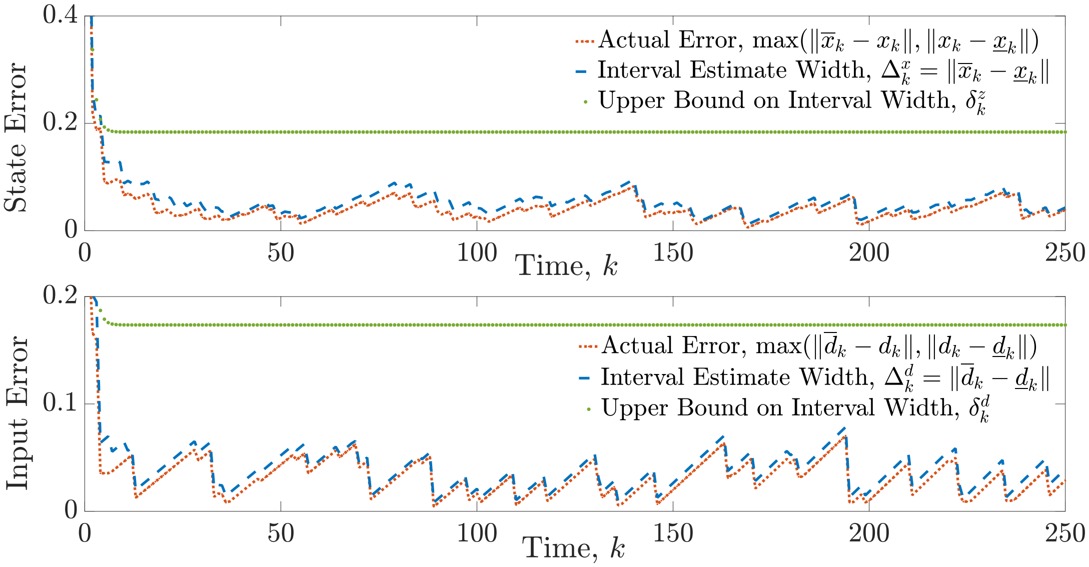


Figure 7.3: Actual Estimation Errors, Interval Estimate Widths and Their Upper Bounds for the Interval Estimates of States and Unknown Inputs

Note that as discussed in the proof of Theorem 7.2.6, since we need to check an *a priori* condition (i.e., offline or before starting to implement the observer) for observer stability, we use global abstraction slopes for stability analysis. However, for the implementation, we iteratively update the framers and consequently, obtain the updated local abstractions, which, in turn, lead to updated local intervals that by construction are tighter than the global ones, as shown in the proof of Theorem 7.2.6. Hence, for a given system, it might be the case that the (relatively conservative) global abstraction-based sufficient conditions for the observer stability given in Theorem 7.2.6 do not hold, while the implemented local-abstraction-based intervals are still uniformly bounded. This is the main benefit of using iterative local affine abstractions, but at the cost of more extensive computational effort.

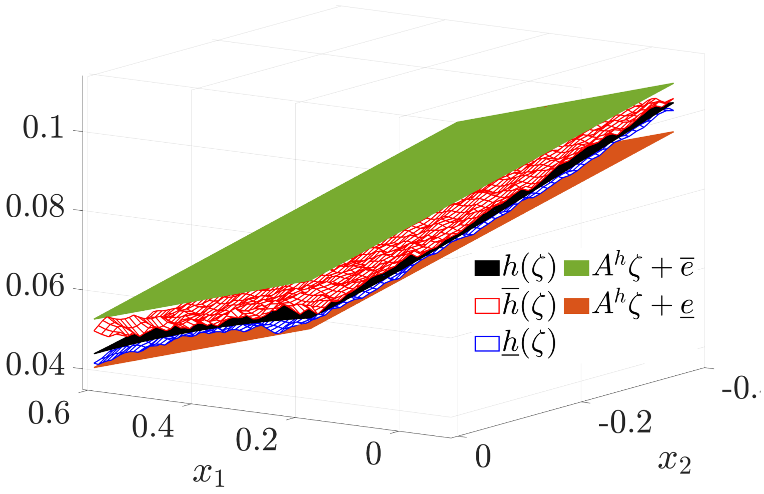


Figure 7.4: Actual Unknown Dynamics Function $h(\zeta)$, Its Upper and Lower Framer Intervals at Time Step $k = 250$, as Well as Its Global Abstraction

7.4 Conclusion

This paper proposed an interval observer for partially unknown nonlinear systems with bounded noise that simultaneously estimates the augmented states and learns the unknown dynamics. By leveraging a combination of nonlinear bounding/decomposition functions, affine abstractions and a data-driven function abstraction method (to over-estimate the unknown dynamics model from noisy input-output data), we introduced a recursive interval observer design whose interval estimates are correct in the sense that the maximal and minimal elements of the interval estimates are guaranteed to frame/bracket the true augmented states. Moreover, using observed output/measurement signals at run time, the observer also iteratively shrinks the intervals by eliminating estimates that are not compatible with the measurements. Further, tractable sufficient conditions for uniform boundedness of the sequence of interval estimate widths, i.e., for stability of the designed observer were provided.

Chapter 8

TIGHT REMAINDER-FORM DECOMPOSITION FUNCTIONS (WITH APPLICATIONS TO CONSTRAINED REACHABILITY AND INTERVAL OBSERVER DESIGN)

In this chapter [a] , we propose a tractable family of remainder-from decomposition functions, that their existence is proven to be sufficient conditions for mixed-monotonicity of a broad-range of not necessarily smooth, constrained and unconstrained, continuous and discrete-time bounded-error dynamical systems. We provide achievable lower and upper bounds for the error of over-approximating the true range of a mapping, using the proposed remainder-form decomposition functions and specify the best/tightest of them. Moreover, we develop a set-inversion algorithm that along with the proposed decomposition functions have several applications, e.g., approximation of the reachable sets for bounded-error, constrained, continuous and/or discrete-time systems, as well as guaranteed state estimation.

8.1 Preliminary Material

Definition 8.1.1 (Inclusion Functions)**.** *[57, Chapter 2.4] Consider a function* $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$*. The interval function* $T^f : \mathbb{IR}^{n_z} \to \mathbb{IR}^{n_x}$ *is an inclusion function for* $f(\cdot)$*, if*

$$\forall \mathcal{Z} \in \mathbb{IR}^{n_z}, f(\mathcal{Z}) \subset T^f(\mathcal{Z}),$$

where $f(\mathcal{Z})$ *is the true range of* $f(\cdot)$ *applying on* $\mathcal{Z}$*.*

---

[a] The content of this chapter is documented as a submitted and under review paper in [119].

Proposition 8.1.2 (Natural (T_N) Inclusion Functions)**.** *[57, Theorem 2.2] Consider* $\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$ *and* $f \triangleq [f_1 \ldots f_{n_x}]^\top : \mathcal{Z} \to \mathbb{R}^{n_x}$*, where each* $f_j,\ j = 1, \ldots, n_x$*, expressed as a finite composition of the operators* $+, -, \times, /$ *and elementary functions (sine, cosine, exponential, square root, ...). A natural inclusion function* $T_N^f : \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$ *for* $f(\cdot)$ *is obtained by replacing each real variable* $z_i, i = 1, \ldots, n_z$ *by its corresponding "interval variable"* $[z_i] \triangleq \mathcal{Z}_i = [\underline{z}_i, \overline{z}_i]$*, and each operator or function by its interval counterpart.*

Proposition 8.1.3 (Centered (T_C) and Mixed Centered (T_M) Inclusion Functions)**.** *[57, Sections 2.4.3–2.4.4] Let* $f \triangleq [f_1 \ldots f_{n_x}]^\top : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *be differentiable over the box* $\mathcal{Z} \triangleq [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$*. Then the interval function*

$$T_C^f(\mathcal{Z}) \triangleq f(m) + [J_f^\top](\mathcal{Z})(\mathcal{Z} - m)$$

is an inclusion function for $f(\cdot)$ *and is called the centered inclusion function, where* $m \triangleq \frac{\underline{z}+\overline{z}}{2}$*,* J_f *is the Jacobian matrix of* $f(\cdot)$ *and* $[J_f]$ *is its interval counterpart (natural inclusion). Moreover,*

$$T_M^f(\mathcal{Z}) \triangleq [T_i^f(\mathcal{Z}) \ldots T_{n_x}^f(\mathcal{Z})]^\top, \text{ where } \forall i \in \{1, \ldots, n_x\} :$$
$$T_i^f(\mathcal{Z}) \triangleq f_i(m) + \sum_{j=1}^{n_z} [J_f]_{i,j}(\mathcal{Z}_1, \ldots, \mathcal{Z}_j, m_{j+1}, \ldots, m_{n_z})(z_j - m_j)$$

is also an inclusion function for $f(\cdot)$ *and is called the mixed-centered inclusion function.*

Next, inspired by the work in [35, Section 3], we introduce the notion of remainder-form (additive) inclusion functions.

Definition 8.1.4 (Remainder-Form (Additive) Inclusion Functions)**.** *Consider a function* $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$*. The interval function* $T_R^f : \mathbb{IR}^{n_z} \to \mathbb{IR}^{n_x}$ *is an additive (remainder-form) inclusion function for* $f(\cdot)$*, if there exist two constituent mappings* $g, h : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$*, such that for any* $\mathcal{Z} \in \mathbb{IR}^{n_z}$*:*

$$f(\mathcal{Z}) \subseteq T_R^f(\mathcal{Z}) \triangleq g(\mathcal{Z}) + h(\mathcal{Z}).$$

Definition 8.1.5 (Mixed-Monotonicity, Decomposition Functions). *[2, Definition 1],[128, Definition 4] Consider the dynamical system*

$$x^+ = f(z), \tag{8.1}$$

where $f : \mathcal{Z} \in \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$, $z \triangleq [x^\top \ u^\top \ w^\top]^\top \in \mathcal{Z} \triangleq \mathcal{X} \times \mathcal{U} \times \mathcal{W}$ *with state* $x \in \mathcal{X} \triangleq [\underline{x}, \overline{x}] \subset \mathbb{R}^{n_x}$*, known input* $u \in \mathcal{U} \subset \mathbb{R}^{n_u}$ *and disturbance input* $w \in \mathcal{W} \triangleq [\underline{w}, \overline{w}] \subset \mathbb{R}^{n_w}$*. Suppose* (8.1) *is a discrete-time system and there exists a mapping* $f_d : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{n_x}$ *such that:*

1. $f(\cdot)$ *is embedded on the diagonal of* $f_d(\cdot, \cdot)$*, i.e.,* $f_d(z, z) = f(z)$*.*
2. f_d *is monotone increasing in its first and monotone decreasing in its second argument, i.e.,* $\hat{z} \geq z \implies f_d(\hat{z}, z') \geq f_d(z, z')$ & $f_d(z', \hat{z}) \leq f_d(z', z)$*.*

Then, the mapping $f(\cdot)$ *and consequently system* (8.1) *is mixed-monotone with respect to* f_d*, which is a decomposition function for* f*.*

Moreover, if (8.1) *is a continuous-time system, then everything holds with the slight modification that in 2),* $f_d(\cdot, \cdot)$ *needs to be monotone increasing on its first and monotone decreasing on its second argument, only for "off-diagonal" arguments, i.e., only if* $j \neq i$, $\forall i \in \{1, \ldots, n_x\}, \forall j \in \{1, \ldots, n_z\}$.

Corollary 8.1.6 (Decomposition-Based Inclusion Functions). *As a direct conclusion of Definition 8.1.5, given* $\mathcal{Z} \triangleq [\underline{z}, \overline{z}]$ *and any decomposition function* f_d *for* f, $f(\mathcal{Z}) \subset T^{f_d}(\mathcal{Z}) \triangleq [f_d(\underline{z}, \overline{z}), f_d(\overline{z}, \underline{z})]$*. Hence by Definition 8.1.1, the interval function* $T^{f_d}(\mathcal{Z})$ *is a (decomposition-based) inclusion function for* f *over* $\mathcal{Z}$ *and is called the inclusion function* "induced by" f_d.

Next, we extend the concept of decomposition functions into "one-sided decomposition functions", which will be used later in our results.

Definition 8.1.7 (One-Sided Decomposition Functions)**.** *Consider* $f : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$ *and suppose there exist two mappings* $\overline{f}_d, \underline{f}_d, : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^{n_x}$ *such that for any* $\underline{z}^*, z, \overline{z}^* \in \mathcal{Z}$*, the following statement holds:*

$$\underline{z}^* \leq z \leq \overline{z}^* \implies \underline{f}_d(\underline{z}^*, \overline{z}^*) \leq f(z) \leq \overline{f}_d(\overline{z}^*, \underline{z}^*).$$

Then, $\overline{f}_d$ *and* $\underline{f}_d$ *are called upper and lower decomposition functions for* f *over* $\mathcal{Z}$*, respectively.*

Corollary 8.1.8. *Similar to Corollary 8.1.6, As a direct conclusion of Definition 8.1.7,* $f(\mathcal{Z}) \subset T^{\overline{f}_d}_{\underline{f}_d}(\mathcal{Z}) \triangleq [\underline{f}_d(\underline{z}, \overline{z}), \overline{f}_d(\overline{z}, \underline{z})]$ *and hence* $T^{\overline{f}_d}_{\underline{f}_d}$ *is an inclusion function for* f *over* $\mathcal{Z}$*.*

Corollary 8.1.9. *By slightly generalizing the notion of "embedding system with respect to* f_d*" in [2, (7)], to the "embedding system with respect to* $\overline{f}_d, \underline{f}_d$*, we conclude through the lines of proof of [2, Proposition 2], that if* (8.1) *is a continuous-time system, then at time* t*,* $\underline{x}_t \leq x_t \leq \overline{x}_t$*, where* $\overline{x}_t$ *and* $\underline{x}_t$ *are the solutions to the following "continuous-time generalized embedding system":*

$$\begin{bmatrix} \dot{\overline{x}} \\ \dot{\underline{x}} \end{bmatrix} = \begin{bmatrix} \overline{f}_d(\overline{x}, \underline{x}, u, \overline{w}, \underline{w}) \\ \underline{f}_d(\underline{x}, \overline{x}, u, \underline{w}, \overline{w}) \end{bmatrix},$$

with the initial values $\underline{x}(0) = \underline{x}_0$ *and* $\overline{x}(0) = \overline{x}_0$*. Moreover, if* (8.1) *is a continuous-time system, then the bounds for* x_k *at time step* k*, i.e.,* $\overline{x_k}$ *and* $\underline{x}_k$ *such that* $\underline{x}_k \leq x_k \leq \overline{x}_k$*, can be found by iteratively solving the following "discrete-time generalized embedding system":*

$$\begin{bmatrix} \overline{x}_{k+1} \\ \underline{x}_{k+1} \end{bmatrix} = \begin{bmatrix} \overline{f}_d(\overline{x}_k, \underline{x}_k, u, \overline{w}, \underline{w}) \\ \underline{f}_d(\underline{x}_k, \overline{x}_k, u, \underline{w}, \overline{w}) \end{bmatrix}.$$

Corollary 8.1.10. *Suppose* $\overline{f}^1_d(\cdot, \cdot)$ *and* $\overline{f}^2_d(\cdot, \cdot)$ *are two upper decomposition functions for* $f(\cdot)$*. Then,* $\min\{\overline{f}^1_d, \overline{f}^2_d\}(\cdot, \cdot)$ *is also an upper decomposition functions for* $f(\cdot)$*.*

Similarly, Suppose $\underline{f}_d^1(\cdot,\cdot)$ *and* $\underline{f}_d^2(\cdot,\cdot)$ *are two lower decomposition functions for* $f(\cdot)$*. Then,* $\max\{\underline{f}_d^1, \underline{f}_d^2\}(\cdot,\cdot)$ *is also a lower decomposition functions for* $f(\cdot)$

Proof. The results are implied by the fact that if two mappings with the same domains and image spaces, are both monotone increasing or decreasing in their i'th arguments, then their "point-wise" minimum and maximum are also monotone increasing or decreasing on their i'th arguments. ☐

Definition 8.1.11 (Clarke Generalized Gradient)**.** *[40, Definitions 3.1–3.2] Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be locally Lipschitz. Then, the Clarke generalized gradient or the Clarke sub-differential of* $f(\cdot)$ *at* $x \in \mathbb{R}^n$ *is denoted as* $\partial^o f(x)$ *and is given as*

$$\partial^o f(x) \triangleq \{\xi \in \mathbb{R}^n | \nu^f(x,v) \geq \xi^\top v, \forall v \in \mathbb{R}^n\},$$

where

$$\nu^f(x,v) \triangleq \limsup_{y \to x, \lambda \to 0} \frac{f(y+\lambda v) - f(y)}{\lambda}$$

is called the Clarke generalized directional derivative of $f(\cdot)$ *at* x *in the direction* v*. The set* $\partial^o f(x)$ [b] *is nonempty, convex and compact for each* $x \in \mathbb{R}^n$*. Moreover, when* $f(\cdot)$ *is differentiable, then* $\nabla f(x) \in \partial^o f(x)$*. If* $f(\cdot)$ *is continuously differentiable or strictly differentiable, then* $\partial^o f(x) = \{\nabla f(x)\}$*. Furthermore, it is important to observe that* $\nu^f(x,v)$ *is the "support function" of the set* $\partial^o f(x)$*, i.e.,* $\nu^f(x,v) = \sup_{\xi \in \partial^o f(x)} \xi^\top v$*.*

Next, we slightly generalize the definition of Jacobian sign-stability, compared to the one used in [128], using Clarke sub-differentials instead of partial derivatives.

---

[b] relying on the fact that a locally Lipschitz function $f(\cdot)$ is differentiable "almost everywhere", i.e. the set of points where $f(\cdot)$ is not differentiable is a set of zero measure, the Clarke sub-differential can be given in the following equivalent way:

$$\partial^o f(\overline{x}) = \mathrm{co}\{\xi \in \mathbb{R}^n | \xi = \lim_{k \to \infty} \nabla f(x_k), x_k \to \overline{x}, x_k \in D\},$$

where D is the set of points over which $f(\cdot)$ is differentiable.

Definition 8.1.12 (Jacobian Sign-Stability). *Mapping* $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *is called Jacobian sign-stable (J.S.S.) over* $\mathcal{Z}$*, if* $\forall i \in \{1, \ldots, n_z\}, \forall j \in \{1, \ldots, n_x\}$,

$$\nu_i \leq 0, \forall \nu \in \partial^o f_j(z), \forall z \in \mathcal{Z} \text{ (positive J.S.S.) or} \tag{8.2}$$

$$\nu_i \geq 0, \forall \nu \in \partial^o f_j(z), \forall z \in \mathcal{Z} \text{ (negative J.S.S.)}, \tag{8.3}$$

where $\partial^o f_j(z)$ *is the Clarke generalized gradient (sub-differential) of* $f(\cdot)$ *at* z*, defined in Definition 8.1.11.*

Proposition 8.1.13 (T_L Inclusion Functions). *[128, Theorem 2] Assume* $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *is differentiable and* $\frac{\partial f_j}{\partial z_i}(z) \in (a_{ji}, b_{ji}) \forall z \in \mathcal{Z} \subseteq \mathbb{R}^{n_z}$*. Then,* f *is mixed-monotone on* $\mathcal{Z}$ *with respect to decomposition function* $f_d^L = [f_{d,1}^L \ldots f_{d,n_x}^L]$*, which is computed as follows.* $\forall j = 1, \ldots, n_x$*:*

$$f_{d,j}^L(z, \hat{z}) = f_j(\zeta) + (\alpha_j - \beta_j)^\top (z - \hat{z}) \tag{8.4}$$

where $\zeta = [\zeta_1, \ldots, \zeta_{n_z}]^\top$, $\alpha_j = [\alpha_{j1}, \ldots, \alpha_{jn_z}]^\top$, $\beta_j = [\beta_{j1}, \ldots, \beta_{jn_z}]^\top$, $\zeta_i = \begin{cases} z_i \ \textit{case } 1,2 \\ \hat{z}_i \ \textit{case } 3,4 \end{cases}$,

$\alpha_{ji} = \begin{cases} 0 \ \textit{case } 1,3,4 \\ |a_{ji}| + \epsilon \ \textit{case } 2 \end{cases}$, $\beta_{ji} = \begin{cases} 0 \ \textit{case } 1,2,4 \\ -|b_{ji}| - \epsilon \ \textit{case } 3 \end{cases}$, *case* $1 : a_{ji} \geq 0$*, case* $2 : a_{ji} \leq 0, b_{ji} \geq 0, |a_{ji}| \leq |b_{ji}|$*, case* $3 : a_{ji} \leq 0, b_{ji} \geq 0, |a_{ji}| \geq |b_{ji}|$ *and case* $4 : b_{ji} \leq 0$*. Furthermore,* T_L *is the inclusion function induced by* f_d^L *(cf. Corollary 8.1.6).*

Definition 8.1.14 (Hausdorff Distance). *[35] The Hausdorff Distance function* $q(\cdot, \cdot) : \mathbb{IR} \times \mathbb{IR} \to \mathbb{R}_+$*, between two real intervals* $\mathcal{X}_1 = [\underline{x}_1, \overline{x}_1]$ *and* $\mathcal{X}_2 = [\underline{x}_2, \overline{x}_2]$*, both in* $\mathbb{IR}$*, is defined as follows:*

$$q(\mathcal{X}_1, \mathcal{X}_2) \triangleq \max\{|\underline{x}_1 - \underline{x}_2|_\infty, |\overline{x}_1 - \overline{x}_2|_\infty\}. \tag{8.5}$$

Definition 8.1.15 (Tightness of Decompositions)**.** *[2, Definition 2] A decomposition function* f_d^1 *for system 8.1 is tighter than decomposition function* f_d^2*, if* $\forall z \leq \hat{z}$*:*

$$f_d^2(z, \hat{z}) \leq f_d^1(z, \hat{z}) \; \& \; f_d^1(\hat{z}, z) \leq f_d^1(\hat{z}, z). \tag{8.6}$$

Then, f_d^* *is tight, i.e., is the tightest possible decomposition function for* f*, if* (8.6) *holds with* f_d^* *and any other decomposition function* f_d *for 8.1.*

Proposition 8.1.16 (Tight Decomposition Functions for Mixed-Monotone Systems)**.** *[129, Theorem 2],[2, Theorem 1] Any system of the form of* (8.1)*-discrete-time or continuous-time- is mixed-monotone with respect to a tight decomposition function, which can be described as follows. If* (8.1) *is a continuous-time system, then* $\forall i = 1, \ldots, n_x$*:*

$$f_{d,i}(z, \hat{z}) = \begin{cases} \min\limits_{\zeta \in [z, \hat{z}], \zeta_i = x_i} f_i(\zeta) & \text{if } z \leq \hat{z}, \\ \max\limits_{\zeta \in [\hat{z}, z], \zeta_i = x_i} f_i(\zeta) & \text{if } \hat{z} \leq z. \end{cases} \tag{8.7}$$

Moreover, if (8.1) *is a discrete-time system, then everything holds with relaxing* $\zeta_i = x_i$ *in the constraint sets in* (8.7)*.*

We call the inclusion function induced by the tight (optimal) decomposition functions in (8.7), the T_O.

Corollary 8.1.17 (Tight Decompositions for J.S.S. Vector Fields)**.** *As a corollary of Proposition 8.1.16, if* f *is J.S.S., then the optimization programs in* (8.7) *can be easily and exactly solved by enumerating* $f_i(\cdot)$ *on the vertices of the interval constraint (fixing* ζ *in dimension* i *at* x_i *in continuous-time case) and choosing the minimum and maximum of the obtained values.*

8.2 Problem Statement

Consider the following constrained nonlinear dynamical system:

$$x^+ = f(x, w), \ s.t. \ \mu(x, y, v) \in \mathcal{G}, \tag{8.8}$$

where $x^+ = \dot{x}$ if (8.8) is a continuous-time and $x^+ = x_{k+1}$ if (8.8) is a discrete-time system. Moreover, $x \in \mathcal{X} \triangleq [\underline{x}, \overline{x}] \subset \mathbb{R}^{n_x}$ is the state and $w \in \mathcal{W} \triangleq [\underline{w}, \overline{w}] \subset \mathbb{R}^{n_w}$ is the augmentation of all the exogenous inputs, e.g., known input, bounded disturbance/noise and internal uncertainties such as uncertain parameters, with known bounds $\underline{x}, \overline{x}, \underline{w}, \overline{w}$. Furthermore, $y \in \mathbb{R}^{n_y}$, $v \in \mathcal{V} \triangleq [\underline{v}, \overline{v}] \subset \mathbb{R}^{n_v}$ and $\mathcal{G} \triangleq [\underline{g}, \overline{g}] \subset R^{n_g}$ are the observation (measurement) signal, the measurement noise signal and the enclosing interval of the observation/constraint set, respectively, with $\underline{v}, \overline{v}, \underline{g}, \overline{g}$ being known a prior. Further, the nonlinear vector field $f \triangleq [f_1, \ldots, f_{n_x}]^\top : \mathcal{Z} \triangleq \mathcal{X} \times \mathcal{W} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ and the observation/constraint mapping [c] $\mu \triangleq [\mu_1, \ldots, \mu_{n_\mu}]^\top : \mathcal{T} \triangleq \mathcal{X} \times \{y\} \times \mathcal{V} \subset \mathbb{R}^{n_t} \to \mathbb{R}^{n_\mu}$ are locally Lipschitz continuous, with their Clarke sub-differentials (cf. Definition 8.1.11) being *bounded from one side*, i.e., with known $\underline{a}^f \triangleq [\underline{a}^{f_1}, \ldots, \underline{a}^{f_{n_x}}], \overline{a}^f \triangleq [\overline{a}^{f_1}, \ldots, \overline{a}^{f_{n_x}}] \in \mathbb{R}^{n_z \times n_x}$ and $\underline{a}^\mu \triangleq [\underline{a}^{\mu_1}, \ldots, \underline{a}^{\mu_{n_\mu}}], \overline{a}^\mu \triangleq [\overline{a}^{\mu_1}, \ldots, \overline{a}^{\mu_{n_\mu}}] \in \mathbb{R}^{n_t \times n_\mu}$ such that

$$\begin{aligned} &\underline{a}^{f_j} \leq \nu \text{ or } \nu \leq \overline{a}^{f_j}, \forall \nu \in \partial^o f_j(z), \forall z \in \mathcal{Z}, \forall j \in \mathcal{J}, \\ &\underline{a}^{\mu_s} \leq \xi \text{ or } \xi \leq \overline{a}^{\mu_s}, \forall \xi \in \partial^o \mu_s(\tau), \forall \tau \in \mathcal{T}, \forall s \in \mathcal{S}, \end{aligned} \tag{8.9}$$

where $\mathcal{J} \triangleq \{1, \ldots, n_x\}$, $\mathcal{S} \triangleq \{1, \ldots, n_\mu\}$, and $\partial f_j^o(z)$ and $\partial^o \mu_s(\tau)$ are the Clarke sub-differentials of $f_j(\cdot)$ and $\mu_s(\cdot)$ at z and τ, accordingly, which are always non-empty, compact and convex sets (cf. Definition 8.1.11).

Given the setting in (8.8), we are interested in finding tight and tractable remainder-form upper and lower decomposition functions and their induced inclusion functions

---

[c] Note that mapping $\mu(\cdot)$ describes all the existing and a prior known or even manufactured/redundant constraints over the the states, observations and measurement noise signals or uncertain parameters

(cf. Definitions 8.1.1, 8.1.4, 8.1.7 and corollary 8.1.6) for a wide class of nonlinear (not necessarily J.S.S) vector fields (including $f(\cdot)$ and $\mu(\cdot)$ in (8.8)), as well as developing set-inversion/refinement algorithms, which together can be applied to several applications, e.g., reachability analysis as well as set-valued state estimation for the constrained nonlinear systems in the form of (8.8).

The problem of constructing remainder-form decomposition functions is three fold and can be stated as follows:

Problem 8.2.1. *Given the nonlinear (not necessarily Jacobian sign-stable) vector field* $f : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$, *find sufficient conditions for mixed-monotonicity of* $f(\cdot)$, *by constructing a family of remainder-form (i.e., additive) decomposition functions for* $f(\cdot)$.

Problem 8.2.2. *Derive lower and upper bounds for over-estimation of interval range of* $f(\cdot)$ *over* $\mathcal{Z}$, *using the remainder-form decomposition functions obtained in Problem 8.2.1.*

Problem 8.2.3. *Find the tightest decomposition function(s) in the family of remainder-form decomposition functions obtained in Problem 8.2.1 and compare them with the decomposition function proposed in [128] (recalled in Proposition 8.1.13), as well as natural, centered and mixed-centered natural inclusions (cf. Proposition 8.1.3 for their definitions).*

Further, the problem of decomposition function based set-inversion (refinement) can be cast as follows:

Problem 8.2.4. *Given the observation/constraint function* $\mu(\cdot)$, *the observation interval* $\mathcal{G}$, *the initial sets of states* $\mathcal{X}_0$ *and observation noise* $\mathcal{V}$, *and/or the observation signal* y, *develop an algorithm to refine* $\mathcal{X}_0$ *and find an interval superset of all the* *states in $\mathcal{X}_0$ which are compatible with the constraint set $\mathcal{G}$ and/or with the observation signal y.*

8.3 Main Results

8.3.1 Remainder-Form Decomposition Functions

In this section, we describe our approach to construct remainder (additive)-form decomposition functions for a given locally Lipschitz vector field $f \triangleq [f_1 \dots f_{n_x}]^\top : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$, that satisfies (8.9).

The idea is simple. We decompose each $f_j(\cdot)$ into two function components $g_j(\cdot)$ and $h_j(\cdot)$, i.e., $f_j(x) = g_j(x) + h_j(x), \forall j \in \mathcal{J} \triangleq \{1, \dots, n_x\}$, in a way that $g_j(\cdot)$ becomes a Jacobin sign-stable function in the sense of Definition 8.1.12. Doing this, it is shown (cf. Proposition A.0.4 in Appendix) that the *remainder* function $h_j(\cdot)$ is also turns out to be Jacobian sign-stable by construction, in the opposite direction of $g_j(\cdot)$, i.e., (8.2) holds for $g_j(\cdot)$, if and only if (8.3) holds for $h_j(\cdot)$ and (8.3) holds for $g_j(\cdot)$, if and only if (8.2) holds for $h_j(\cdot)$ (cf. Definition 8.1.12). Moreover, both $g_j(\cdot)$ and $h_j(\cdot)$ have tight decomposition functions (cf. Definition 8.1.15 and Corollary 8.1.17).Taking these facts into consideration, we can construct *bounding/embedding* functions for $f_j(\cdot)$ by tightly bounding $h_j(\cdot)$ and $g_j(\cdot) \triangleq f_j(\cdot) - h_j(\cdot)$, via evaluating them in the extreme (corner) points of the box $\mathcal{Z} = [\underline{z}, \overline{z}]$ and comparing the values (cf. Corollary 8.1.17). The following theorem summarizes this procedure and demonstrates that there are infinite number of remainder functions $h_j(\cdot)$ that can be used for our purpose in each dimension $j \in \mathcal{J}$. This, provides us a “family of decomposition functions” for $f(\cdot)$ in $\mathcal{Z}$, in the sense of Definition 8.1.5.

To increase readability, all proofs will be provided in Appendix.

Theorem 8.3.1 (A Family of Remainder-Form Decomposition Functions)**.** *Consider* *the locally Lipschitz continuous vector field* $f = [f_1 \dots f_{n_x}]^\top : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$ *that satisfies* (8.9) *and governs the dynamics of the state trajectory of the system* (8.8)*. Then,* $f(\cdot)$ *is mixed monotone with respect to the following family of remainder-form decomposition functions:*

$$\begin{cases} f_d(z, \hat{z}; \{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}) \triangleq [f_d^1(z, \hat{z}; \mathbf{m}^1) \ \dots f_d^{n_x}(z, \hat{z}; \mathbf{m}^{n_x})]^\top \\ f_d^j(z, \hat{z}; \mathbf{m}^j) = h_j(z^c_{\mathbf{m}^j}(\hat{z}, z)) + f_j(z^c_{\mathbf{m}^j}(z, \hat{z})) - h_j(z^c_{\mathbf{m}^j}(z, \hat{z})), \\ \{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}, j \in \mathcal{J} \triangleq \{1, \dots, n_x\}, \end{cases} \tag{8.10}$$

where $\mathbf{m}^j \in \mathbf{M}_j$ *is called a "supporting vector" and* $\mathbf{M}_j$ *is the corresponding set of supporting vectors for*

$$\mathcal{H}^j_{\mathbf{M}_j} \triangleq \{\tilde{h}_j(\cdot) : \mathcal{Z} \to \mathbb{R} | \partial^o \tilde{h}_j(z) \subseteq \mathbf{M}_j, \forall z \in \mathcal{Z}\}, \tag{8.11}$$

which is the family of appropriate locally Lipschitz "remainder" functions, i.e. all the locally Lipchitz functions whose Clarke sub-differential set over $\mathcal{Z}$ *is a subset of* $\mathbf{M}_j$*.*

Moreover, $\{h_j(\cdot)\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}^j_{\mathbf{M}_j}\}_{j=1}^{|\mathcal{J}|}$ *and* $\forall j \in \mathcal{J}$*:*

$$\mathbf{M}_j \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} | \mathbf{m}_i \leq \min(\underline{a}_i^{f_j}, 0) \ \textit{or} \ \mathbf{m}_i \geq \max(\overline{a}_i^{f_j}, 0), \forall i \in \{1, \dots, n_z\}\}, \tag{8.12}$$

if (8.8) *is a discrete-time system and*

$$\begin{aligned} \mathbf{M}_j \triangleq \{\mathbf{m} &\in \mathbb{R}^{n_z} | \mathbf{m}_i \leq \min(\underline{a}_i^{f_j}, 0) \ \textit{or} \ \mathbf{m}_i \geq \max(\overline{a}_i^{f_j}, 0), \forall i \in \{1, \dots, n_z\} \ \wedge \ i \neq j, \\ \mathbf{m}_j &= 0\}, \end{aligned} \tag{8.13}$$

if (8.8) *is a continuous-time system. Finally,*

$$\begin{aligned} z^c_{\mathbf{m}^j}(z, \hat{z}) &= D^{\mathbf{m}^j} z + (I_{n_z} - D^{\mathbf{m}^j})\hat{z}, \\ D^{\mathbf{m}^j}_{i,i} &= \mathrm{sgn}(\min(\underline{a}_i^{f_j}, 0) - \mathbf{m}^j{}_i), \forall i \in \{1, \dots, n_z\}, \end{aligned} \tag{8.14}$$

with a diagonal $D^{\mathbf{m}^j} \in \mathbb{D}^{n_z \times n_z}$*,* I_{n_z} *being the identity matrix in* $\mathbb{R}^{n_z \times n_z}$ *and* $\mathrm{sgn}(t) \triangleq 1$*, if* $t \geq 0$ *and* 0 *otherwise.*

It may worth mentioning that the small difference in the definition of the corresponding set of supporting vectors, $\mathbf{M}_j$, between discrete-time versus continuous-time cases in Theorem 8.3.1, i.e., the difference between (8.12) and (8.13), originates from the subtle difference between the definitions of decomposition functions in discrete-time and continuous-time cases (cf. Definition 8.1.5), where in the former case, the decomposition function for each dimension does not need to be monotone in the corresponding diagonal variable, and hence no "compensation" term/remainder is needed when $i = j$.

Although Theorem 8.3.1, "theoretically" introduces a family of decomposition functions in (8.10), but the results are not tractable yet, since to build such a family, we have to search over $\{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}$, a collection of unbounded sets of corresponding support vectors (cf. (8.12) and (8.13)). Hence, through the following lemma, we consider a "finite" sub-family of (8.10), to construct "computable" upper and lower decomposition functions (cf. Definition 8.1.7), by means of intersection, where we also show that considering such a sub-family is "equivalent" to consider the whole family (8.10) in terms of what will be obtained as the resultant decomposition functions.

Lemma 8.3.2 (Upper and Lower decomposition Functions)**.** *Suppose Theorem 8.3.1 holds. Then,*

$$\overline{f}_d(x,y) = \max_{\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}} f_d(x,y;\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}) = \max_{\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}^c_j\}_{j=1}^{|\mathcal{J}|}} f_d(x,y;\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}) \tag{8.15}$$

is an "upper decomposition function" for $f(\cdot)$ on $\mathcal{Z}$, and

$$\underline{f}_d(x,y) = \min_{\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}} f_d(x,y;\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}) = \min_{\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}^c_j\}_{j=1}^{|\mathcal{J}|}} f_d(x,y;\{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}) \tag{8.16}$$

is a "lower decomposition function" for $f(\cdot)$ *on* $\mathcal{Z}$*. In other words, every ordered tuple* $\underline{z} \le z \le \overline{z}$ *in* $\mathcal{Z}$*, satisfies* $\underline{f}_d(\underline{z}, \overline{z}) \le f(z) \le \overline{f}_d(\overline{z}, \underline{z})$ *and* $\underline{f}_d(z, z) = f(z) = \overline{f}_d(z, z)$*, where* $f_d(x, y; \{\mathbf{m}^j\}_{j=1}^{|\mathcal{J}|}$ *is given in* (8.10)*. Moreover,* $\mathbf{M}_j^c$ *is the corresponding set of supporting vectors for* $\mathcal{H}^j_{\mathbf{M}_j^c}$*, that is a "sub-family" of locally Lipschitz remainder functions in* (8.11)*. More precisely,* $\forall j \in \mathcal{J}$*:*

$$\mathcal{H}^j_{\mathbf{M}_j} \supset \mathcal{H}^j_{\mathbf{M}_j^c} \triangleq \{h_j(\cdot) : \mathcal{Z} \to \mathbb{R} | \partial^o h_j(z) \subseteq \mathbf{M}_j^c, \forall z \in \mathcal{Z}\}, \tag{8.17}$$

and $\mathbf{M}_j^c \subset \mathbf{M}_j$*, with*

$$\mathbf{M}_j^c \triangleq \{\mathbf{m} \in \mathbf{M}_j | \mathbf{m}_i = \min(\underline{a}_i^{f_j}, 0) \text{ or } \mathbf{m}_i = \max(\overline{a}_i^{f_j}, 0), \forall i \in \{1, \ldots, n_z\}\}, \tag{8.18}$$

if (8.8) *is a discrete-time system,*

$$\begin{aligned} \mathbf{M}_j^c \triangleq \{\mathbf{m} &\in \mathbb{R}^{n_z} | \mathbf{m}_i = \min(\underline{a}_i^{f_j}, 0) \text{ or } \mathbf{m}_i = \max(\overline{a}_i^{f_j}, 0), \forall i \in \{1, \ldots, n_z\} \text{ and } i \neq j, \\ \mathbf{m}_j &= 0\}, \end{aligned} \tag{8.19}$$

if (8.8) *is a continuous-time system and* $\mathcal{H}^j_{\mathbf{M}_j}, \mathbf{M}_j$ *defined in Theorem 8.3.1.*

As can be observed, to derive upper and lower decomposition functions in (8.15) and (8.16), we can "equivalently" search over $\{\mathbf{M}_j^c\}_{j=1}^{|\mathcal{J}|}$, which is a collection of finitely and countable sets (cf. (8.18) and (8.19)). This makes our computation tractable.

8.3.2 Error Bounds

Next, we evaluate the tightness of our proposed remainder-form decomposition functions, where we use the standard notion of Hausdorff distance-function, $q(.,.) : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}_+$, defined in Definition 8.1.14.

In particular, we are interested to derive lower and upper bounds for the error of approximation (over-estimation) of the range of function $f(\cdot)$, using our proposed family of remainder-form decomposition functions given in (8.10). We first provide a lower bound, as well as two upper bounds, a smaller and a bigger one, via Theorem 8.3.3. Then, through Lemma 8.3.4, we show that the lower bound is indeed achievable using some specific form of decomposition functions in (8.10). Furthermore, we prove that the decomposition function given in [128] belongs to the family in (8.10) and if used, minimizes the bigger upper bound given in Theorem 8.3.3.

Theorem 8.3.3 (Error Bounds)**.** *Suppose that all the assumptions in Theorem 8.3.1 and Lemma 8.3.2 are satisfied. Let* $\mathbb{H} \triangleq \{h_j(\cdot)\}_{j \in \mathcal{J}} \in \{\mathcal{H}^j_{\mathbf{M}_j}\}_{j \in \mathcal{J}}$ *be a set of remainder functions (cf. Theorem 8.3.1),* $V_f(\mathcal{Z}) \triangleq [\underline{v}^f_{\mathcal{Z}}, \overline{v}^f_{\mathcal{Z}}]$ *be the tightest enclosing interval to the true range (image) of* $f(\cdot)$ *over* $\mathcal{Z} = [\underline{z}, \overline{z}] \in \mathbb{IR}^{n_z}$*, and for any corresponding pair of sets of supporting vectors* $\mathbb{M}_l \triangleq \{m^j_l\}_{j=1}^{|\mathcal{J}|}$ *and* $\mathbb{M}_u \triangleq \{m^j_u\}_{j=1}^{|\mathcal{J}|}$*, both in* $\{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}$*,* $W_f^{\mathbb{M}_l,\mathbb{M}_u}(\mathcal{Z})$ *is an over-approximation interval for* $V_f(\mathcal{Z})$*, using the family of decomposition functions in* (8.10)*, (cf. Corollary 8.1.8), i.e.,*

$$V_f(\mathcal{Z}) \subseteq W_f^{\mathbb{M}_l,\mathbb{M}_u}(\mathcal{Z}) \triangleq [f_d(\underline{z}, \overline{z}; \mathbb{M}_l), f_d(\overline{z}, \underline{z}; \mathbb{M}_u)]. \tag{8.20}$$

Then, the following series of inequalities hold:

$$\underline{q}(W, V; \mathbb{H}) \leq q(W_f^{\mathbb{M}_l,\mathbb{M}_u}(\mathcal{Z}), V_f(\mathcal{Z})) \tag{8.21}$$

$$q(W_f^{\mathbb{M}_l,\mathbb{M}_u}(\mathcal{Z}), V_f(\mathcal{Z})) \leq \overline{q}(W, V; \mathbb{H}) \leq \hat{\overline{q}}(W, V; \mathbb{H}), \tag{8.22}$$

where $s(W,V;\mathbb{H}) = \max_{j \in \mathcal{J}} s_j(W,V;h_j), \forall s \in \{\underline{q}, \hat{\overline{q}}, \overline{q}\}$ *and*

$$
\begin{aligned}
\underline{q}_j(W,V;h_j) &\triangleq \frac{1}{2}[\min_{m \in \mathbf{M}_j^c} l_1^j(m) + \min_{m \in \mathbf{M}_j^c} l_2^j(m) + \Delta f_j^{true}], \\
\hat{\overline{q}}_j(W,V;h_j) &\triangleq \min_{m \in \mathbf{M}_j^c} \Delta h_{j;m}^c, && (8.23) \\
\overline{q}_j(W,V;h_j) &\triangleq \min_{m \in \mathbf{M}_j^c} \min\{\Delta h_{j;m}^c, \Delta h_{j;m}^c + \Delta f_{j;m}^c\}, && (8.24) \\
l_1^j(m) &\triangleq \Delta h_{j;m}^c + f_j(\underline{z}_{j;m}^c), \\
l_2^j(m) &\triangleq \Delta h_{j;m}^c - f_j(\overline{z}_{j;m}^c), \\
\overline{z}_{j;m}^c &\triangleq z_m^c(\overline{z}, \underline{z}), \underline{z}_{m;j}^c \triangleq z_m^c(\underline{z}, \overline{z}), \\
\Delta f_{j;m}^c &\triangleq f_j(\underline{z}_{j;m}^c) - f_j(\overline{z}_{j;m}^c), \\
\Delta h_{j;m}^c &\triangleq h_j(\overline{z}_{j;m}^c) - h_j(\underline{z}_{j;m}^c), \\
\Delta f_j^{true} &\triangleq \underline{f}_j^{true} - \overline{f}_j^{true},
\end{aligned}
$$

with $z_m^c(.,.)$ *defined in Theorem 8.3.1,* $\underline{f}_j^{true} \triangleq \min_{z \in \mathcal{Z}} f_j(z)$, $\overline{f}_j^{true} \triangleq \max_{z \in \mathcal{Z}} f_j(z)$ *and* $q(\cdot,\cdot)$ *is the Hausdorff distance function defined in Definition 8.1.14.*

As mentioned before, we show later that "the best of" the remainder-form decomposition functions in (8.10) minimizes the lower bound $\underline{q}(\cdot,\cdot;\cdot)$ derived in Theorem 8.3.3 and attains the lowest value of them and so is tighter than T_L (cf. Proposition 8.1.13) that will be shown that minimizes $\hat{\overline{q}}(\cdot,\cdot;\cdot)$, which is an upper bound (as opposed to $\underline{q}(\cdot,\cdot;\cdot)$ that is a lower bound), and also, is less tight than the smaller upper bound $\overline{q}(\cdot,\cdot;\cdot)$ by (8.22).

8.3.3 Linear Remainders

In this section, the problem of finding the best/tightest decomposition functions among the ones in the family (8.10) is considered. Particularly, through lemma 8.3.4 we show that no set of remainder functions $\mathbb{H} = \{h_j(\cdot)\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}_{\mathbf{M}_j}^j\}_{j=1}^{|\mathcal{J}|}$ along with any set of corresponding supporting vectors $\{m^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}$, can make a decomposition function that achieves better (i.e smaller) lower bound $\underline{q}(.,.;\mathbb{H})$, than the set of linear remainders $\{\tilde{h}_j(x)\}_{j=1}^{|\mathcal{J}|} = \{\tilde{m}^j x\}_{j=1}^{|\mathcal{J}|}$, for some $\tilde{m}^j \in \mathbf{M}_j^c, \forall j \in \mathcal{J}$. This, enables us to restrict our search to "linear" remainder functions (with their slopes belongs to the finite and countable set $\mathbf{M}_j^c$), when finding the best remainder-form decomposition functions from the family (8.10).

Lemma 8.3.4 (Remainder Form Decomposition Functions with Linear Remainders)**.** *Consider* $f : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *that satisfies* (8.9)*, as well as the family of remainder functions in* (8.11)*, and let* $\mathcal{J} \triangleq \{1, \dots, n_x\}$*. Then, no set of remainder functions* $\mathbb{H} \triangleq \{h_j(\cdot)\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}^j_{\mathbf{M}_j}\}_{j=1}^{|\mathcal{J}|}$ *along with any set of corresponding supporting vectors* $\{m^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}$ *can construct a remainder-form decomposition function in the form of* (8.10) *that achieves a smaller lower bound* $\underline{q}(.,.;\mathbb{H})$*, than the set of linear remainders* $\{\tilde{h}_j(x)\}_{j=1}^{|\mathcal{J}|} = \{\tilde{m}^j x\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}^j_{\mathbf{M}_j^c}\}_{j=1}^{|\mathcal{J}|}$*, where* $\tilde{m}^j \in \mathbf{M}_j^c$ *and* $\forall j \in \mathcal{J}, \forall i \in \{1, \dots, n_z\}$

$$\tilde{m}_i^j = \begin{cases} \min(\underline{a}_i^{f_j}, 0), & \textit{if } m_i^j \leq \min(\underline{a}_i^{f_j}, 0), \\ \max(\overline{a}_i^{f_j}, 0), & \textit{if } m_i^j \geq \max(\overline{a}_i^{f_j}, 0), \end{cases} \tag{8.25}$$

with $\mathcal{H}^j_{\mathbf{M}_j}$*,* $\mathbf{M}_j$*,* $\mathcal{H}^j_{\mathbf{M}_j^c}$*,* $\mathbf{M}_j^c$ *and* $\underline{q}(.,.;\mathbb{H})$ *given in* (8.11)–(8.14)*,* (8.17)–(8.19) *and Theorem 8.3.3, respectively.*

Lemma 8.3.4 guarantees that in order to obtain the tightest possible decomposition function in the form of (8.10), it is sufficient to only search over "linear" remainders $h_j(z) = m^{j\top} z, \forall m^j \in \mathbf{M}_j^c$ -where the search space is the finite and countable set $\mathbf{M}_j^c$- and to find the one that returns the tightest possible lower bound. Hence, the optimal search is computable/tractable. This, leads us to Algorithm **??**, which by Theorems 8.3.1–8.3.3 and Lemma 8.3.4, results in the tightest possible remainder-form decomposition function that can be obtained from the family of *linear remainder-form Jacobian sign-stable decomposition functions*, (8.10). For the sake of simplicity and clarity, from now on, we call the optimal remainder-form decomposition functions resulted from Algorithm 7, the T_R inclusion functions.

Algorithm 7 Remainder-Form Decomposition Functions

```
function $T_R(f(\cdot), \overline{a}^f, \underline{a}^f, z_1, z_2)$
    for $j = 1$ to $n_x$ do
        Initialize: $\overline{f}_j \leftarrow \infty, \underline{f}_j \leftarrow -\infty; \overline{g}_j \leftarrow -\infty, \underline{g}_j \leftarrow \infty;$
        if (8.8) is a discrete-time system then
        $\mathbf{M}_j^c \triangleq \{\mathbf{m} \in \mathbf{M}_j | \mathbf{m}_i = \min(\underline{a}_i^{f_j}, 0) \vee \mathbf{m}_i = \max(\overline{a}_i^{f_j}, 0)$
        $\forall i \in \{1, \ldots, n_z\}\}.$
        end if
        if (8.8) is a continuous-time system then
        $\mathbf{M}_j^c \triangleq \{\mathbf{m} \in \mathbf{M}_j | \mathbf{m}_i = \min(\underline{a}_i^{f_j}, 0) \vee \mathbf{m}_i = \max(\overline{a}_i^{f_j}, 0)$
        $\forall i \in \{1, \ldots, n_z\}$ and $i \neq j, \mathbf{m}_j = 0\},$
        end if
        for $m \in \mathbf{M}_j^c$ do
        $\overline{h}(z_1, z_2) \leftarrow \max(m, \mathbf{0}_{n_z})^\top z_1 + \min(m, \mathbf{0}_{n_z})^\top z_2;$
        $\underline{h}(z_1, z_2) \leftarrow \max(m, \mathbf{0}_{n_z})^\top z_2 + \min(m, \mathbf{0}_{n_z})^\top z_1;$
            for $i = 1$ to $n_z$ do
                if $m_i = \min(\underline{a}_i^{f_j}, 0)$ then
            $\overline{z}_i^c \leftarrow z_{1,i};\ \underline{z}_i^c \leftarrow z_{2,i};$
                else
            $\overline{z}_i^c \leftarrow z_{2,i};\ \underline{z}_i^c \leftarrow z_{1,i};$
                end if
        $\overline{g}(z_1, z_2) \leftarrow \max(\overline{g}(z_1, z_2), f_j(\overline{z}^c) - m^\top \overline{z}^c);$
        $\underline{g}(z_1, z_2) \leftarrow \min(\underline{g}, f_j(\underline{z}^c) - m^\top \underline{z}^c);$
            end for
        $\overline{f}_j(z_1, z_2) \leftarrow \min(\overline{f}_j(z_1, z_2), \overline{g}(z_1, z_2) + \overline{h}(z_1, z_2));$
        $\underline{f}_j(z_1, z_2) \leftarrow \max(\underline{f}_j(z_1, z_2), \underline{g}(z_1, z_2) + \underline{h}(z_1, z_2));$
        end for
        return $\overline{f}_j(z_1, z_2), \underline{f}_j(z_1, z_2);$
    end for
end function
```

8.3.4 Convergence Rate and Further Refinement

In this subsection, considering the approximation of the range of a locally Lipschitz function using the T_R, we study the convergence rate at which the approximation error goes to zero, when the domain interval width/diameter shrinks. We show that using the T_R, the error converges at least linearly, i.e., as good as natural inclusions [84, Chapter 6]. Moreover, applying the "subdivision principle" [7], we show that we can further refine the convergence rate of our approximation from linear to exponential by subdividing the interval domain of the function into sub-intervals, applying T_R on all the sub-intervals and taking the union of all the resultant over-approximations of the sub-intervals. We first introduce the notions of convergence rate, inspired by the work in [7].

Definition 8.3.5 (Convergence Rate)**.** *The generic inclusion function* $T : \mathbb{IR}^{n_z} \to \mathbb{IR}^{n_x}$ *has a convergence rate of* $\alpha > 0$*, if for any locally Lipchitz vector field* $f \triangleq [f_1 \dots f_{n_x}]^\top : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$*, there exists* $\beta > 0$*, such that*

$$q(W_f^T(\mathcal{Z}), V^f(\mathcal{Z})) \leq \beta \| d(\mathcal{Z}) \|_\infty^\alpha, \tag{8.26}$$

where $W_f^T(\mathcal{Z})$ *is the interval over-approximation of the range of* $f(\cdot)$*, using* T*,* $V^f(\mathcal{Z})$ *is the tightest interval enclosure to the true range of* $f(\cdot)$*,* $q(\cdot,\cdot)$ *is the Hausdorff distance function (cf. Definition 8.1.14),* $d(\mathcal{Z}) \triangleq \overline{z} - \underline{z}$ *and* $\|d(\mathcal{Z})\|_\infty$ *is the diameter of* $\mathcal{Z}$*.*

Next, similar to [7], we can apply the subdivision principle to further refine the convergence rate. To this end, we can represent $\mathcal{Z} \in \mathbb{IR}^{n_z}$ as the union of k^{n_z} interval vectors $\mathcal{Z}^l, l = 1, \dots, k^{n_z}$, such that $d(\mathcal{Z}_i^l) = \frac{d(\mathcal{Z}_i)}{k}$ for $i = 1, \dots, n_z$ and $l = 1, \dots, k^{n_z}$. Then, defining $V^f(\mathcal{Z}; k) \triangleq \bigcup_{l=1}^{k^{n_z}} V^f(\mathcal{Z}^l)$ and $W_T^f(\mathcal{Z}; k) \triangleq \bigcup_{l=1}^{k^{n_z}} W_T^f(\mathcal{Z}^l)$, the following results hold.

Theorem 8.3.6 (Convergence Rate and Subdivision Principle for T_R). *For any locally Lipchitz vector field* $f \triangleq [f_1 \dots f_{n_x}]^\top : \mathcal{Z} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^{n_z} \to \mathbb{R}^{n_x}$*, there exists* $\beta^f_R > 0$ *such that:*

$$q(W^{T_R}_f(\mathcal{Z}), V^f(\mathcal{Z})) \leq \beta^f_R \|d(\mathcal{Z})\|_\infty. \tag{8.27}$$

Moreover, applying subdivision principle and with $\mathbb{N} \ni k \geq 1$ *number of divisions in each dimension, there exists* $\gamma^f_R > 0$ *such that*

$$q(W^{T_R}_f(\mathcal{Z}; k), V^f(\mathcal{Z}; k)) \leq \frac{\gamma^f_R}{k}. \tag{8.28}$$

The results in Theorem 8.3.6 imply that without subdivision, the approximation converges to the true tightest enclosing interval at least linearly, and with subdivisions and by adding the number of divisions, the convergence rate can increase exponentially.

8.3.5 Set-Inversion

The remainder-form decomposition functions returned by Algorithm 7 can be used directly to over-approximate unconstrained reachable sets of a dynamic system governed by the vector field $f(\cdot)$. However, when additional state constraint information is available (e.g. sensor observations/measurements in state estimation problems, known safety constraints from system design and manufactured constraints [130, 5, 6, 102, 105]), an extra set-inversion (i.e., update or refinement) step will allow us to take the advantage of the constraints to shrink/update the propagated sets, i.e., to obtain a tighter subset of the propagated set that is compatible/consistent with the given constraints. Mathematically speaking, given the constraint/observation function $\mu(\cdot)$, constraint/observation maximal and minimal values $\overline{y}, \underline{y}$ and the propagated interval $\mathcal{Z}^p = [\underline{z}^p, \overline{z}^p]$, we need to find the following updated/refined interval $\mathcal{Z}^* \subseteq \mathcal{Z}^p$, where

$$\mathcal{Z}^p \supseteq \mathcal{Z}^* \triangleq [\underline{z}^*, \overline{z}^*] \supseteq \{z \in \mathcal{Z}^p | \underline{y} \leq \mu(z) \leq \overline{y}\}. \tag{8.29}$$

Finding $\mathcal{Z}^*$ in (8.29) is called the "set-inversion" problem [57]. To the best of our knowledge, existing set-inversion algorithms/operators are either using (conservative) natural inclusions (SIVIA [57, Chapter 3]) or are only applicable if relatively restrictive monotonicity assumptions hold ($\mathcal{I}_G$ [130, Algorithm 1]). In this section, leveraging our proposed decomposition functions, we develop a set-inversion algorithm that is applicable for some general class of locally Lipschitz continuous systems. It is also notable that as opposed to SIVIA and $\mathcal{I}_G$, our set-inversion algorithm can also be used with any applicable inclusion functions (such as T_N, T_C, T_M, T_L, T_O) or the best of them (by intersecting the returned reachable sets by all of them) replacing the T_R. Our developed set-inversion algorithm based on T_R is summarized in Algorithm 8.

The main idea behind Algorithm 8 is simple and intuitive. Starting form the propagated interval and using bisection, for each dimension, it shrinks the compatible interval form below and/or above, if $\underline{\mu}$, the minimal value of the interval approximation of the range of the observation/constraint mapping $\mu(\cdot)$, is strictly greater than $\overline{y}$, the maximal value of the observations/measurements interval, or, if $\overline{\mu}$, the maximal value of the interval approximation of the range of the observation/constraint mapping $\mu(\cdot)$, is strictly smaller than $\underline{y}$, the minimal value of the observations/measurements interval (cf. lines 5 and 10 in Algorithm 8). Repeating this procedure along with the bisection, the boxes that are determined as "inconsistent" with the observation set are being ruled out. Lemma 8.3.7 shows that this algorithm indeed returns $\mathcal{Z}^*$ in (8.29).

Algorithm 8 Set-Inversion Based on Mixed-Monotone Inclusion Functions

```
function SET-INV($\mu(\cdot), \overline{a}^{\mu}, \underline{a}^{\mu}, \overline{z}^{p}, \underline{z}^{p}, \overline{y}, \underline{y}, \epsilon$)
    Initialize: $\overline{z}^{*} \leftarrow \overline{z}^{p}, \underline{z}^{*} \leftarrow \underline{z}^{p}$;
    for $i = 1$ to $n_z$ do
        $z_l \leftarrow \underline{z}_i^{*}$; $z_u \leftarrow \overline{z}_i^{*}$;
        while $z_u - z_l > \epsilon$ do
            $z_m \leftarrow \frac{1}{2}(z_u + z_l)$; $\overline{\xi} \leftarrow \overline{z}^{*}$; $\underline{\xi} \leftarrow \underline{z}^{*}$; $\underline{\xi}_i \leftarrow z_m$;
            $(\overline{\mu}, \underline{\mu}) \leftarrow T_R(\mu(\cdot), \overline{\xi}, \underline{\xi}, \overline{a}^{\mu}, \underline{a}^{\mu})$;
            if $(\overline{\mu} < \underline{y}) \lor (\underline{\mu} > \overline{y})$ then
                $z_u \leftarrow z_m$; $\overline{z}_i^{*} \leftarrow z_u$;
            else
                $z_l \leftarrow z_m$;
            end if
        end while
        $z_l \leftarrow \underline{z}_i^{*}$; $z_u \leftarrow \overline{z}_i^{*}$;
        while $z_u - z_l > \epsilon$ do
            $z_m \leftarrow \frac{1}{2}(z_u + z_l)$; $\overline{\xi} \leftarrow \overline{z}^{*}$; $\underline{\xi} \leftarrow \underline{z}^{*}$; $\overline{\xi}_i \leftarrow z_m$;
            $(\overline{\mu}, \underline{\mu}) \leftarrow T_R(\mu(\cdot), \overline{\xi}, \underline{\xi}, \overline{a}^{\mu}, \underline{a}^{\mu})$;
            if $(\overline{\mu} < \underline{y}) \lor (\underline{\mu} > \overline{y})$ then
                $z_l \leftarrow z_m$; $\underline{z}_i^{*} \leftarrow z_l$;
            else
                $z_u \leftarrow z_m$;
            end if
        end while
    end for
    return $\overline{z}^{*}, \underline{z}^{*}$;
end function
```

Lemma 8.3.7. *Consider function* $\mu : \mathcal{Z}^p \triangleq [\underline{z}^p, \overline{z}^p] \subset \mathbb{R}^{n_z} \to [\underline{y}, \overline{y}] \subset \mathbb{R}^{n_y}$ *that satisfies* (8.9) *with its upper and lower values for its Clarke generalized gradients being* $\overline{a}^{\mu}$ *and* $\underline{a}^{\mu}$, *respectively. Then the pair* $(\overline{z}^*, \underline{z}^*)$ *returned by Algorithm* **??** *satisfies* (8.29).

8.4 Comparison with Other Inclusion Functions

8.4.1 Comparison with the T_L Inclusion Functions

In this subsection we compare the performances of T_R and T_L (cf. Proposition 8.1.13) through the following Theorem 8.4.1. We show that the decomposition function f_d^L introduced in [130] and recalled in Proposition 8.1.13, belongs to the family of the remainder-form decomposition functions (8.10) and hence, T_L cannot be tighter than T_R, which is the tightest decomposition function that belongs to (8.10)

Theorem 8.4.1 (T_L vs T_R)**.** *Suppose all the assumptions in Theorem 8.3.1 hold. Then, the following statements are true.*

(i) There exists a set of supporting vectors $\{m_L^j\}_{j=1}^{|\mathcal{J}|} \in \{\mathbf{M}_j^c\}_{j=1}^{|\mathcal{J}|}$ *such that the decomposition function* $f_d^L(\cdot,\cdot)$ *(introduced in [128, Theorem 2] and recalled in Proposition 8.1.13), coincides with a remainder-form decomposition function in the form of* (8.10) *with the linear remainder functions* $h_j(z) = m_L^j z, \forall j \in \mathcal{J}$*. In other words,* $f_d^L(\cdot,\cdot)$ *belongs to the family of decomposition functions* (8.10).

(ii) $\{h_j^L(x)\}_{j=1}^{|\mathcal{J}|} = \{m_L^j x\}_{j=1}^{|\mathcal{J}|}$ *is a minimizer of the error upper bound* $\hat{\overline{q}}(W, V; \mathbb{H})$*, introduced in* (8.23) *in Theorem 8.3.1, among all the sets of remainder functions* $\{h_j(\cdot)\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}_{\mathbf{M}_j}^j\}_{j=1}^{|\mathcal{J}|}$*, i.e.,* $\hat{\overline{q}}(W, V; \{m_L^j x\}_{j=1}^{|\mathcal{J}|}) = \min_{\mathbb{H} \in \{\mathcal{H}_{\mathbf{M}_j}^j\}_{j=1}^{|\mathcal{J}|}} \hat{\overline{q}}(W, V; \mathbb{H})$.

(iii) The optimal remainder-form decomposition function T_R *is always tighter than (at least as good as) the inclusion function* T_L*, induced by the decomposition function* f_d^L.

8.4.2 Comparison with T_N, T_C and T_M Inclusion Functions

In this section we compare the performance of natural inclusions and their modifications, i.e., T_N, T_C, T_M with T_R, via computing the over-approximation for the range of some example functions. It is worth mentioning that we were not able to derive any theoretical results for these comparisons and in fact our simulation results showed that depending on function and its corresponding considered domain, one of them can be tighter than the other or vice versa. However, in some cases, reflected in the following examples, the T_R most likely returns tighter intervals.

Composition of Non-Elementary Functions.

In case that the considered vector field is not a composition of "elementary functions" (e.g., simple monomials, $\sin(\cdot)$, $\cos(\cdot)$, monotone functions, etc), then it is either impossible to compute corresponding natural inclusion functions or conservative over approximations for bounds of constituent functions are expected, which lead to poor inclusion functions, i.e., large errors. In these cases, it is most likely that T_R returns better bounds. The following example describes one of such functions.

Example 8.4.2. *Consider the vector field* $f : [1,3] \subset \mathbb{R} \to \mathbb{R}$*, where* $f(x) = x \arctan(x^2 - 2x + 5)$*, which is a composition of non-elementary functions. In this case,* T_N, T_C, T_M, T_L *and* T_R *return* $[-4.7124, 4.7124]$*,* $[1.3258, 4.3393], [1.3187, 4.2475]$, $[1.2835, 2.9461]$ *and* $[1.1760, 2, 7468]$*, respectively.*

"Almost" Sign-Stable Functions.

In case that $f(\cdot)$ can be decomposed into a Jacobian sign-stable constituent and a relatively small additive perturbation, most likely T_R returns tighter bounds than T_N, T_C and T_M. For instance, consider the following example.

Example 8.4.3. *Consider the vector field* $f : [-1, 3] \subset \mathbb{R} \to \mathbb{R}$*, where* $f(x) = x^3 - 0.1x$*, that is a monotone increasing (and hence Jacobian sign-stable) function on its domain, except on the short interval* $[-\sqrt{\frac{0.1}{3}}, \sqrt{\frac{0.1}{3}}]$*. For this example,* T_N, T_C, T_M, T_L *and* T_R *return*

$$[-8.9000, 27.0100], [-49.9000, 54.7000], [-49.9000, 54.7000], [-1.0300, 26.0100],$$

and [-1.0300,26.0100], accordingly.

Vector Fields with Several Additive Terms.

It is well-known in the literature that natural, centered and mixed-centered inclusions perform worse for the functions with several additive terms, compared to the ones with the lesser additive terms [57, 84]. This is not necessarily true for the performance of T_R. The following example illustrates this fact, where a function with several additive terms is considered.

Example 8.4.4. *Consider the vector field* $f : [-2, 2] \times [-2, 2] \times [-2, 2] \subset \mathbb{R}^3 \to \mathbb{R}$*, where* $f(x) = x_1x_2x_3 + x_1^2x_2 + x_2^2x_3 + x_3^2x_1 + x_1^2x_3 + x_3^2x_2 + x_2^2x_1 + x_1^3 + x_2^3 + x_3^3$*. For this function,* T_N, T_C, T_M, T_L *and* T_R *result in*

$$[-80, 80], [-76.45, 76.45], [-73.62, 73.62], [-176, 176]$$

and $[-54.4, 54.4]$*, respectively.*

Existence of Closed-form Decomposition Functions.

Finally, using our approach, we are able to provide closed form formulation for a family of decomposition and inclusion functions for a wide class of vector fields. This can be analytically beneficial, e.g., in convergence analysis for reachable sets or stability analysis in interval observer design. This is in contrast with natural, centered and mixed-centered inclusions, where a closed form general formulation, which is independent of the considered function, is not available.

8.5 Applications

8.5.1 Application to Constrained Reachability Analysis

Consider the following constrained bounded-error system:

$$\begin{aligned} x^+ \quad &= f(x_t, u_t, w_t), \\ \mu(x_t, u_t, v_t) \quad &\in G_t, \ \text{at } t = 0, T, 2T, \ldots \end{aligned} \tag{8.30}$$

where $x_t \in \mathbb{R}^{n_x}$ with $x_0 \in [\underline{x}_0, \overline{x}_0]$ and $u_t \in \mathbb{R}^{n_u}$ are state and known input signals, $w_t \in [\underline{w}, \overline{w}] \subset \mathbb{R}^{n_w}$ and $v_t \in [\underline{v}, \overline{v}] \subset \mathbb{R}^{n_v}$ are bounded process disturbance and noise signals, $G_t = [\underline{g}_t, \overline{g}_t] \in \mathbb{R}^{n_g}$ is the time-varying state interval constraint, $f(\cdot) : \mathbb{R}^{n_x+n_u+n_w} \to \mathbb{R}^{n_x}$, $\mu(\cdot) : \mathbb{R}^{n_x+n_u+n_v} \to \mathbb{R}^{n_g}$ are known vector fields and T is the sampling time at which the constraints are measured/observed. The following proposition shows how to apply Algorithms 7–8, i.e. the mixed-monotone decomposition functions and the set-inversion algorithm, to compute approximations of the reachable sets of the states for system (8.30).

Proposition 8.5.1. *Consider the system* (8.30) *and suppose that* f, g *are mixed-monotone and satisfy* (8.9) *with* $(\underline{a}^f, \overline{a}^f), (\underline{a}^g, \overline{a}^g)$*, respectively, and* ϵ *is a small positive chosen threshold. Then, the upper and lower bounds for the state intervals, i.e.,* $\overline{x}_t, \underline{x}_t$*, such that* $\underline{x}_t \leq x_t \leq \overline{x}_t$*, can be found at time* t*, as follows:*

$$\overline{x}_t = \overline{z}^u_t(1 : n_x), \ \underline{x}_t = \underline{z}^u_t(1 : n_x),$$

where $\overline{z}^u_t(1 : n_x)$ *and* $\overline{z}^u_t(1 : n_x)$ *are two vectors, consisting of the first* n_x *arguments of*

$\overline{z}_t$ *and* $\underline{z}_t$*, respectively,*

$$(\overline{z}^u_t, \underline{z}^u_t) = \mathit{SET\text{-}INV}(\mu(\cdot), \overline{a}^\mu, \underline{a}_\mu, \overline{z}^p_t, \underline{z}^p_t, \overline{g}_t, \underline{g}_t, \epsilon),$$
$$\overline{z}^p_t = [\overline{x}^{p\top}_t u^\top_t \overline{v}^\top]^\top, \quad \underline{z}^p_t = [\underline{x}^{p\top}_t u^\top_t \underline{v}^\top]^\top,$$

and

$$(\overline{x}^p_t, \underline{x}^p_t) = T_R(f(\cdot), \overline{a}^f, \underline{a}^f, [\overline{x}^\top_{t-1} u^\top_{t-1} \overline{w}^\top]^\top, [\underline{x}^\top_{t-1} u^\top_{t-1} \underline{w}^\top]^\top),$$

if (8.30) *is a discrete-time system and* $(\overline{x}^p_t, \underline{x}^p_t)$ *is the solution of the following so-called dynamical "embedding system", with the initial values* $(\overline{x}^p_0, \underline{x}^p_0) = (\overline{x}_0, \underline{x}_0)$*, at time* t*:*

$$\begin{bmatrix} \dot{\overline{x}}^p \\ \dot{\underline{x}}^p \end{bmatrix} = \begin{bmatrix} \overline{f}_d(\overline{x}^p, \underline{x}^p, u, \overline{w}, \underline{w}) \\ \underline{f}_d(\underline{x}^p, \overline{x}^p, u, \underline{w}, \overline{w}) \end{bmatrix},$$

if (8.30) *is a continuous-time system, where* $(\overline{f}_d(z_1, z_2), \underline{f}_d(z_1, z_2)) = T_R(f(\cdot), \overline{a}^f, \underline{a}^f, z_1, z_2)$*.*

8.5.2 Application to State Estimation

Now, consider the following bounded-error system:

$$\begin{aligned} x^+ &= f(x_t, u_t, w_t), \\ y_t &= \tilde{\mu}(x_t, u_t) + V v_t, \text{ at } t = 0, T, 2T, \ldots \end{aligned} \tag{8.31}$$

where its state equation is similar to the system (8.30), but instead of state constraint, an observation/measurement signal $y_t \in \mathbb{R}^{n_y}$ is known/measured and given at time steps kT, which is a function of the state x_t, through the known observation function $\tilde{\mu}(\cdot) : \mathbb{R}^{n_x+n_u} \to \mathbb{R}^{n_y}$, known matrix $V \in \mathbb{R}^{n_y \times v_v}$ and measurement noise signal $v_t \in [\underline{v}, \overline{v}] \subset \mathbb{R}^{n_v}$. It can be easily verified that the observation equation can be equivalently written as $\tilde{\mu}(x_t, u_t) \in \tilde{G}_t \triangleq [y_t - \overline{s}, y_t - \underline{s}]$, where $\overline{s} = V^+\overline{v} - V^-\underline{v}$, $\underline{s} = V^+\underline{v} - V^-\overline{v}$, $V^+ \triangleq \max(V, \mathbf{0}_{n_v})$ and $V^- \triangleq V^+ - V$, with $\mathbf{0}_{n_v}$ being a zero vector in $\mathbb{R}^{n_v}$. This, transforms the system (8.31) into the form of (8.30) and hence, Proposition 8.5.1 is applicable, which returns $\underline{x}_t$ and $\overline{x}_t$.

8.6 Simulation Results

In this section, we compare the performance of T_N (natural inclusions, cf. Proposition 8.1.2), T_C, T_M (centered and mixed-centered inclusions, cf. Proposition 8.1.3), T_L (decomposition functions proposed in [128], cf. Proposition 8.1.13), T_R (our proposed remainder-form decomposition function in Algorithm 7), T_{S_1} (the first proposed bounding approach in [130], if applicable), T_{S_2} (the second proposed approach in [130], if applicable) and T_O (the tight decomposition functions proposed in [129, Theorem 2] for discrete-time and in [2, Theorem 1] for continuous-time syatems, if applicable, cf. Proposition 8.1.16) in computing the reachable sets of several unconstrained and constrained dynamical systems in the form of (8.8).

8.6.1 The Van Der Pol System

Consider the following discretized well-known Van der Pol system [105]:

$$\begin{aligned} x_{1,k+1} &= x_{1,k} + \delta_t x_{2,k}, \\ x_{2,k+1} &= x_{2,k} + \delta_t((1 - x_{1,k}^2)x_{2,k} - x_{1,k}), \end{aligned} \tag{8.32}$$

where $\delta_t = 0.1$, $1.15 \leq x_{1,0} \leq 1.4$ and $2.05 \leq x_{2,0} \leq 2.3$. Starting form these initial intervals, the results for computing reachable intervals, using several applicable inclusion functions, are depicted in Fig. 8.1. Unfortunately, T_{S_1} and T_{S_2} are not valid (applicable) here, due to the lack of required monotonicity assumptions (cf. [130, conditions (6) and (16)]). Moreover, as expected, T_O (that is computable here since the corresponding optimization problems can be analytically solved due to computability of all the critical points of the vector fields) returns the tightest bounds. Our proposed T_R returns tighter bounds compared to the ones returned by all the other remaining applicable ones except the T_O and seems to be a significant refinement for natural, centered and mixed-centered inclusions as well as the T_L. We also computed the reachable sets using the best of all approaches except the tightest one to see how close we can reach into the tight bounds, without using the tight T_O.

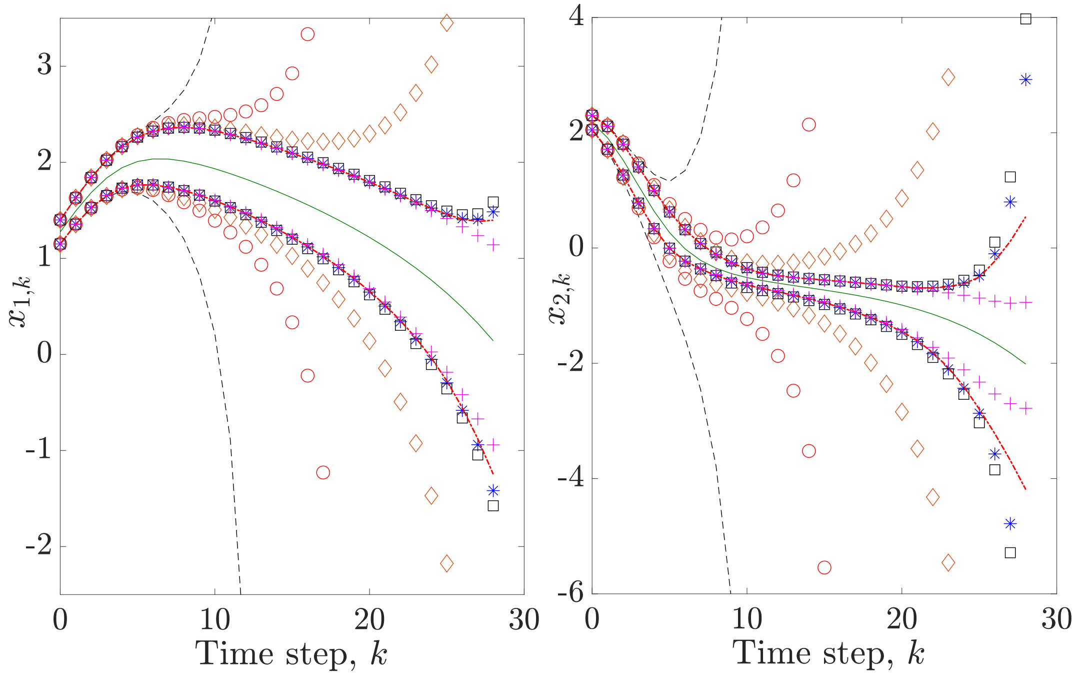


Figure 8.1: Upper and Lower Bounds on x_1 and x_2 in System (8.32) (the Van Der Pol System), Applying T_N(−−), T_C (○), T_M (◇), T_L (□), T_R ($*$), the Best of T_N–T_R (·-) and T_O (+), as Well as the True Trajectory (–)

8.6.2 Example 3 in [130]

Now consider the following discrete-time dynamical system [5, 130], with bounded noise:

$$x_{k+1} = \begin{bmatrix} 0 & -0.5 \\ 1 & 1+0.3v_k \end{bmatrix} x_k + 0.02 \begin{bmatrix} -6 \\ 1 \end{bmatrix} w_k, \tag{8.33}$$

where $w_k, v_k \in [-0.001, 0.001]$ and $x_0 \in [-0.55, -0.445] \times [0.145, 0.248]$. The approximated reachable sets are depicted in Fig. 8.2. Here, T_{S_1} is applicable and valid and returns the exact same bounds as T_N (natural inclusion) does, but T_{S_2} is not applicable due to the lack of monotonicity condition (cf. [130, (23)]). The tight T_O is again computable and as expected returns the tightest possible intervals. Again, our proposed approach, T_R, dominates all the other applicable ones, except for the tight T_O. To further improve our results and inspired by what is proposed in the literature, e.g. in [130], we considered a manufactured *redundant* state z_k as:

$$z_k = x_{1,k} + 6x_{2,k}, \tag{8.34}$$

which implies that:

$$z_{k+1} = z_k + 5x_{1,k} + (1.8v_k - 0.5)x_{2,k}. \tag{8.35}$$

Augmenting (8.35) with the original system (8.33), we considered the approximation of reachable sets for the augmented system, constrained to (8.34). We used our proposed T_R for the propagation step, as well as Algorithm 8 for set-inversion (refinement/update). The results in Fig. 8.2 demonstrate an improvement in applying T_R when considering manufactured redundant variable along with an update step through Algorithm 8, compared to using T_R on the original system without considering any redundancies (cf. Figure 8.2). However, developing a *principled* approach of defining redundant variables is a subject of our future work.

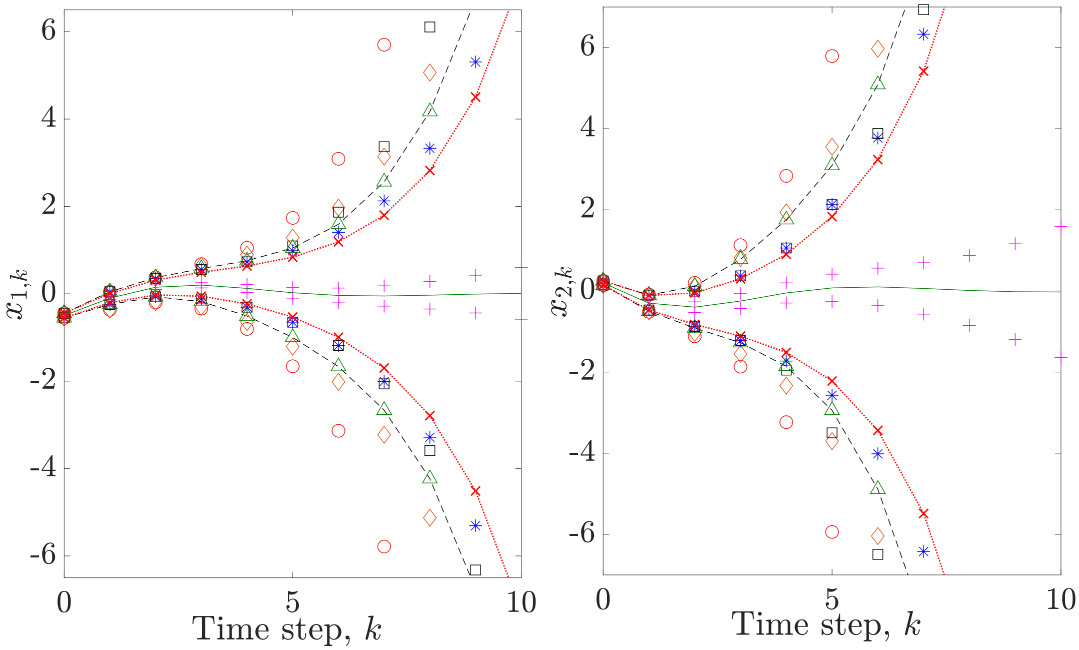


Figure 8.2: Upper and Lower Bounds on x_1 and x_2 in System (8.33), Applying $T_N(--)$, T_{S_1} ($\triangle$), T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), T_R with Manufactured Redundancies ($\times\cdot$) and T_O ($+$), as Well as the True Trajectory ($-$)

Next, consider the following system, which is the system in (8.33) as well as an extra set of information, in the form of observation equation. In other words:

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 0 & -0.5 \\ 1 & 1+0.3v_k \end{bmatrix} x_k + 0.02 \begin{bmatrix} -6 \\ 1 \end{bmatrix} w_k, \\ y_k &= 1.6x_{1,k} + 0.3x_{2,k}, \end{aligned} \tag{8.36}$$

where y_k is a *known* measurement/observation signal at time step k. Fig. 8.3 shows the approximated upper and lower bounds for the states of the system (8.36), using the same inclusion functions that we applied to the system (8.33), as well as applying Algorithm 8 for the set-inversion/refinement procedure. As expected and can be seen by comparing Figures 8.2 and 8.3, using the additional measurement information tightens the resultant intervals from all the inclusion functions.

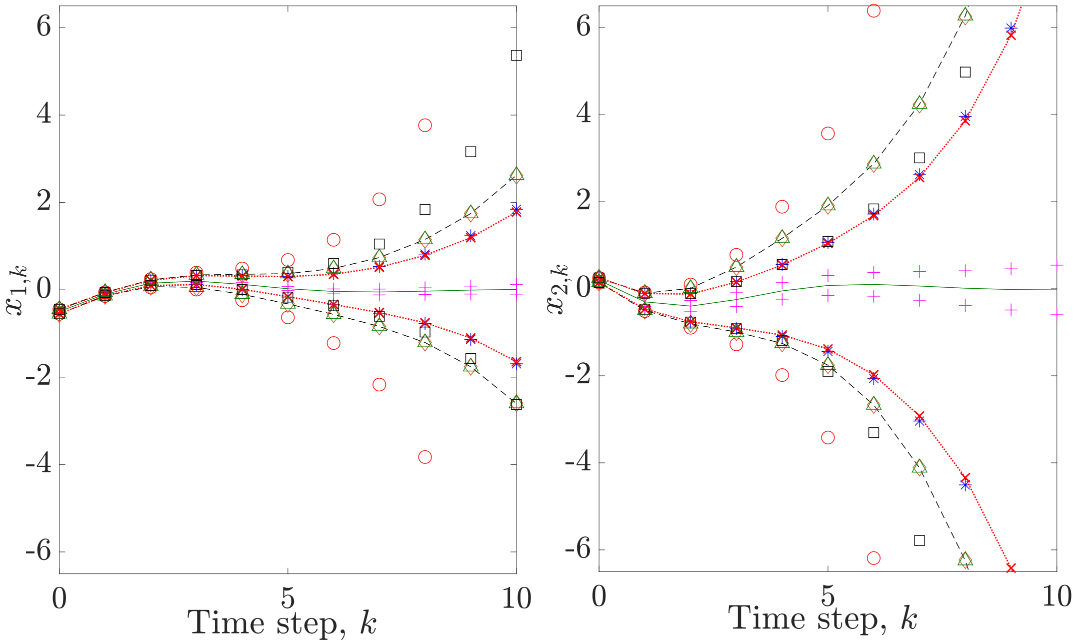


Figure 8.3: Upper and Lower Bounds on x_1 and x_2 in System (8.36), (i.e., (8.33) Plus Observations), Applying $T_N(--)$, T_{S_1} (△), T_C (○), T_M (◇), T_L (□), T_R ($*$), T_R with Manufactured Redundancies (×·) and T_O (+), as Well as the True Trajectory (–).

8.6.3 Example 2.11 in [57]

Next, we consider the following discrete-time dynamical system, e.g., in [57, Example 2.11]:

$$\begin{aligned} x_{1,k+1} &= x_{1,k}^2 + x_{1,k}e^{x_{2,k}} - x_{2,k}^2, \\ x_{2,k+1} &= x_{1,k}^2 - x_{1,k}e^{x_{2,k}} + x_{2,k}^2, \end{aligned} \tag{8.37}$$

where $x_{0,k} \in [0.12, 0.121] \times [0.182, 0.185]$. Here, non of the T_{S_1} and T_{S_2} approaches are applicable, due to lack of the required monotonicity conditions. Further, the tight T_O is not computable, since the corresponding nonlinear equations to find the critical points are not exactly solvable. On the other hand, our proposed remainder-form inclusion function, T_R, is valid and computable. Fig. 8.4 shows the results, where as can be observed, T_R dominates all the other applicable and computable approaches. The corresponding intervals to the best of all the approaches via intersection, is also computed.

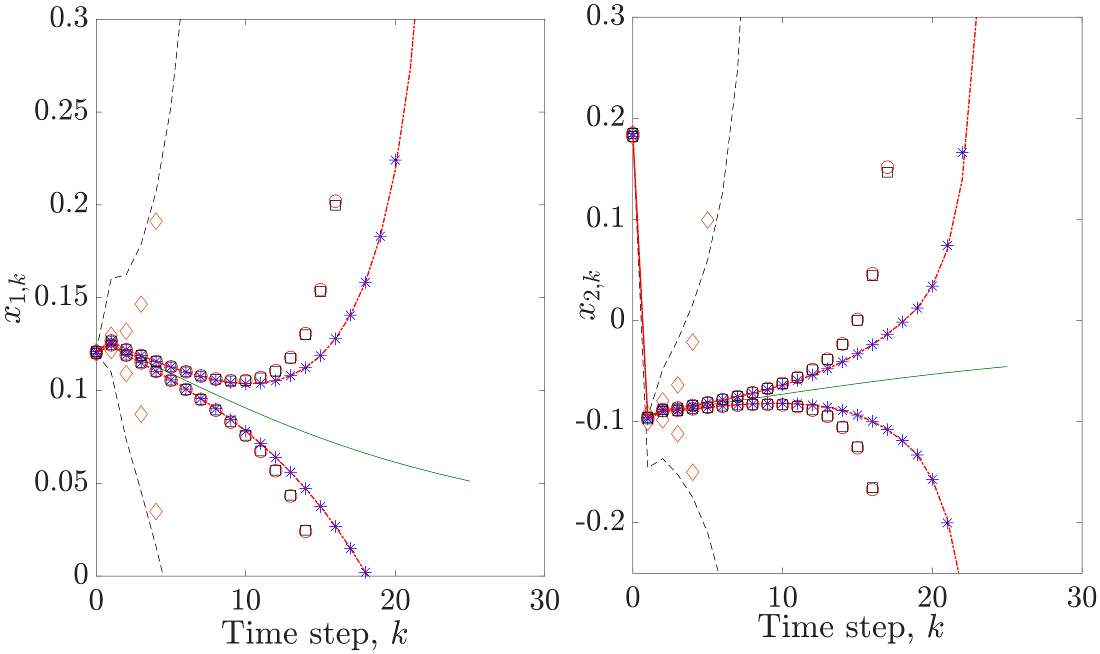


Figure 8.4: Upper and Lower Bounds on x_1 and x_2 in System (8.37), Applying $T_N(--)$, T_C (○), T_M (◇), T_L (□), T_R ($*$) and the Best of T_N–T_R (·-), as Well as the True Trajectory (–)

8.6.4 *Continuous-Time System in [2]*

As the next example, we consider the following continuous-time dynamical system from [2]:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} w_1 x_2^2 - x_2 + w_2 \\ x_3 + 2 \\ x_1 - x_2 - w_1^3 \end{bmatrix}, \tag{8.38}$$

with $\mathcal{X}_0 = [-\frac{1}{2}, \frac{1}{2}]^3$ and $\mathcal{W} = [-\frac{1}{4}, 0] \times [0, \frac{1}{4}]$. Figure 8.5, depicts the approximations for the reachable sets applying T_N, T_C, T_M, T_L, T_R, best of T_N–T_R and T_O. As expected, T_O (implemented using the corresponding embedding functions given in [2, Section VI]), returns the tightest approximations. However, among the others, T_R has the the best performance and it can be slightly improved, using the best of T_N–T_R.

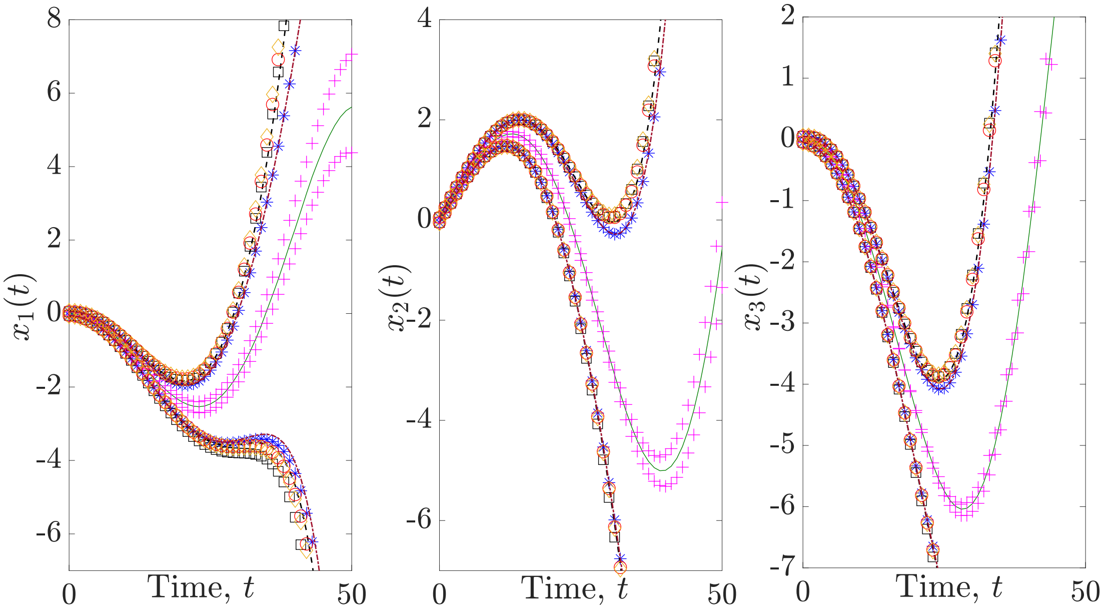


Figure 8.5: Upper and Lower Bounds for x_1, x_2 and x_3 in System (8.38), applying $T_N(--)$, T_C (○), T_M (◇), T_L (□), T_R ($*$), the Best of T_N–T_R ($\cdot$-) and T_O ($+$), as Well as the True Trajectory (–)

8.6.5 The Unicycle System

We are interested in computing the reachable sets for the well-known continuos-time system, the unicycle-like mobile robot, e.g., in [25, 58], with two driving wheels, mounted on the left and right sides of the robot, with their common axis passing through the center of the robot. The dynamics of such a system can be described as

[58]:

$$
\begin{aligned}
\dot{s}_x(t) &= \phi_\omega(t)\cos\theta(t) + w_x(t), \\
\dot{s}_y(t) &= \phi_\omega(t)\sin\theta(t) + w_y(t), \\
\dot{\theta}(t) &= \phi_\theta(t) + w_\theta(t),
\end{aligned}
\tag{8.39}
$$

where $s_x(t)$ and $s_y(t)$ are the coordinates of the main axis mid-point between the two driving wheels, $\theta(t)$ is the angle between the robot forward axis and the X-direction, $\phi_\omega(t)$ and $\phi_\theta(t)$ are the displacement and angular velocities of the robot, respectively and $w(t) = [w_x(t)\ w_y(t)\ w_\theta(t)]^\top$ is the process noise vector. Setting $\phi_\omega(t) = 0.3$, $\phi_\theta(t) = 0.15$, $x(t) \triangleq [s_x(t)\ s_y(t)\ \theta(t)]^\top$, $x(0) = [0.1\ 0.2\ 1]^\top$, $w_x(t) = 0.2(0.5\rho_{x_1}(t) - 0.3)$, $w_y(t) = 0.2(0.3\rho_{x_2}(t) - 0.2)$ and $w_\theta(t) = 0.2(0.6\rho_{x_3}(t) - 0.4)$, with $\rho_{x_l}(t) \in [0,1] (l = 1,2,3)$, Figure 8.6 shows the over-approximations of the reachable sets for system (8.39), using different methods. As can be observed, T_O- which is computable for this example- returns the tightest intervals. Moreover, in some intervals natural inclusions and their modifications, return tighter bounds than T_R. However, by taking intersection and computing the best of T_N–T_R, we make an improvement we respect to all of them.

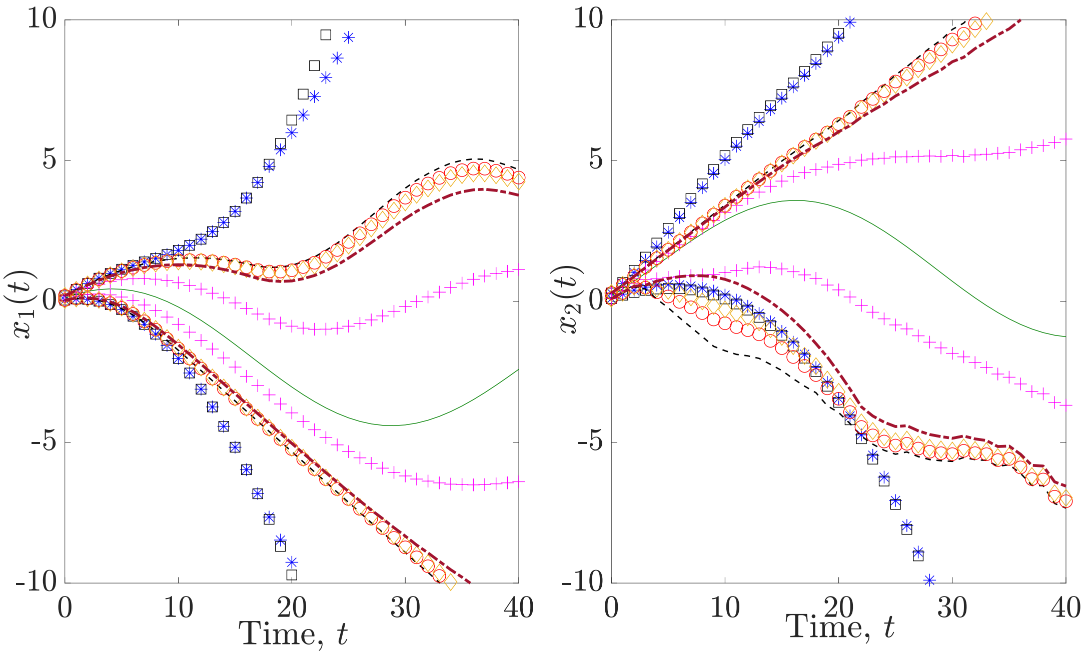


Figure 8.6: Upper and Lower Bounds on x_1 and x_2 in System (8.39) (the Unicycle System Without Observation), Applying $T_N(--)$, T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$), the Best of T_N–T_R ($\cdot$-) and T_O ($+$), as Well as the True Trajectory ($-$)

Then, to better illustrate the effectiveness of our proposed set-inversion algorithm for continuous-time systems, we consider observations/measurements similar to [25], as follows. In X–Y plane, it is considered that two known points, denoted as $(s_{x_i}, s_{y_i})(i = 1, 2)$, are chosen as the markers. Then, the distance from the robot's planner Cartesian coordinates $(s_x(t), s_y(t))$ to each marker (s_{x_i}, s_{y_i}) can be expressed as $d_i(t) = \sqrt{(s_{x_i} - s_x(t))^2 + (s_{y_i} - s_y(t))^2}$. Furthermore, the azimuth $\phi_i(t)$ at time t can be related to the current system state variables $s_x(t), s_y(t)$ and $\theta(t)$ as $\phi_i(t) = \theta(t) - \arctan(\frac{s_{y_i} - s_y(t)}{s_{x_i} - s_x(t)})$. Treating both the distance $d_i(t)$ and $\phi_i(t)$ as the measurements, as well as considering the unpredicted measurement disturbances $v(t)$, the nonlinear observation/measurement equation can be written as:

$$y(t) = [d_1(t)\ \phi_1(t)\ d_2(t)\ \phi_2(t)]^\top + v(t), \tag{8.40}$$

with $y(t)$ being sampled and measured at every $T = 1$ second, $v_1(t) = 0.02\rho_{y_1}(t) - 0.01$, $v_2(t) = 0.03\rho_{y_2}(t) - 0.01$, $v_3(t) = 0.03\rho_{y_3}(t) - 0.02$, $v_4(t) = 0.05\rho_{y_4}(t) - 0.03$ and $\rho_{y_k}(t) \in [0, 1] (k = 1, 2, 3, 4)$. Now, applying all the methods along with the set-inversion approach in Algorithm **??** to the constrained system (8.39)–(8.40), one can observe that taking the advantage of observations, the reachable set approximations have been significantly improved in the Figure 8.7 (with observation) compared to the Figure 8.6 (without observation).

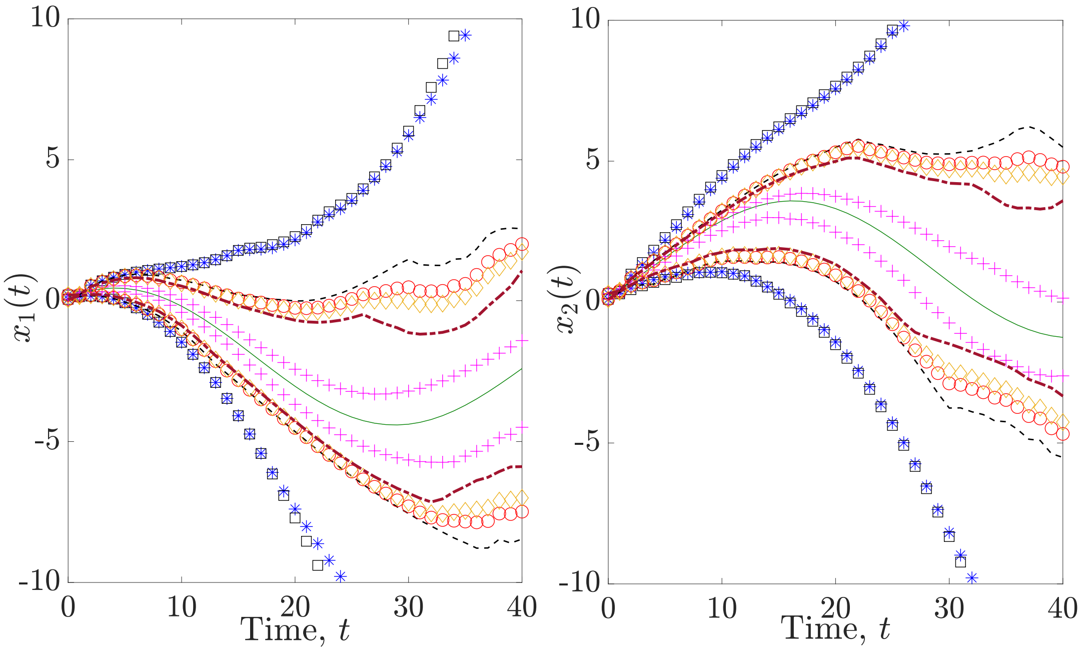


Figure 8.7: Upper and Lower Bounds on x_1 and x_2 in System (8.39)–(8.40) (the Unicycle System with Observation), Applying $T_N(--)$, T_C (○), T_M (◇), T_L (□), T_R ($*$), the Best of T_N–T_R (·-) and T_O (+), as Well as the True Trajectory (–)

8.6.6 Generic Transport Longitudinal Model

Finally, we consider NASA's Generic Transport Model (GTM) [112], that is a remote-controlled commercial aircraft [85], with the following main parameters: wing area $S = 5.902$ ft^2, mean aerodynamic chord $\bar{c} = 0.9153$ ft, mass $m = 1.542$ slugs, pitch axis moment of inertia $I_{yy} = 4.254$ slugs/ft^2, air density $\rho = 0.002375$ slugs/ft^3 and gravitational acceleration $g = 32.17$ ft/s^2. The longitudinal dynamics of the GTM can be described as the following continuous-time dynamical system:

$$\begin{aligned}
\dot{V} &= \tfrac{-D - mg\sin(\theta-\alpha) + T_x\cos\alpha + T_z\sin\alpha}{m}, \\
\dot{\alpha} &= q + \tfrac{-L + mg\cos(\theta-\alpha) - T_x\sin\alpha + T_z\cos\alpha}{mV}, \\
\dot{q} &= \tfrac{M + T_m}{I_{yy}}, \\
\dot{\theta} &= q,
\end{aligned} \tag{8.41}$$

where V, α, q and θ are air speed (ft/s), angle of attack (rad), pitch rate (rad/s) and pitch angle (rad), respectively. Moreover, T_x (lbs), T_z (lbs), T_m (lbs–ft), D (lbs), L (lbs) and M (lbs) denote the projection of the total engine thrust along the body x-axis, the projection of the total engine thrust along the body z-axis, the pitching moment due to both engines, the drag force, the lift force and the aerodynamic pitching moment, respectively, with their nominal values given in [111]. Defining $x \triangleq [V\ \alpha\ q\ \theta]^\top$ with $\mathcal{X}_0 = [147, 158] \times [0.04, 0.05] \times [0.1, 0.2] \times [0.04, 0.05]$, Figure 8.8 depicts the reachable set approximations for x_1 and x_2 in the system (8.41). For this system T_O is not computable, since the critical points of the vector fields cannot be obtained precisely. As can be observed from Figure 8.8, the T_R results in tighter approximations than the T_N, T_C and T_M, and the best of T_N–T_R, still shows an improvement compared to all of them.

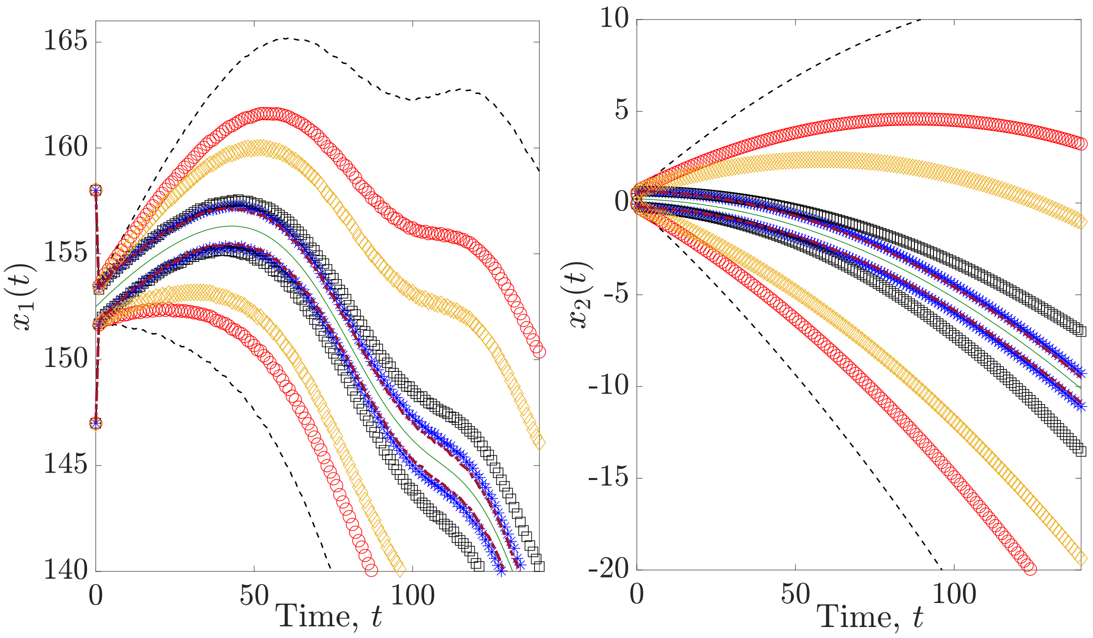


Figure 8.8: Upper and Lower Bounds on x_1 and x_2 in System (8.41) (the GTM System Without Observation), Applying $T_N(--)$, T_C (○), T_M (◇), T_L (□), T_R ($*$) and the Best of T_N–T_R (·-), as Well as the True Trajectory (–)

Next, we consider an additional set of measurements, in the form of a linear observation equation as:

$$y(t) = x_1(t) + x_2(t) - x_3(t), \tag{8.42}$$

with $y(t)$ being sampled and measured at every $T = 1$ second. Then applying T_N–T_R along with the set-inversion (update) approach in Algorithm 8 to the constrained system (8.41)–(8.42), we observe considerable tighter approximations for all approaches in Figure 8.9 (with observation) compared to the approximations of reachable sets in Figure 8.8 (without observation) .

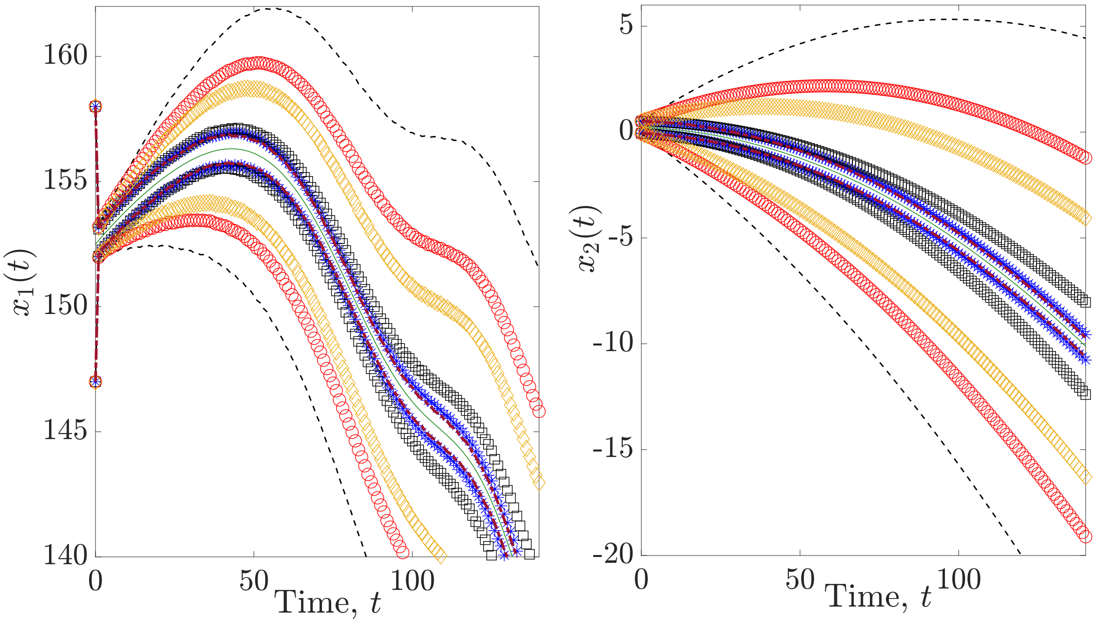


Figure 8.9: Upper and Lower Bounds on x_1 and x_2 in System (8.41)–(8.42) (the GTM System with Observation), Applying $T_N(--)$, T_C ($\circ$), T_M ($\diamond$), T_L ($\square$), T_R ($*$) and the Best of T_N–T_R ($\cdot$-), as Well as the True Trajectory (–)

8.7 Conclusion

A tractable family of remainder-from decomposition functions was proposed, that their existence is proven to be sufficient conditions for mixed-monotonicity of a broad-range of not necessarily smooth, constrained and unconstrained, continuous and discrete-time bounded-error dynamical systems. Specifying the tightest decomposition function belong to the above family, attainable lower and upper bounds for the error of over-approximating the true range of a mapping, applying the proposed remainder-form decomposition functions were provided. Furthermore a set-inversion algorithm was developed that along with the proposed decomposition functions can be effectively applied to several applications, such as reachable set over-approximation for bounded-error, constrained, continuous and/or discrete-time systems, as well as guaranteed state estimation.

Chapter 9

GUARANTEED STATE ESTIMATION VIA INDIRECT POLYTOPIC SET COMPUTATION FOR NONLINEAR DISCRETE-TIME SYSTEMS

This chapter [a] proposes novel set-theoretic approaches for recursive state estimation in bounded-error discrete-time nonlinear systems subject to nonlinear observations/constraints. By transforming the polytopes that are characterized as zonotope bundles (ZB) and/or constrained zonotopes (CZ), from the state space to the space of the generators of ZB/CZ, we leverage a recent result on remainder-form mixed-monotone decomposition functions to compute the propagated set, i.e., a ZB/CZ that is guaranteed to enclose the set of the state trajectories of the considered system. Further, by applying the remainder-form decomposition functions to the nonlinear observation function, we derive the updated set, i.e., an enclosing ZB/CZ of the intersection of the propagated set and the set of states that are compatible/consistent with the observations/constraints. In addition, we show that the mean value extension result in [99] for computing propagated sets can also be extended to compute the updated set when the observation function is nonlinear.

9.1 Preliminaries

In this section, we briefly introduce some of the main concepts that we use throughout the paper, as well as some important existing results that will be used for deriving our main results and for comparison.

Definition 9.1.1 (Intervals, H-Polytopes, Constrained Zonotopes (CZ) and Zono-

---

[a] The content of this chapter is documented as a submitted and under review paper in [114].

tope Bundles (ZB)). *A set $\mathcal{Z} \subset \mathbb{R}^n$ is (i) an interval, (ii) a polytope in hyperplane representation (H-polytope), (iii) a polytope in constrained zonotope representation (CZ), or (iv) a polytope in zonotope bundle representation (ZB), if*

- *(i)* $\exists \underline{z}, \overline{z} \in \mathbb{R}^n$ *such that* $\mathcal{Z} = [\underline{z}, \overline{z}] \triangleq \{z \in \mathbb{R}^n \mid \underline{z} \le z \le \overline{z}\}$. *An interval matrix can be defined similarly, in an element-wise manner;*
- *(ii)* $\exists A_p \in \mathbb{R}^{n_p \times n}, b_p \in \mathbb{R}^{n_p}$ *such that* $\mathcal{Z} = \{A_p, b_p\}_P \triangleq \{z \in \mathbb{R}^n \mid A_p z \le b_p\}$;
- *(iii)* $\exists \tilde{G} \in \mathbb{R}^{n \times n_g}, \tilde{c} \in \mathbb{R}^n, \tilde{A} \in \mathbb{R}^{n_c \times n_g}, \tilde{b} \in \mathbb{R}^{n_c}$ *such that* $\mathcal{Z} = \{\tilde{G}, \tilde{c}, \tilde{A}, \tilde{b}\}_{CZ} \triangleq \{\tilde{G}\xi + \tilde{c} \mid \xi \in \mathbb{B}_\infty^{n_g}, \tilde{A}\xi = \tilde{b}\}$. n_g *and* n_c *are called the number of generators and constraints of the CZ, respectively;*
- *(iv)* $\mathcal{Z}$ *can be represented as an intersection of* $S \in \mathbb{N}$ *zonotopes, i.e.,* $\exists \{G_s \in \mathbb{R}^{n \times \hat{n}_s}, c_s \in \mathbb{R}^n\}_{s=1}^S$ *such that* $\mathcal{Z} = \bigcap_{s=1}^{S} \{G_s, c_s\}_Z \triangleq \bigcap_{s=1}^{S} \{G_s \zeta + c_s \mid \zeta \in \mathbb{B}^{\hat{n}_g}\}$, *with* $\hat{n}_s, s = 1, \ldots, S$, *being called the number of generators for each zonotope.*

It is worth mentioning that a polytope $\mathcal{Z}$ can be equivalently given in the H-polytope, CZ or ZB representations and can be exactly transformed among these representations using off-the-shelf tools, e.g., CORA 2020 [8]. This is represented throughout this paper as:

$$\mathcal{Z} = \{A_p, b_p\}_P \equiv \{\tilde{G}, \tilde{c}, \tilde{A}, \tilde{b}\}_{CZ} \equiv \bigcap_{s=1}^{S} \{G_s, c_s\}_Z.$$

Proposition 9.1.2. *Consider an interval vector* $\mathbb{IZ} \triangleq [\underline{z}, \overline{z}] \subset \mathbb{IR}^n$ *and an interval matrix* $\mathbb{J} \in \mathbb{IR}^{n \times m}$. *Then,* $\mathbb{IZ}$ *and* $\mathbb{J}$ *can be equivalently represented as*

$$\mathbb{IZ} \triangleq [\underline{z}, \overline{z}] \equiv \mathrm{mid}(\mathbb{IZ}) \oplus \tfrac{1}{2}\mathrm{diag}(\mathrm{diam}(\mathbb{IZ}))\mathbb{B}_\infty^n, \tag{9.1}$$

$$\mathbb{J} \triangleq [\underline{J}, \overline{J}] \equiv \mathrm{mid}(\mathbb{J}) \oplus \mathbb{J}_\Delta, \tag{9.2}$$

where for $q \in \{\mathbb{IZ}, \mathbb{J}\}$, $\mathrm{mid}(q) \triangleq \frac{1}{2}(\overline{z} + \underline{z})$, $\mathrm{diam}(q) \triangleq (\overline{z} - \underline{z})$, *and* $\mathbb{J}_\Delta \in \mathbb{IR}^{n \times m}$ *is an interval matrix that is defined as* $[\mathbb{J}_\Delta]_{ij} \triangleq \frac{1}{2}[-\mathrm{diam}(\mathbb{J})_{ij}\ \mathrm{diam}(\mathbb{J})_{ij}], \forall i \in \mathbb{N}_n, \forall j \in \mathbb{N}_m$.

Proposition 9.1.3. *[99, Theorem 1] Let $\mathcal{X} = \{G, c, A, b\}_{CZ} \subset \mathbb{R}^m$ be a constrained zonotope with n_g generators and n_c constraints, and $\mathbb{J} \in \mathbb{IR}^{n\times m}$ be an interval matrix. Consider the set $S = \mathbb{J}\mathcal{X} \triangleq \{Jx | J \in \mathbb{J}, x \in \mathcal{X}\} \subset \mathbb{R}^n$. Let $\overline{\mathcal{X}} = \{\overline{G}, \overline{c}\}_Z$ be a zonotope satisfying $\mathcal{X} \subseteq \overline{\mathcal{X}}$ and $\overline{c} \in \mathbb{R}^{\overline{n}_g}$. Let $\mathbf{m} \in \mathbb{R}^n$ be an interval vector such that $\mathbf{m} \supset (\mathbb{J} - \mathrm{mid}(\mathbb{J}))\overline{c}$ and $\mathrm{mid}(\mathbf{m}) = \mathbf{0}_n$. Let $P \in \mathbb{R}^{n\times n}$ be a diagonal matrix defined as follows. $\forall i = 1, \ldots, n$:*

$$P_{ii} = \frac{1}{2}\mathrm{diam}(\mathbf{m}_i) + \tfrac{1}{2}\textstyle\sum_{j=1}^{\overline{n}_g}\sum_{k=1}^{m}\mathrm{diam}(\mathbb{J}_{ik})|\overline{G}_{kj}|. \tag{9.3}$$

Then, $\mathcal{S} \subseteq \mathrm{mid}(\mathbb{J})\mathcal{X} \oplus P\mathbb{B}_\infty^n$

$$= \{\begin{bmatrix}\mathrm{mid}(\mathbb{J})G & P\end{bmatrix}, \mathrm{mid}(\mathbb{J})c, \begin{bmatrix}A & 0_{n_g\times n}\end{bmatrix}, b\}_{CZ}. \tag{9.4}$$

Proposition 9.1.4 (RRSR Propagation Approach)**.** *[99, Theorem 2] Let $f : \mathbb{R}^n \times \mathbb{R}^{n_w} \to \mathbb{R}^n$ be continuously differentiable and $\nabla_x f$ denote the gradient of f with respect to its first argument. Let $\mathcal{X} = \{G_x, c_x, A_x, b_x\}_{CZ} \subset \mathbb{R}^n$ and $\mathcal{W} \subset \mathbb{R}^{n_w}$ be constrained zonotopes. Choose any $h \in \mathcal{X}$. If $\mathcal{Z}$ is a constrained zonotope such that $f(h, \mathcal{W}) \subseteq \mathcal{Z}$ and $\mathbb{J} \in \mathbb{IR}^{n\times n}$ is an interval matrix satisfying $\nabla_x^\top f(\mathcal{X}, \mathcal{W}) \subseteq \mathbb{J}$, then*

$$f(\mathcal{X}, \mathcal{W}) \subseteq \mathcal{Z} \oplus \mathrm{mid}(\mathbb{J})(\mathcal{X} \ominus \{h\}) \oplus \tilde{P}\mathbb{B}_\infty^n, \tag{9.5}$$

where $\tilde{P}$ can be computed using (9.3) with $\mathbb{J}$ and an enclosing zonotope $\overline{\mathcal{X}} = \{\overline{G}, \overline{c}\}_Z$ of $\mathcal{X} \ominus \{h\} \subseteq \overline{\mathcal{X}}$.

Definition 9.1.5 (Mixed-Monotone (One-Sided) Decomposition Functions For Discrete-Time Systems)**.** *[63, Definitions 3,4] A mapping $f_d : \mathcal{Z} \times \mathcal{Z} \subset \mathbb{R}^{2n} \to \mathbb{R}^m$ is a discrete-time mixed-monotone decomposition function with respect to $f : \mathcal{Z} \subset \mathbb{R}^n \to \mathbb{R}^m$, over the set $\mathcal{Z}$, if it satisfies the following: (i) $f_d(x, x) = f(x)$, (ii) $x \geq x' \Rightarrow f_d(x, y) \geq f_d(x', y)$, and (iii) $y \geq y' \Rightarrow f_d(x, y) \leq f_d(x, y'), \forall x, y, x', y' \in \mathcal{Z}$. Further, if there exists two mixed-monotone mappings $\overline{f}_d, \underline{f}_d : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}^m$, such* *that for any* $\underline{z}, z, \overline{z} \in \mathcal{Z}$*, the following holds:* $\underline{z} \le z \le \overline{z} \Rightarrow \underline{f}_d(\underline{z},\overline{z}) \le f(z) \le \overline{f}_d(\overline{z},\underline{z})$*, then* $\overline{f}_d$ *and* $\underline{f}_d$ *are called upper and lower decomposition functions for* f *over* $\mathcal{Z}$*, respectively.*

It is trivial to see that $\forall x \in [\underline{x},\overline{x}], \underline{f}_d(\underline{x},\overline{x}) \le f(x) \le \overline{f}_d(\overline{x},\underline{x})$, where $\underline{f}_d, \overline{f}_d$ are lower and upper decomposition functions of f.

Proposition 9.1.6 (Tight and Tractable Remainder-Form Upper and Lower Decomposition Decomposition Functions)**.** *[63, Theorems 1,2,3] Consider a locally Lipschitz vector field* $f_i : \mathbb{IZ} \triangleq [\underline{z},\overline{z}] \subseteq \mathbb{IR}^{n_z} \to \mathbb{R}$*. Let* $\mathbb{N}_{n_z} \triangleq \{1,\dots,n_z\}$ *and* $\overline{J}^{\tilde{f}}_i, \underline{J}^{\tilde{f}}_i \in \mathbb{R}^{n_z}$ *denote the upper and lower bounds for the Jacobian matrix (vector) of* f_i *over* $\mathbb{IZ}$*. Suppose that Assumption 9.2.2 in Section 9.2 holds. Then,* $f_i(\cdot)$ *admits a family of mixed-monotone remainder-form decomposition functions denoted as* $\{f_{d,i}(z,\hat{z};m,h(\cdot))\}_{\mathbf{m}\in\mathbf{M}_i,h(\cdot)\in\mathcal{H}_{\mathbf{M}^c_i}}$*, that is parametrized by a set of supporting vectors* $\mathbf{m}\in\mathbf{M}^c_i$

$$\mathbf{m}\in\mathbf{M}^c_i \triangleq \{\mathbf{m}\in\mathbb{R}^{n_z} | \mathbf{m}_j = \min(\underline{J}^f_{ij},0) \vee \mathbf{m}_j = \max(\overline{J}^f_{i,j},0), \forall j \in \mathbb{N}_{n_z}\}, \tag{9.6}$$

and a locally Lipschitz remainder function $h(\cdot)\in\mathcal{H}_{M^c_i}$*, where*

$$f_{d,i}(z,\hat{z};\mathbf{m},h(\cdot)) = h(\zeta_{\mathbf{m}}(\hat{z},z)) + f_i(\zeta_{\mathbf{m}}(z,\hat{z}))) - h_i(\zeta_{\mathbf{m}}(z,\hat{z})), \tag{9.7}$$

$\zeta_{\mathbf{m}}(z,\hat{z}) = [\zeta_{\mathbf{m},1}(z,\hat{z}),\dots,\zeta_{\mathbf{m},n_z}(z,\hat{z})]^\top$, $\forall j \in \mathbb{N}_{n_z}$*:*

$$\zeta_{\mathbf{m},j}(z,\hat{z}) = \begin{cases} \hat{z}_j, & \text{if } \mathbf{m}_j = \max(\overline{J}^f_{i,j},0) \\ z_j, & \text{if } \mathbf{m}_j = \min(\underline{J}^f_{i,j},0) \end{cases}, \tag{9.8}$$

and $\mathcal{H}_{\mathbb{M}_i} \triangleq \{h : \mathbb{IZ} \to \mathbb{R} | [\underline{J}^h(z), \overline{J}^h_C(z)] \subseteq \mathbb{M}_i, \forall z \in \mathbb{IZ}\}$*. Moreover, the search for the tightest mixed-monotone upper and lower remainder-form decomposition functions in the form of* (8.10) *can be equivalently restricted to the set of "linear remainders", parametrized by* $\mathbf{m}\in\mathbf{M}^c_i$*, i.e., linear remainders* $\{h(\cdot)\}_{\mathbf{m}\in\mathbf{M}^c_i} = \{\langle\mathbf{m}^i,\cdot\rangle\}_{\mathbf{m}\in\mathbf{M}^c_i}$*.*

Corollary 9.1.7. *Consider a locally Lipschitz mapping* $\tilde{f}(\cdot) : \mathbb{I}\Xi \triangleq [\underline{\xi}, \overline{\xi}] \subseteq \mathbb{IR}^{n_\xi} \to \mathbb{R}^{n_x}$ *that satisfies the assumptions in Proposition 9.1.6. Let us define:* $\mathbb{N}_{n_x} \triangleq \{1, \ldots, n_x\}$ *and*

$$\mathbf{H}_{\tilde{f}} \triangleq \{H \in \mathbb{R}^{n_x \times n_\xi} | H_{i,:}^\top \in \mathbf{M}_i^c, \forall i \in \mathbb{N}_{n_x}\}, \tag{9.9}$$

where $\mathbf{M}_i^c$ *is defined in* (8.12)*. Then,* $\forall \xi \in \mathbb{I}\Xi, \forall H \in \mathbf{H}^{\tilde{f}}, \tilde{g}^H(\xi) \triangleq \tilde{f}(\xi) - H\xi$ *is proven to be a Jacobian sign-stable (JSS) function, i.e.,* $\forall i \in \mathbb{N}_{n_x}, \forall j \in \mathbb{N}_{n_z}, J_{ij}^H(\xi) \triangleq \frac{\partial \tilde{g}_i^H}{\partial \xi_j}(\xi) \geq 0, \forall \xi \in \mathbb{I}\Xi$ *or* $J_{ij}^H(\xi) \triangleq \frac{\partial \tilde{g}_i^H}{\partial \xi_j}(\xi) \leq 0, \forall \xi \in \mathbb{I}\Xi.$*. Consequently,* $\tilde{g}^H(\cdot)$ *can be tightly bounded in each dimension* $i \in \mathbb{N}_{n_x}$ *by remainder-form decomposition functions* $\tilde{g}_{d,i}(\cdot, \cdot; H_{i,:}^\top, \langle H_{i,:}^\top, \cdot \rangle)$*, constructed using* (9.7)*–*(9.8)*, as follows:*

$$\tilde{g}_{d,i}(\underline{\xi}, \overline{\xi}; H_{i,:}^\top, \langle H_{i,:}^\top, \cdot \rangle) \leq \tilde{g}_i(\xi) \leq \tilde{g}_{d,i}(\overline{\xi}, \underline{\xi}; H_{i,:}^\top, \langle H_{i,:}^\top, \cdot \rangle),$$

where, by [63, Lemma 3] and defining $m \triangleq H_{i,:}^\top$*, we obtain* $\tilde{g}_{d,i}(\overline{\xi}, \underline{\xi}; m, \langle m, \cdot \rangle) = \tilde{f}_i(\zeta_m^+) + m^\top(\zeta_m^- - \zeta_m^+)$*,* $\tilde{g}_{d,i}(\underline{\xi}, \overline{\xi}; m^\top, \langle m, \cdot \rangle) = \tilde{f}_i(\zeta_m^-) + m^\top(\zeta_m^+ - \zeta_m^-)$*,* $\zeta_m^+ \triangleq \zeta_m(\overline{\xi}, \underline{\xi})$*,* $\zeta_m^- \triangleq \zeta_m(\underline{\xi}, \overline{\xi})$*, with* $\zeta_m(\cdot, \cdot)$ *given in* (9.8)*.*

9.2 Problem Formulation

System Assumptions. Consider the following bounded-error nonlinear constrained discrete-time system:

$$\begin{aligned} x_{k+1} \quad &= \hat{f}(x_k, w_k, u_k) = f(z_k), \\ \hat{\mu}(x_k, u_k) \quad &= \mu(x_k) \in \mathcal{Y}_k, \ x_0 \in \hat{\mathcal{X}}_0, w_k \in \mathcal{W}_k, \end{aligned} \tag{9.10}$$

where $z_k \triangleq [x_k^\top w_k^\top]^\top$, $x_k \in \mathbb{R}^{n_x}$ is the state vector, $w_k \in \mathcal{W}_k \subset \mathbb{R}^{n_w}$ is the augmentation of all exogenous uncertain inputs, e.g., bounded process disturbance/noise and system uncertainties such as uncertain parameters and $u_k \in \mathcal{U}_k \subseteq \mathbb{R}^{n_u}$ is the *known* input signal. Furthermore, $f : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ (with $n_z \triangleq n_x + n_w$) and $\mu : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\mu}$ are nonlinear state vector field and observation/constraint mapping, respectively, which are well-defined, given $\hat{f}(\cdot,\cdot)$ and $\hat{\mu}(\cdot,\cdot)$, as well as the fact that u_k is known. Note that the mapping $\mu(\cdot)$ along with the set $\mathcal{Y}_k$ characterize all existing and known or even manufactured/redundant constraints over the states, observations and measurement noise signals or uncertain parameters at time step k..

The unknown initial state x_0 is assumed to be in a given set $\hat{\mathcal{X}}_0$ and moreover, we assume the following:

Assumption 9.2.1. *The initial state set* $\hat{\mathcal{X}}_0$*, as well as* $\mathcal{W}_k, \mathcal{U}_k, \forall k \geq 0$ *are prior known H-polytopes, or equivalently constrained zonotopes or zonotope bundles (cf. Definition 9.1.1).*

Assumption 9.2.2. *The nonlinear vector fields* $f(\cdot)$ *and* $\mu(\cdot)$ *are locally Lipschitz on their domains. Consequently, they are differentiable and have bounded Jacobian matrix elements, almost everywhere. We further assume that given any* $\mathcal{Z} \subset \mathbb{R}^{n_z}$ *and* $\mathcal{X} \subset \mathbb{R}^{n_x}$*, some upper and lower bounds for all elements of Jacobian matrices for* $f(\cdot)$ *and* $\mu(\cdot)$ *over* $\mathcal{Z}$ *and* $\mathcal{X}$ *are available or can be computed. In other words,* $\exists \underline{J}^f, \overline{J}^f \in \mathbb{R}^{n_x \times n_z}, \underline{J}^\mu, \overline{J}^\mu \in \mathbb{R}^{n_\mu \times n_x}$*, such that:* $\underline{J}^f \leq J^f(z) \leq \overline{J}^f, \underline{J}^\mu \leq J^\mu(x) \leq \overline{J}^\mu, \forall z \in \mathcal{Z}, \forall x \in \mathcal{X}$*, where* $J^f(z)$ *and* $J^\mu(x)$ *denote the Jacobian matrices of the mappings* $f(\cdot)$ *and* $\mu(\cdot)$ *at the points* z *and* x*, respectively.*

In this chapter, we aim to propose novel set-membership approaches for obtaining polytope-valued state estimates for bounded-error nonlinear systems (9.10), using indirect polytope representations, namely using zonotope bundles (ZBs) and constrained zonotopes (CZs). More formally, given the initial state set estimate $\hat{\mathcal{X}}_0$, where $x_0 \in \hat{\mathcal{X}}_0$, we consider a two-step approach for recursive state estimation by solving the following problems for the propagation and update steps, respectively, at each time step $k \in \mathbb{N}$:

Problem 9.2.3 (Propagation)**.** *Given the 'updated set'* $\mathcal{X}^u_{k-1}$ *from the previous time* *step and $\mathcal{W}_{k-1}$ (with $\mathcal{Z}_{k-1} \triangleq \mathcal{X}^u_{k-1} \times \mathcal{W}_{k-1}$), find the 'propagated set' $\mathcal{X}^p_k$ that satisfies*

$$f(\mathcal{Z}_{k-1}) \triangleq \{\hat{f}(x, w, u_{k-1}) \mid x \in \mathcal{X}^u_{k-1}, w \in \mathcal{W}_k\} \subseteq \mathcal{X}^p_k. \tag{9.11}$$

Problem 9.2.4 (Update)**.** *Given the 'propagated set' $\mathcal{X}^p_k$ and the uncertain observation/constraint set $\mathcal{Y}_k$ at time step k, find the 'updated set' $\mathcal{X}^u_k$ that satisfies*

$$\mathcal{X}^p_k \cap_\mu \mathcal{Y}_k \triangleq \{x \in \mathcal{X}^p_k \mid \mu(x) \in \mathcal{Y}_k\} \subseteq \mathcal{X}^u_k. \tag{9.12}$$

9.3 Indirect Polytopic Set Computation

We consider a *recursive* two-step state estimation approach consisting of state propagation (prediction) and measurement update (refinement) steps, by solving Problems 9.2.3 and 9.2.4 in Sections 9.3.1 and 9.3.2, respectively. Our recursive algorithm can be either initialized at time step 0 with the initial polytopic state estimate $\mathcal{X}_0$ as $\mathcal{X}^u_0 = \mathcal{X}_0$ or if $\mathcal{Y}_0$ is available/measured, with $\mathcal{X}^p_0 = \hat{\mathcal{X}}_0$ and the application of the update step by solving Problem 9.2.4 at time 0 to obtain $\mathcal{X}^u_0$.

9.3.1 Decomposition-Based ZB/CZ Propagation Step

In this section, we address Problem 9.2.3, assuming that the state estimate set from the previous time step is a zonotope bundle (Lemma 9.3.1) or a constraint zonotope (Lemma 9.3.2). The main idea is to "transform" the ZBs/CZs from the z-space, i.e., the space of augmented state x and process uncertainty w, to intervals in the ξ-space, i.e., the space of ZB/CZ generators. Then, based on our recent results in [63], we decompose the transformed vector fields in the ξ-space into two components, a Jacobian sign stable (JSS) and a linear remainder mapping (cf. Corollary 9.1.7). Finally, we apply our recently developed approach to find a family of mixed-monotone remainder-form decomposition functions to compute enclosures to the JSS components, which are proven to be tight by Corollary 9.1.7 for interval domains. Using these tight bounds and thanks to the linearity of the remainders, we show that by augmenting and intersecting all the obtained enclosures, the resultant set is a ZB/CZ. We formally summarize our proposed Decomposition-Based ZB/CZ approaches in the following Lemmas 9.3.1 and 9.3.2.

Lemma 9.3.1 (Decomposition-Based ZB Propagation). *Suppose* $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *satisfies Assumption 9.2.2. Let* $\mathcal{Z}$ *be a ZB in* $\mathbb{R}^{n_z}$*, i.e.,* $\mathcal{Z} = \bigcap_{s=1}^{S} \{G_s, c_s\}_Z$*, and* $\forall s \in \mathbb{N}_S \triangleq \{1, \ldots, S\}$*,* n_s *be the number of generators of the corresponding zonotope. Then, the following set inclusion holds:*

$$f(\mathcal{Z}) \subseteq \mathcal{ZB}_f \triangleq \bigcap_{s=1}^{S} \bigcap_{H_s \in \mathbf{H}_{\tilde{f}_s}} \{G_s^{H_s}, c_s^{H_s}\}_Z, \tag{9.13}$$

where $G_s^{H_s} \triangleq [H_s \ \tfrac{1}{2}\mathrm{diag}(\overline{g}_s^{H_s} - \underline{g}_s^{H_s})]$*,* $c_s^{H_s} \triangleq \tfrac{1}{2}(\overline{g}_s^{H_s} + \underline{g}_s^{H_s})$*,*

$$\overline{g}_{s,i}^{H_s} \triangleq g_{i,d}^{s}(\mathbf{1}_{n_s}, -\mathbf{1}_{n_s}; {H_s^\top}_{(i,:)}, \langle {H_s^\top}_{(i,:)}, \cdot\rangle), \tag{9.14}$$

$$\underline{g}_{s,i}^{H_s} \triangleq g_{i,d}^{s}(-\mathbf{1}_{n_s}, \mathbf{1}_{n_s}; {H_s^\top}_{(i,:)}, \langle {H_s^\top}_{(i,:)}, \cdot\rangle), \tag{9.15}$$

while $g_{i,d}^{s}(\cdot, \cdot; {H_s^\top}_{(i,:)}, \langle {H_s^\top}_{(i,:)}, \cdot\rangle)$ *is the tight mixed-monotone decomposition function (cf. Proposition 9.1.6) for the JSS mapping* $g_{s,i}^{H_s}(\xi) \triangleq \tilde{f}_{s,i}(\xi) - \langle {H_s^\top}_{(i,:)}, \xi\rangle : \mathbb{B}_\infty^{n_s} \to \mathbb{R}^{n_x}$*,* $\mathbf{H}_{\tilde{f}_s}$ *is defined in Corollary 9.1.7 (with the corresponding function being* $\tilde{f}_s$*) and* $\tilde{f}_s(\xi) \triangleq f(c_s + G_s\xi)$*.*

Lemma 9.3.2 (Decomposition-Based CZ Propagation). *Suppose* $f : \mathcal{Z} \subset \mathbb{R}^{n_z} \to \mathbb{R}^{n_x}$ *satisfies Assumption 9.2.2 and let* $\mathcal{Z}$ *be a CZ in* $\mathbb{R}^{n_z}$*, i.e.,* $\mathcal{Z} = \{\tilde{G}, \tilde{c}, \tilde{A}, \tilde{b}\}_{CZ}$*, and* n_g *be the number of generators of* $\mathcal{Z}$*. Then, the following set inclusion holds:*

$$f(\mathcal{Z}) \subseteq \mathcal{CZ}_f \triangleq \bigcap_{H \in \mathbf{H}_{\tilde{f}}} \{\tilde{G}^H, \tilde{c}^H, \mathbb{A}, \tilde{b}\}_{CZ}, \tag{9.16}$$

where $\tilde{G}^H \triangleq [H \quad \tfrac{1}{2}\mathrm{diag}(\overline{g}^H - \underline{g}^H)], \mathbb{A} \triangleq [\tilde{A} \quad 0_{n_g \times n_x}]$*,*

$$\overline{g}_i^H \triangleq \tilde{g}_{i,d}(\bar{\mathbf{l}}_{n_g}, \underline{\mathbf{l}}_{n_g}; H_{i,:}^\top, \langle H_{i,:}^\top, \cdot\rangle), \tilde{c}^H \triangleq \frac{1}{2}(\overline{g}^H + \underline{g}^H), \tag{9.17}$$

$$\underline{g}_i^H \triangleq \tilde{g}_{i,d}(\underline{\mathbf{l}}_{n_g}, \bar{\mathbf{l}}_{n_g}; H_{i,:}^\top, \langle H_{i,:}^\top, \cdot\rangle), \tag{9.18}$$

$\bar{\mathbf{l}}_{n_g} \triangleq \min(\mathbf{1}_{n_g}, \tilde{A}^\dagger\tilde{b} + \kappa\mathbf{r}_{n_g}), \underline{\mathbf{l}}_{n_g} \triangleq \max(-\mathbf{1}_{n_g}, \tilde{A}^\dagger\tilde{b} - \kappa\mathbf{r}_{n_g}), \tilde{g}_{i,d}(\cdot,\cdot;H^\top_{i,:},\langle H^\top_{i,:},\cdot\rangle)$ *is the tight mixed-monotone decomposition function (cf. Proposition 9.1.6) for the JSS mapping* $\tilde{g}_i(\xi) \triangleq \tilde{f}_i(\xi) - \langle H^\top_{i,:}, \xi\rangle : \mathbb{B}^{n_g}_\infty \to \mathbb{R}^{n_x}$*,* $\mathbf{H}_{\tilde{f}}$ *is defined in Corollary 9.1.7,* $\tilde{f}(\xi) \triangleq f(\tilde{c}+\tilde{G}\xi)$*,* $\mathbf{r}_{n_g} \triangleq \mathrm{rowsupp}(I_{n_g} - \tilde{A}^\dagger\tilde{A})$ *and* κ *is a very large positive real number (infinity).*

Finally, for further improvement, we can take the intersection of the resulting propagated sets in Lemmas 9.3.1 and 9.3.2. This is formally summarized in the following Theorem 9.3.3.

Theorem 9.3.3 (Decomposition-Based ZB/CZ Propagation)**.** *Suppose all the assumptions in Lemmas 9.3.1 and 9.3.2 hold. Then,* $f(\mathcal{Z}) \subseteq \mathcal{ZB}_f \cap \mathcal{CZ}_f$*, where* $\mathcal{ZB}_f, \mathcal{CZ}_f$ *are computed in Lemmas 9.3.1 and 9.3.2, respectively.*

9.3.2 Decomposition-Based CZ/ZB Update Step

In this section, we address Problem 9.2.4 for a given a locally Lipschitz nonlinear vector field $\mu(\cdot)$ and assuming that the initial propagated and the observation/constraint sets at each time step k are zonotope bundles (Lemma 9.3.4) or constraint zonotopes (Lemma 9.3.5). Using a similar idea as in Section 9.3.1, i.e, considering the space of generators, decomposing the transformed observation function into a JSS and a linear component, applying the tight remainder-form decomposition functions [63] to bound the JSS component, augmenting and intersecting, as well as taking the advantage of linear remainder functions, we obtain ZB/CZ enclosures to the nonlinear generalized intersection in (9.12). The results of this section are summarized in Lemmas 9.3.4 and 9.3.5 and Theorem 9.3.6.

Lemma 9.3.4 (Decomposition-Based ZB Update)**.** *Suppose* $\mu : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\mu}$ *satisfies Assumption 9.2.2. Let* $\mathcal{Z}_f \subset \mathbb{R}^{n_x}$ *and* $Z_\mu \subset \mathbb{R}^{n_\mu}$ *be two ZB sets, i.e.,* $\mathcal{Z}_f = \mathcal{ZB}_f =$

$\bigcap_{r=1}^{R}\{G_f^r, c_f^r\}_Z$ and $\mathcal{Z}_\mu = \mathcal{ZB}_\mu = \bigcap_{t=1}^{T}\{G_\mu^t, c_\mu^t\}_Z$, and $\forall r \in \mathbb{N}_R \triangleq \{1, \ldots, R\}, \forall t \in \mathbb{N}_T \triangleq \{1, \ldots, T\}$, let n_r, n_t be the number of generators of the corresponding zonotopes, respectively. Then, the following set inclusion holds:

$$\mathcal{ZB}_f \cap_\mu \mathcal{ZB}_\mu \subseteq \mathcal{ZB}_u \triangleq \bigcap_{r=1}^{R} \bigcap_{t=1}^{T} \bigcap_{Q_r \in \mathbf{Q}_{\tilde{\mu}_r}} \{\hat{G}_r^t, \hat{c}_r, \hat{A}_{r,t}^{Q_r}, \hat{b}_{r,t}^{Q_r}\}_{CZ}, \tag{9.19}$$

where $\hat{G}_r^t \triangleq [G_f^r \ \mathbf{0}^t], \hat{c}_r \triangleq c_f^r,\ \hat{b}_{r,t}^{Q_r} \triangleq c_\mu^t - \frac{1}{2}(\overline{p}_r^{Q_r} + \underline{p}_r^{Q_r})$,

$$\hat{A}_{r,t}^{Q_r} \triangleq \begin{bmatrix} Q_r & -G_\mu^t & \frac{1}{2}\mathrm{diag}(\overline{p}_r^{Q_r} - \underline{p}_r^{Q_r}) \end{bmatrix},$$

$$\overline{p}_{r,i}^{Q_r} \triangleq p_{i,d}^r(\mathbf{1}_{n_r}, -\mathbf{1}_{n_r}; Q_{r(i,:)}^\top, \langle Q_{r(i,:)}^\top, \cdot \rangle), \tag{9.20}$$

$$\underline{p}_{r,i}^{Q_r} \triangleq p_{i,d}^r(-\mathbf{1}_{n_r}, \mathbf{1}_{n_r}; Q_{r(i,:)}^\top, \langle Q_{r(i,:)}^\top, \cdot \rangle), \tag{9.21}$$

$p_{i,d}^r(\cdot, \cdot; Q_r, \langle Q_{r(i,:)}^\top, \cdot \rangle)$ is the tight mixed-monotone decomposition function (cf. Proposition 9.1.6) for the JSS mapping $p_{r,i}^{Q_r}(\alpha) \triangleq \tilde{\mu}_{r,i}(\alpha) - \langle Q_{r(i,:)}^\top, \alpha \rangle : \mathbb{B}_\infty^{n_r} \to \mathbb{R}^{n_\mu}$, $\mathbf{Q}_{\tilde{\mu}_r}$ is defined similar to $\mathbf{H}_f$ in Corollary 9.1.7 (with the corresponding function being $\tilde{\mu}_r(\alpha) \triangleq \mu(c_f^r + G_f^r \alpha)$) and $\mathbf{0}^t$ is a zero matrix in $\mathbb{R}^{n_x \times (n_t + n_\mu)}$.

Lemma 9.3.5 (Decomposition-Based CZ Update)**.** *Suppose $\mu : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\mu}$ satisfies Assumption 9.2.2. Let $\mathcal{Z}_f \subset \mathbb{R}^{n_x}$ and $\mathcal{Z}_\mu \subset \mathbb{R}^{n_\mu}$ be two CZ sets, i.e., $\mathcal{Z}_f = \mathcal{CZ}_f = \{\tilde{G}_f, \tilde{c}_f, \tilde{A}_f, \tilde{b}_f\}_{CZ}$ and $\mathcal{Z}_\mu = \mathcal{CZ}_\mu = \{\tilde{G}_\mu, \tilde{c}_\mu, \tilde{A}_\mu, \tilde{b}_\mu\}_{CZ}$, and n_c, n_τ be the number of generators of $\mathcal{Z}_f, \mathcal{Z}_\mu$, respectively. Then, the following set inclusion holds:*

$$\mathcal{CZ}_f \cap_\mu \mathcal{CZ}_\mu \subseteq \mathcal{CZ}_u \triangleq \bigcap_{\Omega \in \mathbf{\Omega}_\lambda} \{\mathbb{G}, \tilde{c}_f, \mathbb{A}_\Omega, \tilde{b}_\Omega\}_{CZ}, \tag{9.22}$$

where $\mathbb{G} \triangleq [\tilde{G}_f \ 0 \ 0]$, $\tilde{b}_\Omega \triangleq [\tilde{b}_f^\top \ \tilde{b}_\mu^\top \ (\tilde{c}_f - \frac{1}{2}(\overline{\nu}^\Omega + \underline{\nu}^\Omega))^\top]^\top$,

$$\mathbb{A}_\Omega \triangleq \begin{bmatrix} \tilde{A}_f & 0 & 0 \\ 0 & \tilde{A}_\mu & \\ \Omega & -\tilde{G}_\mu & \frac{1}{2}\mathrm{diag}(\overline{\nu}^\Omega - \underline{\nu}^\Omega) \end{bmatrix},$$

$$\overline{\nu}_i^\Omega \triangleq \nu_{i,d}(\bar{\mathbf{l}}_{n_c}, \underline{\mathbf{l}}_{n_c}; \Omega^\top_{(i,:)}, \langle \Omega^\top_{(i,:)}, \cdot \rangle), \tag{9.23}$$

$$\underline{\nu}_i^\Omega \triangleq \nu_{i,d}(\underline{\mathbf{l}}_{n_c}, \bar{\mathbf{l}}_{n_c}; \Omega^\top_{(i,:)}, \langle \Omega^\top_{(i,:)}, \cdot \rangle), \tag{9.24}$$

$\bar{\mathbf{l}}_{n_c} \triangleq \min(\mathbf{1}_{n_c}, \tilde{A}_f^\dagger \tilde{b}_f + \kappa \mathbf{r}_{n_c}), \underline{\mathbf{l}}_{n_c} \triangleq \max(-\mathbf{1}_{n_c}, \tilde{A}_f^\dagger \tilde{b}_f - \kappa \mathbf{r}_{n_c}), \nu_{i,d}(\cdot, \cdot; \Omega^\top_{(i,:)}, \langle \Omega^\top_{(i,:)}, \cdot \rangle)$ *is the tight mixed-monotone decomposition function (cf. Proposition 9.1.6) for the JSS mapping* $\nu_i^\Omega(\beta) \triangleq \lambda_i(\beta) - \langle \Omega^\top_{(i,:)}, \beta \rangle : \mathbb{B}_\infty^{n_c} \to \mathbb{R}^{n_\mu}$, $\mathbf{\Omega}_\lambda$ *is defined similar to* $\mathbf{H}_f$ *in Corollary 9.1.7 (with the corresponding function being* $\lambda(\beta) \triangleq \mu(\tilde{c}_f + \tilde{G}_f \beta)$*),* $\mathbf{r}_{n_c} \triangleq \mathrm{rowsupp}(I_{n_c} - \tilde{A}_f^\dagger \tilde{A}_f)$ *and* κ *is a very large positive real number (infinity).*

We conclude this subsection by combining the results in Lemmas 9.3.4 and 9.3.5 via the following Theorem 9.3.6.

Theorem 9.3.6 (Decomposition-Based ZB/CZ Update)**.** *Suppose all the assumptions in Lemmas 9.3.4 and 9.3.5 hold. Then*

$$\mathcal{Z}_f \cap_\mu \mathcal{Z}_\mu \subseteq \mathcal{ZB}_u \cap \mathcal{CZ}_u,$$

where $\mathcal{ZB}_u, \mathcal{CZ}_u$ *are given in Lemmas 9.3.4 and 9.3.5, respectively.*

9.3.3 Modifications to the Approach in [99]

The purpose of this subsection is twofold. i) We propose a potential refinement/improvement to the propagation approach in [99, Theorem 2] (recapped in Proposition 9.1.4) through the following Proposition 9.3.7, by applying our previously developed remainder-form decomposition functions to compute potentially tighter enclosing intervals to Jacobian matrix of $f(\cdot)$; ii) We propose an update method via Lemma 9.3.8, that is based on the "CZ-inclusion" introduced in [99, Theorem 1] (recapped in Proposition 9.1.3). The proposed update method is applicable to general nonlinear observation functions (similar to the proposed methods in Lemmas 9.3.4 and 9.3.5), as opposed to the update (i.e, linear intersection) approach in [99] that is only applicable when the observation function is linear.

Proposition 9.3.7 (Refinement to the Propagation Approach in [99])**.** *Suppose all the assumptions in Proposition 9.1.6 (i.e, [99, Theorem 2]) hold. Then the set inclusion in* (9.5) *also holds when replacing* $\mathbb{J}$ *with* $\tilde{\mathbb{J}}$ *(or the best (tightest) of them), where* $\tilde{\mathbb{J}}$ *is an enclosing interval to* $g(x) \triangleq \nabla_x^\top f(X, W)$ *that can be computed by applying Proposition 9.1.6 to the function* $g(\cdot)$*.*

Lemma 9.3.8 (Update Based on "CZ-Inclusion" in [99])**.** *Suppose all the assumptions in Lemma 9.3.5 hold. Let* $x_0 \in \mathcal{CZ}_f$ *and* $\mathbb{J}^\mu, \mathbb{J}^\mu_\Delta \in \mathbb{R}^{n_\mu \times n_x}$ *be interval matrices satisfying* $J^\mu(\mathcal{CZ}_f) \subseteq \mathbb{J}^\mu$ *and* $\forall i \in \mathbb{N}_{n_\mu}, \forall j \in \mathbb{N}_{n_x}$, $[\mathbb{J}^\mu_\Delta]_{ij} \triangleq \frac{1}{2}\begin{bmatrix} -\mathrm{diam}(\mathbb{J}^\mu)_{ij} & \mathrm{diam}(\mathbb{J}^\mu)_{ij} \end{bmatrix}$*, where* J^μ *denotes the Jacobian of* $\mu(\cdot)$*. Let* $\overline{\mathcal{Z}}_f = \{\overline{G}^f, \overline{c}^f\}_Z$ *be a zonotope satisfying* $\mathcal{CZ}_f \ominus \{x_0\} \subseteq \overline{\mathcal{Z}}_f$*, with* $\overline{c}^f \in \mathbb{R}^{\overline{n}}$*, let* $\mathbf{m}^\mu \in \mathbb{R}^{n_\mu}$ *be an interval vector such that* $\mathbf{m}^\mu \supset \mathbb{J}^\mu_\Delta \overline{c}^f$ *and* $\mathrm{mid}(\mathbf{m}^\mu) = \mathbf{0}_{n_\mu}$ *and let* $P^\mu \in \mathbb{R}^{n_\mu \times n_\mu}$ *be a diagonal matrix defined as follows:* $\forall i = 1, \ldots, n_\mu$*:*

$$P^\mu_{ii} = \frac{1}{2}\mathrm{diam}(\mathbf{m}^\mu)_i + \tfrac{1}{2}\textstyle\sum_{j=1}^{\overline{n}} \sum_{k=1}^{n_x} \mathrm{diam}(\mathbb{J}^\mu_\Delta)_{ik} |\overline{G}^f_{kj}|. \tag{9.25}$$

Then, the following set inclusion holds:

$$\mathcal{CZ}_f \cap_\mu \mathcal{CZ}_\mu \subseteq \mathcal{CZ}^R_u \triangleq \{G_u, c_u, A_u, b_u\}_{CZ}, \tag{9.26}$$

where

$$A_u \triangleq \begin{bmatrix} \mathrm{mid}(\mathbb{J}^\mu)\tilde{G}_f & -\tilde{G}_\mu & G_R \\ \tilde{A}_f & 0 & 0 \\ 0 & \tilde{A}_\mu & 0 \\ 0 & 0 & A_R \end{bmatrix}, b_u \triangleq \begin{bmatrix} \tilde{c}_\mu - \mu(x_0) - c_R + \mathrm{mid}(\mathbb{J}^\mu)(x_0 - \tilde{c}_f) \\ \tilde{b}_f \\ \tilde{b}_\mu \\ b_R \end{bmatrix},$$

$$G_R \triangleq [0\ P^\mu],\ c_R \triangleq 0,\ A_R \triangleq [\tilde{A}_f\ 0],\ b_R \triangleq \tilde{b}_f, G_u \triangleq [\tilde{G}_f\ 0\ 0]. \tag{9.27}$$

9.4 Simulation Results

In this section we compare the performance of five approaches to guaranteed state estimation: i) RRSR, i.e., the mean value extension-based propagation introduced in [99] (recapped in Proposition 9.1.4) in addition to the update approach in [99] for the case when the observation function is linear (for Example I below) and its extension in Lemma 9.3.8 to nonlinear measurements (for Example II below), ii) D-RRSR, i.e, a modification to RRSR where the bounds for Jacobian matrices are computed using the reminder-form decomposition functions (cf. Proposition 9.3.7), iii) D-ZB, i.e., decomposition-based propagation and update with ZBs (cf. Lemmas 9.3.1 and 9.3.4), iv) D-CZ, i.e., decomposition-based propagation and update with CZs (cf. Lemmas 9.3.2 and 9.3.5) and v) COMB, i.e., a combination of i)–v) via intersection (based on a similar idea as Theorems 9.3.3, 9.3.6). All simulations are performed on a 1.8 GHz (8 CPUS) i5-8250U, using MATLAB version 2020a and CORA 2020 [8].

9.4.1 Example I

Consider the following nonlinear discrete-time system from [99, Example 1]

$$\begin{aligned}
x_{1,k} &= 3x_{1,k-1} - \tfrac{x_{1,k-1}^2}{7} - \tfrac{4x_{1,k-1}x_{2,k-1}}{4+x_{1,k-1}} + w_{1,k-1}, \\
x_{2,k} &= -2x_{2,k-1} + \tfrac{3x_{1,k-1}x_{2,k-1}}{4+x_{1,k-1}} + w_{2,k-1}, \\
\begin{bmatrix} y_{1,k} \\ y_{2,k} \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} v_{1,k} \\ v_{2,k} \end{bmatrix},
\end{aligned} \tag{9.28}$$

with $\|w_k\|_\infty \leq 0.1$, an unknown initial state $x_0 \in \mathcal{X}_0 = \left\{ \begin{bmatrix} 0.1 & 0.2 & -0.1 \\ 0.1 & 0.1 & 0 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \right\}$ and $\mathcal{Y}_k \triangleq \{y_k - v_k \mid \|v_k\|_\infty \leq 0.4\}$.

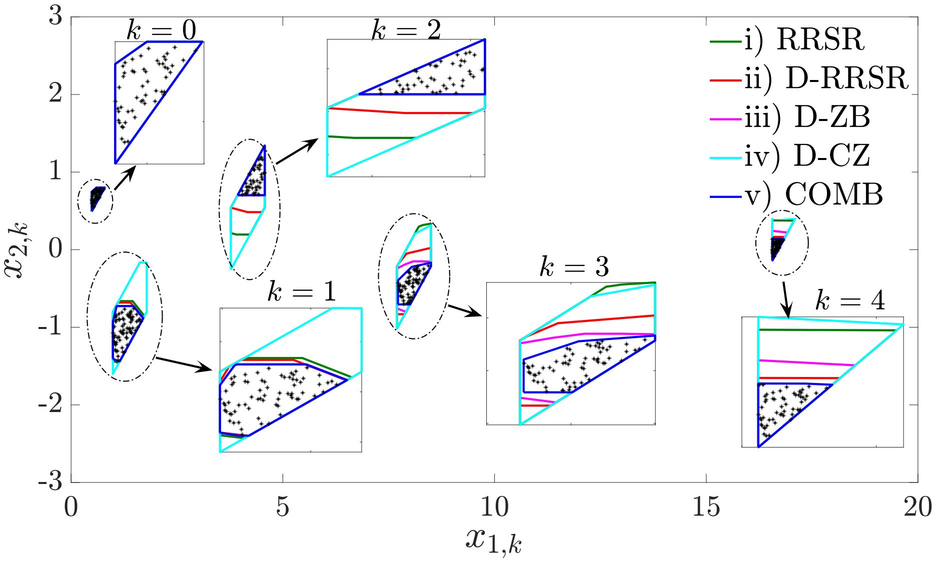


Figure 9.1: Results for Example I from the First Five Time Steps of Set-valued State Estimation, Using Five Different Approaches: Black Dots Are Obtained from Uniform Sampling of the Initial State and Noise Signals, and Propagating Through the System Dynamics

As can be seen from Figure 9.1, D-ZB provides less conservative enclosures compared to the other individual approaches, and further, the COMB approach results in significant improvement by taking advantage of all approaches via intersection. Moreover, a more systematic comparison of the average computation times and enclosure set volumes of the five approaches is given in Table 9.1. It can be observed that D-ZB is the fastest computationally, while the combination of all approaches, i.e., COMB, took the longest, as expected. Furthermore, RRSR and D-RRSR took approximately the same time on average. In terms of average set volumes, D-ZB and D-RRSR generate the least conservative (smallest) enclosures when compared to the other approaches, while a further improvement is obtained using the intersection of all approaches (COMB).

Table 9.1: Average Total Times (Seconds) and Average Total Volumes at Each Time Step for Five State Estimators in Example I: Each Average Is Taken Over 50 Simulations with Uniformly Sampled Noise and Initial State

Methods:		$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
RRSR	Time:	0.0869	0.2496	0.1926	0.1960	0.2042
	Vol.:	0.2012	0.5002	0.6205	0.4811	0.3340
D-RRSR	Time:	0.0866	0.2251	0.1809	0.1977	0.2005
	Vol.:	0.2012	0.4758	0.6008	0.4385	0.1472
D-ZB	Time:	0.0882	0.0949	0.0906	0.0907	0.1226
	Vol.:	0.2012	0.4518	0.5729	0.32721	0.3175
D-CZ	Time:	0.0869	2.8245	2.9200	2.1183	3.3176
	Vol.:	0.2012	0.5673	0.6310	0.5061	0.4169
COMB	Time:	0.0872	6.1929	6.8815	6.2782	6.908
	Vol.:	0.2012	0.4485	0.5659	0.2841	0.1465

9.4.2 Example II (The Unicycle System)

Now consider the following discretized unicycle-like mobile robot system [25]:

$$\begin{aligned} s_{x,k+1} &= s_{x,k} + T_0\phi_w \cos(\theta_k) + w_{1,k}, \\ s_{y,k+1} &= s_{y,k} + T_0\phi_w \sin(\theta_k) + w_{2,k}, \\ \theta_{k+1} &= \theta_k + T_0\phi_\theta + w_{3,k}, \\ y_k &= [d_{1,k}\ \phi_{1,k}\ d_{2,k}\ \phi_{2,k}]^\top + v_k, \end{aligned} \tag{9.29}$$

where $x_k \triangleq [s_{x,k}\ s_{y,k}\ \theta_k]^\top$, $w_k = [w_{x,k}\ w_{y,k}\ w_{\theta,k}]^\top, \phi_{\omega,k} = 0.3, \phi_{\theta,k} = 0.15, w_{x,k} = 0.2(0.5\rho_{x_{1,k}} - 0.3), w_{y,k} = 0.2(0.3\rho_{x_{2,k}} - 0.2)$ and $w_{\theta,k} = 0.2(0.6\rho_{x_{3,k}} - 0.4)$, with $\rho_{x_{l,k}} \in [0,1]$ $(l = 1,2,3)$ and initial state $x_0 = [0.1\ 0.2\ 1]^\top$. Moreover, $\forall i \in \{1,2\}$, $d_{i,k} = \sqrt{(s_{x_i} - s_{x,k})^2 + (s_{y_i} - s_{y,k})^2}$ and $\phi_{i,k} = \theta_k - \arctan(\frac{s_{y_i} - s_{y,k}}{s_{x_i} - s_{x,k}})$, with s_{x_i}, s_{y_i} being two known values. Furthermore, $\mathcal{Y}_k \triangleq \{y_k - v_k \mid v_{1,k} = 0.02\rho_{y_1,k} - 0.01, v_{2,k} = 0.03\rho_{y_2,k} - 0.01, v_{3,k} = 0.03\rho_{y_3,k} - 0.02, v_{4,k} = 0.05\rho_{y_4,k} - 0.03, \rho_{y_k,k} \in [0,1], \forall k = \{1,2,3,4\}\}$.

Applying all methods i) through v), one can observe from Figure 9.2 that the resulting set estimates appear comparable for all approaches. Upon closer examination, Table 9.2 shows that D-CZ takes the least average computation time followed by RRSR, D-RRSR, COMB and D-ZB, while in terms of average set volumes, the COMB approach results in the smallest volume followed by D-ZB, D-CZ, RRSR and D-RRSR. Note that the computation time for D-ZB is exceptionally large, presumably because of the specific implementation in CORA 2020 [8] for converting a polytope to its ZB representation that could result in a higher number of zonotopes than the minimal needed to exactly represent the same polytope. Thus, the reduction of the number of zonotopes in the bundle could be an interesting future topic, which could significantly decrease the computation time of the D-ZB approach.

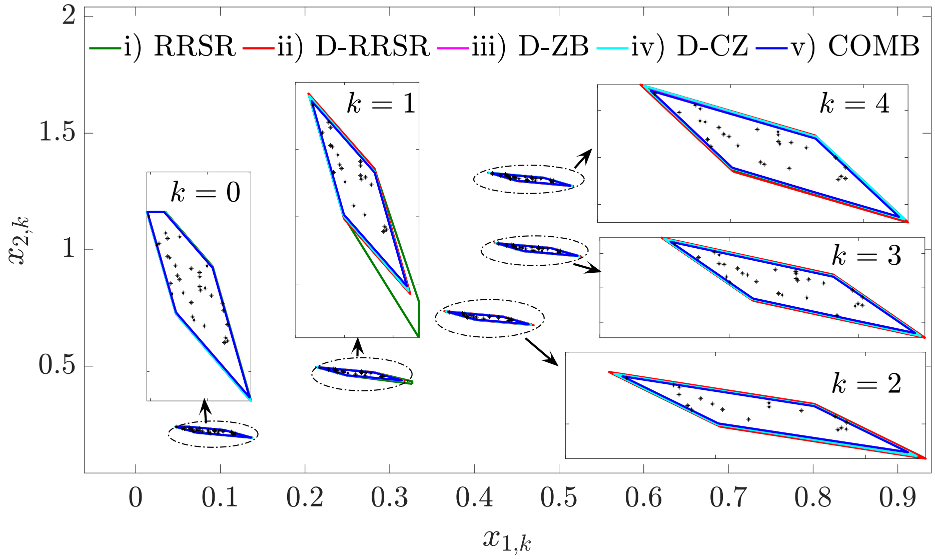


Figure 9.2: Results for Example II from the First Five Time Steps of Set-valued State Estimation, Using Five Different Approaches: Black Dots Are Obtained from Uniform Sampling of the Initial State and Noise Signals, and Propagating Through the System Dynamics

9.5 Conclusion

Novel methods were presented in this chapter for guaranteed state estimation in bounded-error discrete-time nonlinear systems subject to nonlinear observations and/or constraints using indirect polytopic representations, i.e., using ZBs/CZs. By considering polytopes in the space of ZB/CZ's generators, our recent results on remainder-form mixed-monotone decomposition functions can be applied to compute enclosures that are guaranteed to enclose the set of all possible state trajectories. Further, the decomposition functions were leveraged to bound the nonlinear observation function to derive the updated set, i.e., to return enclosures to the intersection of the propagated set and the set of states that are consistent with noisy measurements. Finally, the mean value extension-based approach in [99] was also generalized to

Table 9.2: Average Total Times (Seconds) and Average Total Volumes (10^{-5}) at Each Time Step for Five State Estimators in Example II: Each Average Is Taken Over 20 Simulations with Uniformly Sampled Noise and Initial State

Methods:		$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
RRSR	Time:	0.7719	4.2557	4.1883	2.9950	3.6747
	Vol.:	4.2924	3.7834	1.6171	4.5738	4.5558
D-RRSR	Time:	1.6690	42.905	45.571	28.642	50.539
	Vol.:	4.0527	3.2943	1.6600	5.1036	4.8001
D-ZB	Time:	1.3967	34.207	163.08	147.75	131.94
	Vol.:	4.2551	2.9697	1.3248	4.2917	4.2519
D-CZ	Time:	0.6020	2.1824	2.0195	2.2281	2.5908
	Vol.:	4.2551	3.0793	1.4337	4.6017	4.4564
COMB	Time:	0.2361	34.902	65.501	62.371	57.728
	Vol.:	4.0527	2.7726	1.2220	3.9126	3.9914

compute the updated set when the observation functions are nonlinear.

REFERENCES

[1] M. Abate and S. Coogan. Computing robustly forward invariant sets for mixed-monotone systems. *arXiv preprint arXiv:2003.05912*, 2020.

[2] M. Abate, M. Dutreix, and S. Coogan. Tight decomposition functions for continuous-time mixed-monotone systems with disturbances. *arXiv preprint arXiv:2003.07975*, 2020.

[3] M. Abolhasani and M. Rahmani. Robust deterministic least-squares filtering for uncertain time-varying nonlinear systems with unknown inputs. *Systems & Control Letters*, 122:1–11, 2018.

[4] B. Açıkmeşe and M. Corless. Observers for systems with nonlinearities satisfying incremental quadratic constraints. *Automatica*, 47(7):1339–1348, 2011.

[5] T. Alamo, J.M. Bravo, and E.F. Camacho. Guaranteed state estimation by zonotopes. *Automatica*, 41(6):1035–1043, 2005.

[6] T. Alamo, J.M. Bravo, M.J. Redondo, and E.F. Camacho. A set-membership state estimation algorithm based on dc programming. *Automatica*, 44(1):216–224, 2008.

[7] G. Alefeld and G. Mayer. Interval analysis: theory and applications. *Journal of computational and applied mathematics*, 121(1-2):421–464, 2000.

[8] M. Althoff. https://tumcps.github.io/cora/data/cora2020manual.pdf. 2020.

[9] M. Althoff and B.H Krogh. Zonotope bundles for the efficient computation of reachable sets. In *2011 50th IEEE conference on decision and control and European control conference*, pages 6814–6821. IEEE, 2011.

[10] B.D.O. Anderson and J.B. Moore. Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM Journal on Control and Optimization*, 19(1):20–32, 1981.

[11] D. Angeli and E.D. Sontag. Monotone control systems. *IEEE Transactions on automatic control*, 48(10):1684–1698, 2003.

[12] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1.*, 2017.

[13] G. Beliakov. Interpolation of Lipschitz functions. *Journal of computational and applied mathematics*, 196(1):20–44, 2006.

[14] O. Bernard and J-L. Gouzé. Closed loop observers bundle for uncertain biotechnological models. *Journal of Process Control*, 14(7):765–774, 2004.

[15] D.P. Bertsekas, A. Nedich, A.E Ozdaglar, et al. Convex analysis and optimization. 2003.

[16] F. Blanchini and S. Miani. *Set-theoretic methods in control.* Springer, 2008.

[17] F. Blanchini and M. Sznaier. A convex optimization approach to synthesizing bounded complexity ℓ^∞ filters. *IEEE Transactions on Automatic Control*, 57(1):216–221, 2012.

[18] H.L. Bodlaender, P. Gritzmann, V. Klee, and J. Van Leeuwen. Computational complexity of norm-maximization. *Combinatorica*, 10(2):203–225, 1990.

[19] J.P. Calliess. *Conservative decision-making and inference in uncertain dynamical systems.* PhD thesis, University of Oxford, 2014.

[20] M. Canale, L. Fagiano, and M.C. Signorile. Nonlinear model predictive control from data: a set membership approach. *International Journal of Robust and Nonlinear Control*, 24(1):123–139, 2014.

[21] A.A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Conference on Hot Topics in Security*, pages 6:1–6:6, 2008.

[22] A. Chakrabarty and M. Corless. Estimating unbounded unknown inputs in nonlinear systems. *Automatica*, 104:57–66, 2019.

[23] A. Chakrabarty, S.H. Żak, and S. Sundaram. State and unknown input observers for discrete-time nonlinear systems. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7111–7116. IEEE, 2016.

[24] K. Chandrasekar and M.S Hsiao. Implicit search-space aware cofactor expansion: A novel preimage computation technique. In *2006 International Conference on Computer Design*, pages 280–285. IEEE, 2006.

[25] B. Chen and G. Hu. Nonlinear state estimation under bounded noises. *Automatica*, 98:159–168, 2018.

[26] J. Chen and C.M. Lagoa. Observer design for a class of switched systems. In *IEEE Conference on Decision and Control European Control Conference*, pages 2945–2950, 2005.

[27] L. Chisci, G.A., and G. Zappa. Recursive state bounding by parallelotopes. *Automatica*, 32(7):1049–1055, 1996.

[28] M.S. Chong, M. Wakaiki, and J.P. Hespanha. Observability of linear systems under adversarial attacks. In *IEEE American Control Conference (ACC)*, pages 2439–2444, 2015.

[29] T. Chu and L. Huang. Mixed monotone decomposition of dynamical systems with application. *Chinese science bulletin*, 43(14):1171–1175, 1998.

[30] C. Combastel. A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 7228–7234. IEEE, 2005.

[31] C. Combastel. Merging kalman filtering and zonotopic state bounding for robust fault detection under noisy environment. *IFAC-PapersOnLine*, 48(21):289–295, 2015.

[32] S. Coogan and M. Arcak. Efficient finite abstraction of mixed monotone systems. In *Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control*, pages 58–67, 2015.

[33] S. Coogan and M. Arcak. Stability of traffic flow networks with a polytree topology. *Automatica*, 66:246–253, 2016.

[34] S. Coogan, M. Arcak, and A.A. Kurzhanskiy. Mixed monotonicity of partial first-in-first-out traffic flow models. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 7611–7616. IEEE, 2016.

[35] H. Cornelius and R. Lohner. Computing the range of values of real functions with accuracy higher than second order. *Computing*, 33(3-4):331–347, 1984.

[36] G. De Nicolao, G. Sparacino, and C. Cobelli. Nonparametric input estimation in physiological systems: Problems, methods, and case studies. *Automatica*, 33(5):851–870, 1997.

[37] C.E. de Souza. Robust $\mathcal{H}_\infty$ filtering for a class of discrete-time lipschitz nonlinear systems. *Automatica*, 103:69–80, 2019.

[38] C.E. De Souza, K.A. Barbosa, and A.T. Neto. Robust $\mathcal{H}_\infty$ filtering for discrete-time linear systems with uncertain time-varying parameters. *IEEE Transactions on Signal Processing*, 54(6):2110–2118, 2006.

[39] S.S Delshad, A. Johansson, M. Darouach, and T. Gustafsson. Robust state estimation and unknown inputs reconstruction for a class of nonlinear systems: Multiobjective approach. *Automatica*, 64:1–7, 2016.

[40] J. Dutta. Generalized derivatives and non-smooth optimization, a finite dimensional tour. *Top*, 13(2):185–279, 2005.

[41] D. Efimov, T. Raïssi, S. Chebotarev, and A. Zolghadri. Interval state observer for nonlinear time varying systems. *Automatica*, 49(1):200–205, 2013.

[42] N. Ellero, D. Gucik-Derigny, and D. Henry. An unknown input interval observer for LPV systems under L_2-gain and L_∞-gain criteria. *Automatica*, 103:294–301, 2019.

[43] G.A. Enciso, H.L. Smith, and E.D. Sontag. Nonmonotone systems decomposable into monotone systems with negative feedback. *Journal of Differential Equations*, 224(1):205–227, 2006.

[44] J.P. Farwell and R. Rohozinski. Stuxnet and the future of cyber war. *Survival*, 53(1):23–40, 2011.

[45] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, June 2014.

[46] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica*, 43(1):111–116, January 2007.

[47] S. Gillijns and B. De Moor. Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica*, 43(5):934–937, 2007.

[48] A. Girard and G.C. Le. Efficient reachability analysis for linear systems using support functions. *IFAC Proceedings Volumes*, 41(2):8966–8971, 2008.

[49] J.L. Gouzé and K.P. Hadeler. Monotone flows and order intervals. *Nonlinear World*, 1(1):23–34, 1994.

[50] J.F. Grcar. A matrix lower bound. *Linear Algebra and its Applications*, 433(1):203–220, 2010.

[51] Q.P. Ha and H. Trinh. State and input simultaneous estimation for a class of nonlinear systems. *Automatica*, 40(10):1779–1785, 2004.

[52] M.W. Hirsch and H. Smith. Monotone dynamical systems. In *Handbook of differential equations: ordinary differential equations*, volume 2, pages 239–357. Elsevier, 2006.

[53] M. James. The generalised inverse. *The Mathematical Gazette*, 62(420):109–114, 1978.

[54] L. Jaulin. Nonlinear bounded-error state estimation of continuous-time systems. *Automatica*, 38(6):1079–1082, 2002.

[55] L. Jaulin. A nonlinear set membership approach for the localization and map building of underwater robots. *IEEE Transactions on Robotics*, 25(1):88–98, 2009.

[56] L. Jaulin. Inner and outer set-membership state estimation. *Reliable Computing*, 22:47–55, 2016.

[57] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied interval analysis. 2001. *ed: Springer, London*, 2001.

[58] L. Jetto, S. Longhi, and G. Venturini. Development and experimental validation of an adaptive extended kalman filter for the localization of mobile robots. *IEEE Transactions on Robotics and Automation*, 15(2):219–229, 1999.

[59] Z. Jin, **Khajenejad, M.**, and S.Z. Yong. Data-driven model invalidation for unknown lipschitz continuous systems via abstraction. In *American Control Conference (ACC)*, pages 2975–2980. IEEE, 2020.

[60] M. Khajenejad and S.Z Yong. Simultaneous mode, input and state set-valued observers with applications to resilient estimation against sparse attacks. In *IEEE Conference on Decision and Control (CDC), Accepted*, 2019.

[61] M. Khajenejad and S.Z. Yong. Simultaneous input and state interval observers for nonlinear systems with full-rank direct feedthrough. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5443–5448. IEEE, 2020.

[62] M. Khajenejad and S.Z. Yong. Simultaneous input and state interval observers for nonlinear systems with rank-deficient direct feedthrough. In *European Control Conference, accepted*, 2021.

[63] M. Khajenejad and S.Z. Yong. Tight remainder-form decomposition functions with applications to constrained reachability and interval observer design. *IEEE Transactions on Automatic Control, submitted, https://arxiv.org/pdf/2103.08638.pdf*, 2021.

[64] H.K. Khalil. Nonlinear systems. *Upper Saddle River*, 2002.

[65] M. Kieffer, L. Jaulin, and E. Walter. Guaranteed recursive non-linear state bounding using interval analysis. *International Journal of Adaptive Control and Signal Processing*, 16(3):193–218, 2002.

[66] M. Kieffer and E. Walter. Guaranteed nonlinear state estimator for cooperative systems. *Numerical algorithms*, 37(1-4):187–198, 2004.

[67] H. Kim, P. Guo, M. Zhu, and P. Liu. Attack-resilient estimation of switched nonlinear cyber-physical systems. In *American Control Conference (ACC)*, pages 4328–4333. IEEE, 2017.

[68] P.K. Kitanidis. Unbiased minimum-variance linear state estimation. *Automatica*, 23(6):775–778, November 1987.

[69] J. Korbicz, M. Witczak, and V. Puig. LMI-based strategies for designing observers and unknown input observers for non-linear discrete-time systems. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 2007.

[70] M. Kulenović and O. Merino. A global attractivity result for maps with invariant boxes. *Discrete & Continuous Dynamical Systems-B*, 6(1):97, 2006.

[71] V.T.H. Le, C. Stoica, T. Alamo, E.F. Camacho, and D. Dumur. Zonotopic guaranteed state estimation for uncertain systems. *Automatica*, 49(11):3418–3424, 2013.

[72] W. Liu and I. Hwang. Robust estimation and fault detection and isolation algorithms for stochastic linear hybrid systems with unknown fault input. *IET control theory & applications*, 5(12):1353–1368, 2011.

[73] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[74] P. Lu, E-J. Van Kampen, C.C. De Visser, and Q. Chu. Framework for state and unknown input estimation of linear time-varying systems. *Automatica*, 73:145–154, 2016.

[75] T.T. Lu and S.H. Shiou. Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications*, 43(1-2):119–129, 2002.

[76] F. Mazenc and O. Bernard. Interval observers for linear time-invariant systems with disturbances. *Automatica*, 47(1):140–147, 2011.

[77] F. Mazenc, T-N. Dinh, and S-I. Niculescu. Robust interval observers and stabilization design for discrete-time systems with input and output. *Automatica*, 49(11):3490–3497, 2013.

[78] F. Mazenc, T.N. Dinh, and S.I. Niculescu. Interval observers for discrete-time systems. *International journal of robust and nonlinear control*, 24(17):2867–2890, 2014.

[79] P.J. Meyer, A. Devonport, and M. Arcak. Tira: Toolbox for interval reachability analysis. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pages 224–229, 2019.

[80] P.J. Meyer and D.V. Dimarogonas. Hierarchical decomposition of ltl synthesis problem for nonlinear control systems. *IEEE Transactions on Automatic Control*, 64(11):4676–4683, 2019.

[81] M. Milanese and C. Novara. Set membership identification of nonlinear systems. *Automatica*, 40:957–975, 2004.

[82] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009, 1991.

[83] M. Moisan, O. Bernard, and J-L. Gouzé. Near optimal interval observers bundle for uncertain bioreactors. In *European Control Conference (ECC)*, pages 5115–5122. IEEE, 2007.

[84] R.E. Moore, R.B. Kearfott, and M.J. Cloud. *Introduction to interval analysis*. SIAM, 2009.

[85] A. Murch and J. Foster. Recent nasa research on aerodynamic modeling of post-stall and spin dynamics of large transport airplanes. In *45th AIAA aerospace sciences meeting and exhibit*, page 463, 2007.

[86] Y. Nakahira and Y. Mo. Dynamic state estimation in the presence of compromised sensory data. In *IEEE Conference on Decision and Control (CDC)*, pages 5808–5813, 2015.

[87] C.H. Nien and F.J Wicklin. An algorithm for the computation of preimages in noninvertible mappings. *International Journal of Bifurcation and Chaos*, 8(02):415–422, 1998.

[88] R.C. Oliveira and P.L. Peres. Robust stability analysis and control design for time-varying discrete-time polytopic systems with bounded parameter variation. In *American Control Conference*, pages 3094–3099. IEEE, 2008.

[89] M. Pajic, P. Tabuada, I. Lee, and G.J. Pappas. Attack-resilient state estimation in the presence of noise. In *IEEE Conference on Decision and Control (CDC)*, pages 5827–5832, 2015.

[90] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, November 2013.

[91] R. Patton, R. Clark, and P.M. Frank. *Fault diagnosis in dynamic systems: theory and applications*. Prentice Hall, 1989.

[92] M.A. Peters and P.A. Iglesias. A spectral test for observability and reachability of time-varying systems. *SIAM Journal on Control Optimization*, 37(5):1330–1345, August 1999.

[93] B.T. Polyak, S.A. Nazin, Cé. Durieu, and E. Walter. Ellipsoidal parameter or state estimation under model uncertainty. *Automatica*, 40(7):1171–1179, 2004.

[94] Dimitrios Pylorof, Efstathios Bakolas, and Kevin S Chan. Design of robust lyapunov-based observers for nonlinear systems with sum-of-squares programming. *IEEE Control Systems Letters*, 4(2):283–288, 2019.

[95] S. Raghavan and J.K. Hedrick. Observer design for a class of nonlinear systems. *International Journal of Control*, 59(2):515–528, 1994.

[96] T. Raïssi, D. Efimov, and A. Zolghadri. Interval state estimation for a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 57(1):260–265, 2011.

[97] T. Raïssi, G. Videau, and A. Zolghadri. Interval observer design for consistency checks of nonlinear continuous-time systems. *Automatica*, 46(3):518–527, 2010.

[98] H. Ratschek and J. Rokne. *Computer methods for the range of functions*. Horwood, 1984.

[99] B.S. Rego, G.V. Raffo, J.K. Scott, and D.M Raimondo. Guaranteed methods based on constrained zonotopes for set-valued state estimation of nonlinear discrete-time systems. *Automatica*, 111:108614, 2020.

[100] G. Richards. Hackers vs slackers. *Engineering Technology*, 3(19):40–43, November 2008.

[101] R.T. Rockafellar. *Convex analysis*. Princeton university press, 2015.

[102] J.K. Scott and P.I. Barton. Bounds on the reachable sets of nonlinear control systems. *Automatica*, 49(1):93–100, 2013.

[103] J.K. Scott, D.M. Raimondo, G.R. Marseglia, and R.D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, 2016.

[104] J.S. Shamma and K. Tu. Set-valued observers and optimal disturbance rejection. *IEEE Trans. on Automatic Control*, 44(2):253–264, 1999.

[105] K. Shen and J.K. Scott. Rapid and accurate reachability analysis for nonlinear dynamic systems by exploiting model redundancy. *Computers & Chemical Engineering*, 106:596–608, 2017.

[106] S. Sheng and M. Hsiao. Efficient preimage computation using a novel success-driven atpg. In *2003 Design, Automation and Test in Europe Conference and Exhibition*, pages 822–827. IEEE, 2003.

[107] Y. Shoukry, P. Nuzzo, A. Puggelli, A.L. Sangiovanni-Vincentelli, S.A. Seshia, M. Srivastava, and P. Tabuada. Imhotep-SMT: A satisfiability modulo theory solver for secure state estimation. In *13th International Workshop on Satisfiability Modulo Theories (SMT)*, pages 3–13, 2015.

[108] Kanishka Raj Singh, Qiang Shen, and Sze Zheng Yong. Mesh-based affine abstraction of nonlinear systems with tighter bounds. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 3056–3061. IEEE, 2018.

[109] J. Slay and M. Miller. Lessons learned from the Maroochy water breach. In *International Conference on Critical Infrastructure Protection*, pages 73–82. Springer, 2007.

[110] H.L. Smith. Global stability for mixed monotone systems. *Journal of Difference Equations and Applications*, 14(10-11):1159–1164, 2008.

[111] B.L. Stevens, F.L. Lewis, and E.N. Johnson. *Aircraft control and simulation: dynamics, controls design, and autonomous systems*. John Wiley & Sons, 2015.

[112] E. Summers, A. Chakraborty, W. Tan, U. Topcu, P. Seiler, G. Balas, and A. Packard. Quantitative local l2-gain and reachability analysis for nonlinear systems. *International Journal of Robust and Nonlinear Control*, 23(10):1115–1135, 2013.

[113] **Khajenejad, M.**, Z. Jin, and S.Z. Yong. Interval observers for simultaneous state and model estimation of partially known nonlinear systems. In *2021 American Control Conference (ACC)*, pages 2848–2854. IEEE, 2021.

[114] **Khajenejad, M.**, F. Shoaib, and S.Z. Yong. Set-valued state estimation of discrete-time nonlinear systems via mixed-monotone decomposition. In *2021 IEEE 60th Conference on Decision and Control (CDC), accepted.* IEEE, 2021.

[115] **Khajenejad, M.** and S.Z. Yong. Simultaneous input and state set-valued $\mathcal{H}_\infty$-observers for linear parameter-varying systems. In *American Control Conference (ACC)*, pages 4521–4526, 2019.

[116] **Khajenejad, M**. and S.Z. Yong. Simultaneous mode, input and state set-valued observers with applications to resilient estimation against sparse attacks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1544–1550. IEEE, 2019.

[117] **Khajenejad, M**. and S.Z. Yong. Simultaneous input and state interval observers for nonlinear systems with full-rank direct feedthrough. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5443–5448. IEEE, 2020.

[118] **Khajenejad, M**. and S.Z. Yong. Simultaneous state and unknown input set-valued observers for nonlinear dynamical systems. *arXiv preprint arXiv:2001.10125, Submitted to Automatica, under review*, 2020.

[119] **Khajenejad, M**. and S.Z. Yong. Tight remainder-form decomposition functions (with applications to constrained reachability and interval observer design). *In preparation for IEEE Transaction on Automatic Control*, 2020.

[120] **Khajenejad, M**. and S.Z. Yong. Simultaneous input and state interval observers for nonlinear systems with rank-deficient direct feedthrough. *European Control Conference (ECC), accepted, arXiv preprint arXiv:2004.01861*, 2021.

[121] **Khajenejad, M**. and S.Z. Yong. Simultaneous mode, state and input set-valued observers for switched nonlinear systems. *arXiv preprint arXiv:2102.10793*, 2021.

[122] **Khajenejad, M**., Jin Z., and S.Z. Yong. Set-valued state and unknown terrain estimation for planetary rovers. *Advanced Intelligent Systems, (accepted)*, 2021.

[123] K.C. Veluvolu and Y.C. Soh. Discrete-time sliding-mode state and unknown input estimations for nonlinear systems. *IEEE Transactions on Industrial Electronics*, 56(9):3443–3452, 2008.

[124] A. Vicino and G. Zappa. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control*, 41(6):774–785, 1996.

[125] Y. Wang, D-M. Bevly, and R. Rajamani. Interval observer design for LPV systems with parametric uncertainty. *Automatica*, 60:79–85, 2015.

[126] Y. Wang, L. Xie, and C.E. De Souza. Robust control of a class of uncertain nonlinear systems. *Systems & Control Letters*, 19(2):139–149, 1992.

[127] L. Yang, A. Karnik, B. Pence, M.T.B Waez, and N. Ozay. Fuel cell thermal management: Modeling, specifications, and correct-by-construction control synthesis. *IEEE Transactions on Control Systems Technology*, 2019.

[128] L. Yang, O. Mickelin, and N. Ozay. On sufficient conditions for mixed monotonicity. *IEEE Transactions on Automatic Control*, 64(12):5080–5085, 2019.

[129] L. Yang and N. Ozay. Tight decomposition functions for mixed monotonicity. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 5318–5322. IEEE, 2019.

[130] X. Yang and J. Scott. Accurate uncertainty propagation for discrete-time nonlinear systems using differential inequalities with model redundancy. *IEEE Transactions on Automatic Control*, 2020.

[131] S.Z. Yong. Simultaneous input and state set-valued observers with applications to attack-resilient estimation. In *American Control Conference (ACC)*, pages 5167–5174. IEEE, 2018.

[132] S.Z. Yong, M.Q. Foo, and E. Frazzoli. Robust and resilient estimation for cyber-physical systems under adversarial attacks. In *American Control Conference (ACC)*, pages 308–315. IEEE, 2016.

[133] S.Z. Yong, M. Zhu, and E. Frazzoli. On strong detectability and simultaneous input and state estimation with a delay. In *IEEE Conference on Decision and Control (CDC)*, pages 468–475, 2015.

[134] S.Z. Yong, M. Zhu, and E. Frazzoli. Resilient state estimation against switching attacks on stochastic cyber-physical systems. In *IEEE Conference on Decision and Control (CDC)*, pages 5162–5169, 2015.

[135] S.Z. Yong, M. Zhu, and E. Frazzoli. A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automatica*, 63:321–329, 2016.

[136] S.Z. Yong, M. Zhu, and E. Frazzoli. Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation, and attack mitigation. *ACM Transactions on Cyber-Physical Systems*, 2(2):9, 2018.

[137] S.Z. Yong, M. Zhu, and E. Frazzoli. Switching and data injection attacks on stochastic cyber-physical systems: Modeling, resilient estimation, and attack mitigation. *ACM Transactions on Cyber-Physical Systems*, 2(2):9, 2018.

[138] Z.B Zabinsky, R.L Smith, and B.P Kristinsdottir. Optimal estimation of univariate black-box Lipschitz functions with upper and lower error bounds. *Computers & Operations Res.*, 30(10):1539–1553, 2003.

[139] K. Zetter. Inside the cunning, unprecedented hack of Ukraine's power grid. Wired Magazine, 2016.

[140] G. Zheng, D. Efimov, and W. Perruquetti. Design of interval observer for a class of uncertain unobservable nonlinear systems. *Automatica*, 63:167–174, 2016.

[141] Q. Zheng, S. Xu, and Z. Zhang. Asynchronous non-fragile $\mathcal{H}_\infty$ filtering for discrete-time nonlinear switched systems with quantization. *Nonlinear Analysis: Hybrid Systems*, 37:100911, 2020.

APPENDIX A

PROOFS

Proof of Lemma 2.3.2

(2.6)-(2.10) and plugging $M_1 = \Sigma^{-1}$ into (2.7) imply that

$$\begin{aligned}
\hat{d}_{1,k} &= M_1(C_1\tilde{x}_{k|k} + \Sigma d_{1,k} + \textstyle\sum_{i=1}^{N}\lambda_{i,k}v^i_{1,k}), && \text{(A.1)}\\
\hat{d}_{2,k-1} &= M_2(C_2(\textstyle\sum_{i=1}^{N}\lambda_{i,k-1}A^i\tilde{x}_{k-1|k-1} + G_1\tilde{d}_{1,k-1} \\
&\quad + G_2 d_{2,k-1} + \textstyle\sum_{i=1}^{N}\lambda_{i,k-1}w^i_{k-1}) + \textstyle\sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k}). && \text{(A.2)}\\
\tilde{d}_{1,k} &= d_{1,k} - \hat{d}_{1,k} = -M_1(C_1\tilde{x}_{k|k} + \textstyle\sum_{i=1}^{N}\lambda_{i,k}v^i_{1,k}). && \text{(A.3)}
\end{aligned}$$

(A.3) and setting $M_2 = (C_2G_2)^\dagger$ (Lemma 2.3.1) in (A.2), return

$$\begin{aligned}
\tilde{d}_{2,k-1} &= -M_2(C_2\hat{A}_{k-1}\tilde{x}_{k-1|k-1} - C_2G_1M_1\textstyle\sum_{i=1}^{N}\lambda_{i,k-1}v^i_{1,k-1} \\
&\quad +C_2\textstyle\sum_{i=1}^{N}\lambda_{i,k-1}w^i_{k-1} + \textstyle\sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k}).
\end{aligned} \tag{A.4}$$

Defining $\tilde{x}^\star_{k|k} \triangleq x_k - \hat{x}^\star_{k|k}$, from (2.1), (2.10) and (2.11) we obtain

$$\tilde{x}^\star_{k|k} = \textstyle\sum_{i=1}^{N}\lambda_{i,k-1}(A^i\tilde{x}_{k-1|k-1} + w^i_{k-1}) + G_1\tilde{d}_{1,k-1} + G_2\tilde{d}_{2,k-1} \tag{A.5}$$

In addition, from (2.6) and (2.12) and (A.3)-(A.5) we conclude:

$$\begin{aligned}
\tilde{x}_{k|k} &= (I - \tilde{L}C_2)\tilde{x}^\star_{k|k} - \tilde{L}\textstyle\sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k}. && \text{(A.6)}\\
\tilde{x}^\star_{k|k} &= \overline{A}_{k-1}\tilde{x}_{k-1|k-1} - (I - G_2M_2C_2)(G_1M_1\textstyle\sum_{i=1}^{N}\lambda_{i,k-1}(v^i_{1,k-1} - w^i_{k-1})) \\
&\quad - G_2M_2\sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k}. && \text{(A.7)}
\end{aligned}$$

Now, defining $\overline{w}_{k-1} \triangleq -G_2M_2\sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k} - (I - G_2M_2C_2)(G_1M_1\sum_{i=1}^{N}\lambda_{i,k-1}(v^i_{1,k-1} - w^i_{k-1}))$ and $\overline{v}_{k-1} \triangleq \sum_{i=1}^{N}\lambda_{i,k}v^i_{2,k}$, (A.6)-(A.7) imply that

$$\begin{aligned}
\tilde{x}^\star_{k|k} &= \overline{A}_{k-1}\tilde{x}_{k-1|k-1} + \overline{w}_{k-1},\\
\tilde{x}_{k|k} &= (I - \tilde{L}C_2)\overline{A}_{k-1}\tilde{x}_{k-1|k-1} + (I - \tilde{L}C_2)\overline{w}_{k-1} - \tilde{L}\overline{v}_{k-1}.
\end{aligned} \tag{A.8}$$

Now, consider the following linear time-varying system:

$$x_{k+1} = \overline{A}_k x_k + \overline{w}_k, y_k = C_2x_k + \overline{v}_k. \tag{A.9}$$

Systems (A.8) and (A.9) are equivalent from the viewpoint of estimation, since the estimation error equations for both problems are the same, hence they both have the same objective. Therefore, the pair $(\overline{A}_k, C_2)$ needs to be uniformly detectable such that the observer is stable [10, Section 5]. □

Proof of Theorem 2.3.3

Starting from (A.8), we have

$$\tilde{x}_{k|k} = (I - \tilde{L}C_2)\overline{A}_{k-1}\tilde{x}_{k-1|k-1} + (I - \tilde{L}C_2)\overline{w}_{k-1} - \tilde{L}\overline{v}_{k-1},$$

from which we can define a system with $\tilde{x}_{k|k}$ as its state and $\tilde{z}_{k|k} = \tilde{x}_{k|k}$ as the output:

$$\begin{aligned} \tilde{x}_{k|k} &= (I - \tilde{L}C_2)\overline{A}_{k-1}\tilde{x}_{k-1|k-1} + \begin{bmatrix} I - \tilde{L}C_2 & -\tilde{L} \end{bmatrix} \begin{bmatrix} \overline{w}_{k-1} \\ \overline{v}_{k-1} \end{bmatrix}, \\ \tilde{z}_{k|k} &= \tilde{x}_{k|k}. \end{aligned}$$

By [38, Lemma 3], this system has an $\mathcal{H}_\infty$ performance bounded by η, if there exists a symmetric positive definite matrix P with rank n such that:

$$\begin{bmatrix} P & (I - \tilde{L}C_2)\overline{A^i}P & \begin{bmatrix} I - \tilde{L}C_2 & -\tilde{L} \end{bmatrix} & 0 \\ * & P & 0 & P \\ * & * & \eta I & 0 \\ * & * & * & \eta I \end{bmatrix} \succ 0, \forall i \in \{1, 2, \ldots, N\}. \tag{A.10}$$

Notice that the referenced lemma requires the existence of a *bounded matrix sequence*, which in our case is a sequence of time-invariant matrices (P is the same for each k), that is obviously bounded. By plugging $S = P^{-1} \succ 0$ and applying some similarity transformations, we obtain

$$\begin{aligned} &\begin{bmatrix} 0 & S & 0 & 0 \\ * & 0 & 0 & 0 \\ * & * & I & 0 \\ * & * & * & I \end{bmatrix} \begin{bmatrix} P & (I - \tilde{L}C_2)\overline{A^i}P & \begin{bmatrix} I - \tilde{L}C_2 & -\tilde{L} \end{bmatrix} & 0 \\ * & P & 0 & P \\ * & * & \eta I & 0 \\ * & * & * & \eta I \end{bmatrix} \begin{bmatrix} 0 & S & 0 & 0 \\ * & 0 & 0 & 0 \\ * & * & I & 0 \\ * & * & * & I \end{bmatrix} \\ &= \begin{bmatrix} S & \overline{A^i}^\top (I - C_2^\top \tilde{L}^\top)S & 0 & I \\ * & S & \begin{bmatrix} I - \tilde{L}C_2 & -\tilde{L} \end{bmatrix} & 0 \\ * & * & I & 0 \\ * & * & * & \eta I \end{bmatrix} \succ 0 \ \forall i \in \{1, 2, \ldots, N\}. \end{aligned}$$

Setting $Y \triangleq S\tilde{L}$ completes the proof. □

Proof of Theorem 2.3.4

Suppose, for contradiction, that there exists an $\mathcal{H}_\infty$-observer for system (2.1) with any sequence $\{\lambda_{i,k}\}_{k=0}^{\infty}$ for all $i \in \{1, 2, \ldots, N\}$ that satisfies $0 \leq \lambda_{i,k} \leq 1, \sum_{i=1}^{N} \lambda_{i,k} = 1, \forall k$, but one of the constituent linear time-invariant systems (e.g., (A^j, G, C, H)) is not strongly detectable. Since the $\mathcal{H}_\infty$-observer exists for any sequence of $\lambda_{i,k}$, particularly it exists when $\lambda_{j,k} = 1$ and $\lambda_{ij,k} = 0$, $\forall i \neq j$ for all k. However, we know from [131] that strong detectability is necessary for the stability of the linear time-invariant system (A^j, G, C, H), which is a contradiction. □

Proof of Theorem 2.3.5

To prove Theorem 2.3.5, we first find closed form expressions for the state and input estimation errors in the following:

Lemma A.0.1. *The state and input estimation errors are*

$$\begin{aligned}\tilde{x}_{k|k} &= (\textstyle\prod_{j=0}^{k-1} A_{e,k-j})\tilde{x}_{0|0} + \textstyle\sum_{i=1}^{k}(\prod_{j=0}^{i-2} A_{e,k-j})(\Psi\overline{w}_{k-i} - \tilde{L}\overline{v}_{k-i}),\\ \tilde{d}_{k-1} &= \textstyle\sum_{i=1}^{N} \lambda_{i,k-1}(-V_1M_1C_1 - V_2M_2C_2A_{e,i})\tilde{x}_{k-1|k-1}\\ &\quad + (-V_1M_1 + V_2M_2C_2G_1M_1)T_1 \textstyle\sum_{i=1}^{N} \lambda_{i,k-1}v_{k-1}^{i}\\ &\quad - V_2M_2C_2 \textstyle\sum_{i=1}^{N} \lambda_{i,k-1}w_{k-1}^{i} - V_2M_2T_2 \textstyle\sum_{i=1}^{N} \lambda_{i,k}v_{k}^{i}.\end{aligned}$$

Proof. From (A.8), we have

$$\tilde{x}_{k|k} = \Psi\overline{A}_k\tilde{x}_{k-1|k-1} + \Psi\overline{w}_{k-1} - \tilde{L}\overline{v}_{k-1}. \tag{A.11}$$

We use induction and (A.11) to obtain

$$\begin{aligned}\tilde{x}_{1|1} &= \Psi\overline{A}_1\tilde{x}_{0|0} + \Psi\overline{w}_0 - \tilde{L}\overline{v}_0 = A_{e,1}\overline{x}_{0|0} + \Psi\overline{w}_{1-1} - \tilde{L}\overline{v}_{1-1}\\ \tilde{x}_{k|k} &= (\prod_{j=0}^{k-1} A_{e,k-1})\tilde{x}_{0|0} + \sum_{i=1}^{k}(\prod_{j=0}^{i-2} A_{e,1-j})(\Psi\overline{w}_{k-1} - \tilde{L}\overline{v}_{1-i})\tilde{x}_{k+1|k+1}\\ &= \Psi\overline{A}_{k+1}\tilde{x}_{k|k} + \Psi\overline{w}_k - \tilde{L}\overline{v}_k\\ &= \Psi\overline{A}_{k+1}\big[(\prod_{j=0}^{k-1}A_{e,k-j})\tilde{x}_{0|0} + \sum_{i=1}^{k}(\prod_{j=0}^{i-2} A_{e,k-j})(\Psi\overline{w}_{k-i} - \tilde{L}\overline{v}_{k-i})\big] + \Psi\overline{w}_k - \tilde{L}\overline{v}_k\\ &= (A_{e,k+1}A_{e,k}...A_{e,1})\tilde{x}_{0|0} + \Psi\overline{w}_k - \tilde{L}\overline{v}_k\\ &+ \sum_{i=1}^{k}(A_{e,k+1}A_{e,k}...A_{e,k-(i-2)})(\Psi\overline{w}_{k-i} - \tilde{L}\overline{v}_{k-i}))\\ &= (\prod_{j=0}^{k+1} A_{e,k+1-j})\tilde{x}_{0|0} + \textstyle\sum_{i=0}^{k}(\displaystyle\prod_{j=0}^{i-2} A_{e,k-j})(\Psi\overline{w}_{k-i} - \tilde{L}\overline{v}_{k-i})\\ &= (\prod_{j=0}^{k+1} A_{e,k+1-j})\tilde{x}_{0|0} + \textstyle\sum_{i=1}^{k+1}(\displaystyle\prod_{j=0}^{i-2} A_{e,k+1-j})(\Psi\overline{w}_{k+1-i} - \tilde{L}\overline{v}_{k+1-i}).\end{aligned}$$

As for $\tilde{d}_{k-1}$, (A.3)-(A.4) imply

$$\begin{aligned}\tilde{d}_{k-1} &= V_1\tilde{d}_{1,k-1} + V_2\tilde{d}_{2,k-1} = \sum_{i=1}^{N} \lambda_{i,k-1}(-V_1M_1C_1 - V_2M_2C_2A_{e,i})\tilde{x}_{k-1|k-1}\\ &+(V_2M_2C_2G_1M_1 - V_1M_1)T_1 \sum_{i=1}^{N} \lambda_{i,k-1}v_{k-1}^{i}\\ &-V_2M_2C_2 \sum_{i=1}^{N} \lambda_{i,k-1}w_{k-1}^{i} - V_2M_2T_2 \sum_{i=1}^{N} \lambda_{i,k}v_{k}^{i}.\end{aligned} \tag{A.12}$$

□

Now, we are ready to prove Theorem 2.3.5. First, we define

$$B_{e,k} \triangleq \prod_{j=0}^{k-1} A_{e,k-j}, C_{e,k}^{i} \triangleq \prod_{j=0}^{i-2} A_{e,k-j}, \bar{t}_k \triangleq \Psi\overline{w}_k - \tilde{L}\overline{v}_k \tag{A.13}$$

for $1 \le i \le k$. Then, from Lemma A.0.3, it follows that

$$\|\tilde{x}_{k|k}\| = \|B_{e,k}\tilde{x}_{0|0} + \sum_{i=1}^{k} C^i_{e,k}\bar{t}_{k-i}\| \le \|B_{e,k}\|\|\tilde{x}_{0|0}\| + \|\sum_{i=1}^{k} C^i_{e,k}\bar{t}_{k-i}\|. \tag{A.14}$$

Moreover, by similar reasoning, we obtain:

$$\begin{aligned}
\|B_{e,k}\| &= \|\prod_{j=0}^{k-1} A_{e,k-j}\| \le \prod_{j=0}^{k-1} \|A_{e,k-j}\| = \prod_{j=0}^{k-1} \|\Psi\Phi\hat{A}_{k-j}\| \\
&= \prod_{j=0}^{k-1} \|\Psi\Phi \sum_{i=1}^{N} \lambda^i_{k-j}(A^i - G_1 M_1 C_1)\| = \prod_{j=0}^{k-1} \|\sum_{i=1}^{N} \lambda^i_{k-j}\Psi\Phi(A^i - G_1 M_1 C_1)\| \\
&\le \prod_{j=0}^{k-1} \sum_{i=1}^{N} \lambda^i_{k-j}\|\Psi\Phi(A^i - G_1 M_1 C_1)\| \le \prod_{j=0}^{k-1} \theta = \theta^k,
\end{aligned} \tag{A.15}$$

$$\|\sum_{i=1}^{k} C^i_{e,k}\bar{t}_{k-i}\| \le \sum_{i=1}^{k} \|C^i_{e,k}\bar{t}_{k-i}\| \le \sum_{i=1}^{k} \|C^i_{e,k}\|\|\bar{t}_{k-i}\|, \tag{A.16}$$

$$\|C^i_{e,k}\| = \|\prod_{j=0}^{i-2} A_{e,k-j}\| \le \prod_{j=0}^{i-2} \|A_{e,k-j}\| = \prod_{j=0}^{i-2} \|\sum_{s=1}^{N} \lambda_{s,k-j}A_{e,s}\| \le \prod_{j=0}^{i-2} \theta \le \theta^{i-1}. \tag{A.17}$$

Furthermore, from the definition of $\overline{w}_k$ and (A.13) we have

$$\begin{aligned}
\overline{w}_{k-i} &= -\Phi(G_1 M_1 \sum_{s=1}^{N} \lambda_{s,k-i}v^s_{1,k-i} - \sum_{s=1}^{N} \lambda_{s,k-i}w^s_{k-i}) - G_2 M_2 \sum_{s=1}^{N} \lambda_{s,k-i}v^s_{2,k-i}, \\
\|\bar{t}_{k-i}\| &= \|\Psi\overline{w}_{k-i} - \tilde{L}\overline{v}_{k-i}\| \\
&= \| - \Psi\Phi G_1 M_1 T_1 \sum_{s=1}^{N} \lambda_{s,k-i}v^s_{k-i} + \Psi\Phi \sum_{s=1}^{N} \lambda_{s,k-i}w^s_{k-i} \\
&\quad - \Psi G_2 M_2 T_2 \textstyle\sum_{s=1}^{N} \lambda_{s,k-i}v^s_{k-i} - \tilde{L}T_2 \displaystyle\sum_{s=1}^{N} v^s_{k-i}\| \\
&= \|\sum_{s=1}^{N} \lambda_{s,k-i}(\Gamma v^s_{k-i} + (\Psi\Phi)w^s_{k-i})\| \le \overline{\eta},
\end{aligned}$$

from which, as well as (A.14)-(A.17), we conclude that

$$\|\tilde{x}_{k|k}\| \le \|\tilde{x}_{0|0}\|\theta^k + \overline{\eta}\sum_{i=1}^{k} \theta^{i-1} = \|\tilde{x}_{0|0}\|\theta^k + \overline{\eta}\tfrac{1-\theta^k}{1-\theta} \triangleq \delta^x_k. \tag{A.18}$$

As for δ^d_{k-1}, using Lemma A.0.3 and (A.12), triangle inequality and the facts that $0 \le \lambda_{i,k} \le 1, \sum_{i=1}^{N} \lambda_{i,k} = 1$ and submultiplicativity of matrix norms, we obtain the result. □

Proof of Theorem 2.3.6

Notice that $0 \le \|A_{e,i}\| \le \theta < 1$ for all $i \in \{1, 2, \ldots, N\}$ by assumption. So, θ^k in (A.18) vanishes in steady state, which gives us the following steady state estimation radius: $\lim_{k\to\infty} \delta^x_k = \lim_{k\to\infty}\left(\|\tilde{x}_{0|0}\|\theta^k + \overline{\eta}\frac{1-\theta^k}{1-\theta}\right) = \frac{\overline{\eta}}{1-\theta}$. Using this and starting from the expression for δ^d_{k-1} in Theorem 2.3.5, it converges to steady state, as follows: $\lim_{k\to\infty} \delta^d_{k-1} = (\lim_{k\to\infty} \beta\delta^x_{k-1}) + \|V_2 M_2 C_2\|\eta_w + (\|(V_2 M_2 C_2 G_1 - V_1)M_1 T_1\| + \|V_2 M_2 T_2\|)\eta_v = \frac{\overline{\eta}\beta}{1-\theta} + \eta_w\|V_2 M_2 C_2\| + \eta_v(\|V_2 M_2 T_2\| + \|R\|)$. $\square$

Proof of Lemma 3.3.4

To show (3.10), we first find a lower bound for $\delta^{q,inf}_{r,k}$. Then, we show that the lower bound diverges and so does $\delta^{q,inf}_{r,k}$. Define $\tilde{t}^\star_k \triangleq t^\star_k/\eta^t_k$, where η^t_k is defined in Corollary 3.2.8. Now consider

$$\eta^t_k \sigma_{min}(\mathbb{A}^q_k) = \sigma_{min}(\eta^t_k \mathbb{A}^q_k) = \min_{\|t\|_2 \le 1} \|\eta^t_k \mathbb{A}^q_k t\|_2 \le \|\eta^t_k \mathbb{A}^q_k \tilde{t}^\star_k\|_2 = \|\mathbb{A}^q_k t^\star_k\|_2 = \delta^{q,inf}_{r,k},$$

where $\sigma_{min}(A)$ is the least non-trivial singular value of matrix A, the first equality holds since $\sigma_{min}(.)$ is a linear operator, the second equality is a special case of a *matrix lower bound* [50] when 2-norms are considered, the inequality holds since $\|\tilde{t}^\star_k\|_2 = 1$ by Corollary 3.2.8, so $\tilde{t}^\star_k$ is a feasible point for the minimization in the third statement and the last equality holds by Theorem 3.2.7. So far we have shown that $\eta^t_k \sigma_{min}(\mathbb{A}^q_k)$ is a lower bound for $\delta^{q,inf}_{r,k}$. Next, we will prove that $\eta^t_k \sigma_{min}(\mathbb{A}^q_k)$ is unbounded. First, it is trivial that η^t_k is unbounded by its definition in Corollary 3.2.8. Second, consider the block matrix $\mathbb{A}^q_k$ in Lemma 3.2.5. By the strong detectability assumption, matrix A^q_e is stable [131, Theorem 3 and Appendix C], so all the block matrices of $\mathbb{A}^q_k$, except three of them which are constant matrices with respect to time, converge to zero matrices when time goes to infinity. Hence $\mathbb{A}^q_k$ converges to an infinite dimensional sparse matrix, with only three non-zero finite dimensional constant blocks and so the limit matrix has a finite rank and clearly has a bounded minimum non-trivial singular value. Henceforth, $\eta^t_k \sigma_{min}(\mathbb{A}^q_k)$ is unbounded, since the product of the bounded and non-zero $\sigma_{min}(\mathbb{A}^q_k)$ and unbounded η^t_k is unbounded. As for (3.11), the first equality holds by definition of $\hat{\delta}^q_{r,k}$ (cf. Theorem 3.2.7) and (3.10), the first inequality holds since $\delta^{q,tri}_{r,k} \le \overline{\delta}^{q,r}_{r,k}$ by triangle and sub-multiplicative inequalities and the last equality, i.e., convergence of $\delta^{q,tri}_{r,k}$, follows from strong detectability assumption which implies the stability of A^q_e [131, Theorem 3]. $\square$

Proof of Lemma 3.3.5

Suppose, for contradiction, that none of q and q' are eliminated. Then

$$\begin{aligned}
&\|C^q_2 \hat{x}^{\star,q}_{k|k} + D^q_2 u^q_k - C^{q'}_2 \hat{x}^{\star,q'}_{k|k} - D^{q'}_2 u^{q'}_k\|_2 = \|r^{q'}_k - r^q_k + z^q_{2,k} - z^{q'}_{2,k})\|_2 \\
&\le \|r^{q'}_k\|_2 + \|r^q_k\|_2 + \|z^q_{2,k} - z^{q'}_{2,k}\|_2 \le \delta^q_{r,k} + \delta^{q'}_{r,k} + R_y\|T^q_2 - T^{q'}_2\|_2,
\end{aligned}$$

where the equality holds by Definition 3.2.2, the first inequality holds by triangle inequality and the last inequality holds by the assumption that none of q and q' can be eliminated, as well as the boundedness assumption for the measurement space. This last inequality contradicts with the inequality in the lemma, thus the result holds. $\square$

Proof of Lemma 3.3.6

The result can be obtained by applying Proposition 3.2.3, (3.7) and the closed-form output signal:

$$y_k = \begin{bmatrix} \begin{bmatrix} (CA^k)^\top \\ (CA^{k-1})^\top \\ \vdots \\ C^\top \\ I \end{bmatrix}^\top & \begin{bmatrix} H^\top \\ (CG)^\top \\ (CAG)^\top \\ \vdots \\ (CA^{k-1}G)^\top \end{bmatrix}^\top & \begin{bmatrix} D^\top \\ (CB)^\top \\ (CAB)^\top \\ \vdots \\ (CA^{k-1}B)^\top \end{bmatrix}^\top \end{bmatrix} \begin{bmatrix} t_k \\ d^{q*}_{0:k} \\ u^{q*}_{0:k} \end{bmatrix},$$

which can be derived by using (3.1) and simple induction. □

Proof of Theorem 3.3.7

To show that (iii) is sufficient for asymptotic mode detectability, consider Lemma 3.3.5 with $\delta^{q,tri}_{r,k}$ as the upper bound. It suffices to show $\exists K \in \mathbb{N}$, such that (3.12) holds for $k \geq K, \forall q \neq q' \in \mathbb{Q}$. Notice that by Definition 3.2.2, $C^q_2 \hat{x}^{\star,q}_{k|k} = C^q_2 x_k + T^q_2 v_k - r^{q|*}_k$. Plugging this into (3.12), we need to show $\exists K \in \mathbb{N}$ such that:

$$\begin{aligned} &\|W^{q,q'} s^{q,q'}_k\|_2 > \delta^{q,tri}_{r,k} + \delta^{q',tri}_{r,k} + R^{q,q'}_z, \forall k \geq K, && \text{(A.19)} \\ &s^{q,q'}_k \triangleq \begin{bmatrix} x_k^\top & v_k^\top & r^{q|*\top}_k & r^{q'|*\top}_k & u^{q\top}_k & u^{q'\top}_k \end{bmatrix}^\top, \forall q \neq q' \in \mathbb{Q}. \end{aligned}$$

A sufficient condition to satisfy (A.19) is that $\exists K \in \mathbb{N}$ such that $\forall k \geq K$, (A.19) holds for all $s^{q,q'}_k$. Equivalently, it suffices

$$\begin{aligned} &\min_{x_k, v_k, r^q_k, r^{q'}_k} \|W^{q,q'} s^{q,q'}_k\|_2 > \delta^{q,tri}_{r,k} + \delta^{q',tri}_{r,k} + R^{q,q'}_z \\ &s.t.\ \|x_k\|_2 \leq R_x, \|v_k\|_2 \leq \eta_v, \|r^{q|*}_k\|_2 \leq \delta^{q,tri}_{r,k}, \|r^{q'|*}_k\|_2 \leq \delta^{q',tri}_{r,k},\ \forall k \geq K, \forall q \neq q' \in \mathbb{Q}. \end{aligned}$$

By expanding the constraint set, it is sufficient to require that $\exists K \in \mathbb{N}$ such that:

$$\begin{aligned} &\min_{s^{q,q'}_k} \|W^{q,q'} s^{q,q'}_k\|_2 > \delta^{q,tri}_{r,k} + \delta^{q',tri}_{r,k} + R^{q,q'}_z \\ &s.t.\ \|s^{q,q'}_k\|_2^2 \leq R_x^2 + \eta_v^2 + (\delta^{q,tri}_{r,k})^2 + (\delta^{q',tri}_{r,k})^2 + (u^q_k)^2 + (u^{q'}_k)^2,\ \forall k \geq K, \forall q \neq q' \in \mathbb{Q}. \end{aligned}$$

Now, by *matrix lower bound* theorem [50] and similar argument as in the proof of Lemma 3.3.4, it is sufficient to be satisfied that $\exists K \in \mathbb{N}\ s.t.\ \forall k \geq K, \forall q \neq q' \in \mathbb{Q}:$

$$\sigma^2_{min}(W^{q,q'}) > \frac{(\delta^{q,tri}_{r,k} + \delta^{q',tri}_{r,k} + R^{q,q'}_z)^2}{R_x^2 + \eta_v^2 + (\delta^{q,tri}_{r,k})^2 + (\delta^{q',tri}_{r,k})^2 + (u^q_k)^2 + (u^{q'}_k)^2}. \tag{A.20}$$

(A.20) provides us a *time-dependent* sufficient condition for mode detectability. In order to find a *time-independent* sufficient condition, notice that $\frac{(\overline{\delta}^{q,tri}_{r,k} + \overline{\delta}^{q',tri}_{r,k} + R^{q,q'}_z)^2}{R_x^2 + \eta_v^2}$ is an upper bound for the right hand side of (A.20), since the latter's denominator is smaller than the former's and the numerator of the latter is an upper bound signal for the former's by triangle and sub-multiplicative inequalities. So a sufficient condition for (A.20) is $\exists K \in \mathbb{N}$ *s.t.* $\forall k \geq K, \forall q \neq q' \in \mathbb{Q}$:

$$\sigma^2_{min}(W^{q,q'}) > \frac{(\overline{\delta}^{q,tri}_{r,k} + \overline{\delta}^{q',tri}_{r,k} + R^{q,q'}_z)^2}{R^2_x + \eta^2_v}. \tag{A.21}$$

Then, for the above to hold, it suffices that

$$\sigma^2_{min}(W^{q,q'}) > \lim_{k\to\infty} \frac{(\overline{\delta}^{q,tri}_{r,k} + \overline{\delta}^{q',tri}_{r,k} + R^{q,q'}_z)^2}{R^2_x + \eta^2_v},$$

which is equivalent to (iii) by (3.11). As for the sufficiency of (i), notice that by Theorems 3.2.4 and 3.2.7, Lemma 3.2.5 and Definition 3.3.1, for mode detectability, it suffices that for any specific mode q, the true mode q^* and large enough k,

$$\|r^q_k\|_2 = \| \begin{bmatrix} \mathbb{T}^{q,q^*}_k & \mathbb{B}^{q,q^*}_k & \mathbb{D}^{q,q^*}_k \end{bmatrix} \begin{bmatrix} t^\top_k & u^{q^*\top}_{0:k} & d^{q*\top}_{0:k} \end{bmatrix}^\top \|_2 > \delta^{q,tri}_{r,k},$$

with t_k given in (3.9). Since q^* is unknown, a sufficient condition to satisfy the above equality is $\forall q' \neq q \in Q$:

$$\|r^q_k\|_2 = \| \begin{bmatrix} \mathbb{T}^{q,q'}_k & \mathbb{B}^{q,q'}_k & \mathbb{D}^{q,q'}_k \end{bmatrix} \begin{bmatrix} t^\top_k & u^{q'\top}_{0:k} & d^{q*\top}_{0:k} \end{bmatrix}^\top \|_2 > \delta^{q,tri}_{r,k}.$$

So, it suffices that $\forall q' \neq q \in Q, \exists \overline{d} \in \mathbb{R}$, such that:

$$\begin{aligned}
&\min_{t'_k} \| \begin{bmatrix} \mathbb{T}^{q,q'}_k & \mathbb{B}^{q,q'}_k & \mathbb{D}^{q,q'}_k \end{bmatrix} t'_k \|_2 > \delta^{q,tri}_{r,k} \\
&s.t.\ t'_k = \begin{bmatrix} t^\top_k & u^{q'\top}_{0:k} & d^{q*\top}_{0:k} \end{bmatrix}^\top, t_k = \begin{bmatrix} \tilde{x}^\top_{0|0} & w^\top_0 & \dots & w^\top_{k-1} & v^\top_0 & \dots & v^\top_k \end{bmatrix}, \\
&\quad \|d^{q*}_{0:k}\|_2 \geq \overline{d}, \|\tilde{x}_{0|0}\|_\infty \leq \delta^x_0,\ \|w_i\|_\infty \leq \eta_w,\ \|v_j\|_\infty \leq \eta_v, \\
&\quad \forall i \in \{0, ..., k-1\},\ \forall j \in \{0, ..., k\}.
\end{aligned} \tag{A.22}$$

Again by matrix lower bound theorem, a sufficient condition for the above inequality to hold is that $\exists \overline{d} \in \mathbb{R}$, such that:

$$\begin{aligned}
&\min_{t_k, d_{0:k}} \|t'_k\|_2 > \frac{\delta^{q,tri}_{r,k}}{\sigma_{min} \begin{bmatrix} \mathbb{T}^{q,q'}_k & \mathbb{B}^{q,q'}_k & \mathbb{D}^{q,q'}_k \end{bmatrix}} \\
&s.t.\ t'_k = \begin{bmatrix} t^\top_k & u^{q'\top}_{0:k} & d^{q*\top}_{0:k} \end{bmatrix}^\top, \|d^{q*}_{0:k}\|_2 \geq \overline{d}, \\
&\quad t_k = \begin{bmatrix} \tilde{x}^\top_{0|0} & w^\top_0 & \dots & w^\top_{k-1} & v^\top_0 & \dots & v^\top_k \end{bmatrix}, \\
&\quad \|\tilde{x}_{0|0}\|_\infty \leq \delta^x_0,\ \|w_i\|_\infty \leq \eta_w,\ \|v_j\|_\infty \leq \eta_v, \\
&\quad \forall i \in \{0, ..., k-1\},\ \forall j \in \{0, ..., k\}.
\end{aligned} \tag{A.23}$$

Finally, since $\delta^{q,tri}_{r,k} \leq \overline{\delta}^{q,tri}_{r,k}$ and

$$\|t'_k\|_2 = \| \begin{bmatrix} t^\top_k & u^{q'\top}_{0:k} & d^{q*\top}_{0:k} \end{bmatrix} \|_2 \geq \sqrt{0^2 + 0^2 + \|d^{q*\top}_{0:k}\|^2_2} = \|d^{q*\top}_{0:k}\|_2,$$

then a sufficient condition for (A.23) is that

$$\|d_{0:k}^{q*\top}\|_2 > \frac{\overline{\delta}_{r,k}^{q,tri}}{\sigma_{min}(\begin{bmatrix}\mathbb{T}_k^{q,q'} & \mathbb{B}_k^{q,q'} & \mathbb{D}_k^{q,q'}\end{bmatrix})}. \tag{A.24}$$

Now suppose that $T_2^q \neq T_2^{q'}$ (otherwise the matrix in the denominator of (A.24) is zero and it never holds). Asymptotically speaking, the right hand side of (A.24) converges to $\tilde{\delta} \triangleq \max\{0, (\overline{\delta}_r^{q,tri}/\overline{\sigma}^{q,q'})\}$, since $\overline{\delta}_{r,k}^{q,tri}$ converges to $\overline{\delta}_r^{q,tri}$ and the least singular value in the denominator either diverges or converges to some steady value $\overline{\sigma}^{q,q'}$. So we set $\overline{d}$ equal to any real number strictly grater than $\tilde{\delta}$. By unlimited energy assumption for attack signal, after some large enough time step K, the monotone increasing function $\|d_{0:k}^{q*}\|_2$, exceeds $\overline{d}$ and so the system will be mode detectable. ☐

Proof of Proposition 4.1.7

The results follow from the facts that an inequality in $\mathbb{R}$ is preserved by multiplying the both sides by a non-negative number, or by multiplying the left hand side by a non-negative number that is not greater than 1, or by increasing the right hand side, as well as $A \preceq B \implies x^\top(A-B)x \preceq 0$. ☐

Proof of Proposition 4.1.8

Considering $M = \begin{bmatrix}-I & 0 \\ 0 & L_f^2\end{bmatrix}$, we have $\begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix} M \begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix}^\top = -(\Delta f)^\top \Delta f + L_f^2 (\Delta q)^\top \Delta q \geq 0$, where the inequality is implied by the Lipschitz continuity of $f(\cdot)$. ☐

Proof of Proposition 4.1.9

By definition, f is δ-QC with multiplier matrix M means that $\begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix}$ $M \begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix}^\top \geq 0$. Then, it follows in a straightforward manner that

$$\begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix} (-M) \begin{bmatrix}(\Delta f)^\top & (\Delta q)^\top\end{bmatrix}^\top \leq \gamma$$

for every $\gamma \geq 0$. ☐

Proof of Proposition 4.1.14

We observe that $\begin{bmatrix}(\Delta f)^\top & (\Delta x)^\top\end{bmatrix} \mathcal{M} \begin{bmatrix}(\Delta f)^\top & (\Delta x)^\top\end{bmatrix}^\top = (\Delta f)^\top(\Delta f) \leq L_f^2 \|\Delta x\|^2 \leq L_f^2 (2r)^2 = 4r^2 L_f^2$, where the second and third inequalities hold by Lipschitz continuity of $f(\cdot)$ and boundedness of the state space, respectively. ☐

Proof of Lemma 4.1.15

First, notice that $\Delta f = A\Delta x + \Delta g$. Given this and $\|g(x)\| \leq r$, we can conclude that $\begin{bmatrix}(\Delta f)^\top & (\Delta x)^\top\end{bmatrix} \mathcal{M} \begin{bmatrix}(\Delta f)^\top & (\Delta x)^\top\end{bmatrix}^\top = (\Delta f)^\top(\Delta f) - 2(\Delta x)^\top A^\top (\Delta f) +$

$(\Delta x)^\top A^\top A(\Delta x) = (\Delta f - A\Delta x)^\top \Delta f - A\Delta x) = (\Delta g)^\top(\Delta g) \leq (2r)^2.$ □

Proof of Proposition 4.1.16

By construction, we have the following condition:

$$\mathcal{M} - \begin{bmatrix} I_{n\times n} & -\mathcal{A} \\ -\mathcal{A}^\top & \mathcal{A}^\top\mathcal{A} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{11} - I & 0 \\ 0 & \mathcal{M}_{22} - \mathcal{M}_{12}^\top \mathcal{M}_{12} \end{bmatrix} \succeq 0,$$

since both submatrices on the diagonal are positive semi-definite by assumption. □

Proof of Proposition 4.1.17

The global Lipschitz continuity of LPV systems can be shown as follows:

$$\begin{aligned} \Delta f_k \triangleq \|f(x_1 - x_2)\| &= \|\sum_{i=1}^N \lambda_{i,k} A^i \Delta x_k\| \leq \sum_{i=1}^N \lambda_{i,k} \|A^i \Delta x_k\| \\ &\leq \sum_{i=1}^N \lambda_{i,k} \|A^i\| \|\Delta x_k\| \leq \|A^m\| \|\Delta x_k\|, \end{aligned}$$

with $\|A^m\| = \max_{i\in 1\ldots N} \|A^i\|$, where the first and second inequalities hold by submultiplicative inequality for norms and positivity of $\lambda_{i,k}$, the third inequality holds by the facts that $0 \leq \lambda_{i,k} \leq 1$ and $\sum_{i=1}^N \lambda_{i,k} = 1$. □

Proof of Lemma 4.5.1

Aiming to derive the governing equation for the evolution of the state errors, from (4.5) and (4.6), we obtain

$$\hat{d}_{1,k} = M_1(C_1 \tilde{x}_{k|k} + \Sigma d_{1,k} + v_{1,k}). \tag{A.25}$$

Moreover, from (4.3), (4.5) and (4.7)–(4.10), we have

$$\hat{d}_{2,k-1} = M_2[C_2(\Delta f(x_{k-1}) + G_1 \tilde{d}_{1,k-1} + G_2 d_{2,k-1} + W w_{k-1}) + v_{2,k}], \tag{A.26}$$

and by plugging $M_1 = \Sigma^{-1}$ into (A.25), we obtain

$$\tilde{d}_{1,k} = d_{1,k} - \hat{d}_{1,k} = -M_1(C_1 \tilde{x}_{k|k} + v_{1,k}), \tag{A.27}$$

where $\Delta f(x_k) \triangleq f(x_k) - f(\hat{x}_k)$. Then, by setting $M_2 = (C_2 G_2)^\dagger$ in (A.26) and using (A.27), we have

$$\tilde{d}_{2,k-1} = -M_2[C_2(\Delta f(x_{k-1}) - G_1 M_1(C_1 \tilde{x}_{k-1|k-1} + v_{1,k-1}) + W w_{k-1}) + v_{2,k}]. \tag{A.28}$$

Furthermore, it follows from (4.3),(4.9) and (4.10) that

$$\tilde{x}^\star_{k|k} = \Delta f(x_{k-1}) + G_1 \tilde{d}_{1,k-1} + G_2 \tilde{d}_{2,k-1} + W w_{k-1}. \tag{A.29}$$

In addition, by plugging $\tilde{d}_{k-1}$ and $\tilde{d}_{k-2}$ from (A.27) and (A.28) into (A.29), by (4.5) and (4.11), we obtain

$$\tilde{x}_{k|k} = (I - \tilde{L}C_2)\tilde{x}^{\star}_{k|k} - \tilde{L}\tilde{v}_k. \tag{A.30}$$

$$\tilde{x}^{\star}_{k|k} = \Phi[\Delta f(x_{k-1}) - G_1 M_1 C_1 \tilde{x}_{k-1|k-1}] + \tilde{w}_k, \tag{A.31}$$

where $\tilde{v}_k \triangleq v_{2,k}$, $\tilde{w}_k \triangleq -\Phi(G_1 M_1 v_{1,k-1} - W w_{k-1}) - G_2 M_2 v_{2,k}$ and $\Phi \triangleq I - G_2 M_2 C_2$. Finally, combining (A.30) and (A.31) returns the results. □

Proof of Theorem 4.5.3

Consider the state error dynamics (4.12) without bounded noise signals w_k and v_k

$$\tilde{x}_{k+1|k+1} = (I - \tilde{L}C_2)\Phi(\Delta f_k - \Psi \tilde{x}_k), \tag{A.32}$$

and the positive definite candidate Lyapunov function $V^{wn}_k = \tilde{x}^{\top}_{k|k} P \tilde{x}_{k|k}$ for some $P \succ 0$. We will show that (4.13) implies that $\Delta V^{wn}_k \triangleq V^{wn}_{k+1} - V^{wn}_k \preceq 0$, where

$$\begin{aligned}\Delta V^{wn}_k = {} & \Delta f_k^{\top} \Phi^{\top} (I - \tilde{L}C_2)^{\top} P (I - \tilde{L}C_2) \Phi \Delta f_k \\ & + \tilde{x}^{\top}_{k|k} (\Psi^{\top} \Phi^{\top} (I - \tilde{L}C_2)^{\top} P (I - \tilde{L}C_2) \Phi \Psi - P) \tilde{x}_{k|k} \qquad \text{(A.33)} \\ & - 2\Delta f_k^{\top} \Phi^{\top} (I - \tilde{L}C_2)^{\top} P (I - \tilde{L}C_2) \Phi \Psi \tilde{x}_{k|k}, \qquad \text{(A.34)}\end{aligned}$$

for each case. First, notice that for cases *(0)–II* and considering $Y = P\tilde{L}$, $\Pi \succeq 0 \iff I - \Gamma \succeq 0$ and $\begin{bmatrix} \Gamma & Y^{\top} \\ Y & P \end{bmatrix} \succeq 0$, which by pre- and post-multiplication by $\begin{bmatrix} I & 0 \\ 0 & P^{-1} \end{bmatrix}$, is equivalent to $I - \Gamma \succeq 0$ and $\begin{bmatrix} \Gamma & \tilde{L}^{\top} \\ \tilde{L} & P^{-1} \end{bmatrix} \succeq 0$. Applying Schur complement to the latter, $\Pi \succeq 0$ is equivalent to

$$0 \preceq \tilde{L}^{\top} P \tilde{L} \preceq \Gamma \preceq I. \tag{A.35}$$

On the other hand, defining $S \triangleq P - C_2^{\top} Y^{\top} - Y C_2$, (A.33) becomes

$$\begin{aligned}\Delta V^{wn}_k = {} & \Delta f_k^{\top} \Phi^{\top} (S + C_2^{\top} \tilde{L}^{\top} P \tilde{L} C_2) \Phi \Delta f_k \\ & + \tilde{x}^{\top}_{k|k} (\Psi^{\top} \Phi^{\top} (S + C_2^{\top} \tilde{L}^{\top} P \tilde{L} C_2) \Phi \Psi - P) \tilde{x}_{k|k} \\ & - 2\Delta f_k^{\top} \Phi^{\top} S \Phi \Psi \tilde{x}_{k|k} - 2\Delta f_k^{\top} \Phi^{\top} C_2^{\top} \tilde{L}^{\top} P \tilde{L} C_2 \Phi \Psi \tilde{x}_{k|k}.\end{aligned} \tag{A.36}$$

Now we consider each of the four cases, separately.

- Case (0): Applying Lemma 4.1.19 to (A.36), we obtain

  $$\Delta V^{wn}_k \leq -(\Delta f_k^{\top} \Theta \Delta f_k + \tilde{x}^{\top}_{k|k} \Xi \tilde{x}_{k|k} + 2\Delta f_k^{\top} \Lambda^i \tilde{x}_{k|k}) \tag{A.37}$$

  $$= -\begin{bmatrix} \Delta f_k^{\top} & \tilde{x}^{\top}_{k|k} \end{bmatrix} \Upsilon^i \begin{bmatrix} \Delta f_k^{\top} & \tilde{x}^{\top}_{k|k} \end{bmatrix}^{\top}, \tag{A.38}$$

  with Θ, Ξ and Λ^i defined in (4.14) and Υ^i in (4.13). Finally, (A.37) and (4.13) imply that $\Delta V^{wn}_k \leq 0$.

- Case I: Adding and subtracting $\Delta f_k^\top \Delta f_k$ from the right hand side of (A.36), as well as from the Lipschitz continuity of $f(\cdot)$, we have

$$\begin{aligned}\Delta V_k^{wn} \leq\, & \Delta f_k^\top \Phi^\top (S + C_2^\top \tilde{L}^\top P \tilde{L} C_2 - I)\Phi \Delta f_k \\ & + \tilde{x}_{k|k}^\top (\Psi^\top \Phi^\top (S + C_2^\top \tilde{L}^\top P \tilde{L} C_2)\Phi \Psi - P + L_f^2)\tilde{x}_{k|k} \\ & - 2\Delta f_k^\top \Phi^\top S \Phi \Psi \tilde{x}_{k|k} - 2\Delta f_k^\top \Phi^\top C_2^\top \tilde{L}^\top P \tilde{L} C_2 \Phi \Psi \tilde{x}_{k|k}. \end{aligned} \tag{A.39}$$

  Now, applying Lemma 4.1.19 to (A.39) results in (A.37) with Θ, Ξ and Λ^i defined in (4.16) and Υ^i defined in (4.13), which implies that $\Delta V_k^{wn} \leq 0$.

- Case II: To prove this, we first derive the following lemma.

  Lemma A.0.2. *Suppose $f(\cdot)$ is a Class II function. Then, at each time step k, Δf_k can be decomposed into a linear function of $\tilde{x}_{k|k}$ and a* bounded norm *uncertain nonlinear term, i.e., $\Delta f_k = \mathcal{A}\tilde{x}_{k|k} + s_k$, where $\|s_k\| \leq \gamma$.*

  Proof. Define $s_k \triangleq \Delta f(x_k) - \mathcal{A}\tilde{x}_{k|k}$. Then, notice that

$$\begin{aligned}\|s_k\|^2 &= s_k^\top s_k = (\Delta f(x_k) - \mathcal{A}\tilde{x}_{k|k})^\top (\Delta f(x_k) - \mathcal{A}\tilde{x}_{k|k}) \\ &= \Delta f(x_k)^\top \Delta f_k - 2\tilde{x}_{k|k}^\top \mathcal{A}^\top \Delta f(x_k) + \tilde{x}_{k|k}^\top \mathcal{A}^\top \mathcal{A}\tilde{x}_{k|k} \\ &= \begin{bmatrix}(\Delta f(x_k))^\top & \tilde{x}_{k|k}^\top\end{bmatrix} \mathcal{M} \begin{bmatrix}(\Delta f(x_k))^\top & \tilde{x}_{k|k}^\top\end{bmatrix}^\top \leq \gamma^2,\end{aligned}$$

  where the last inequality holds since $f(\cdot)$ is a DQC* function. □

  Now, from Lemma A.0.2 and (A.32), we have $\tilde{x}_{k+1|k+1} = (I - \tilde{L}C_2)\Phi(s_k - (\Psi - \mathcal{A})\tilde{x}_k)$. Comparing this with (A.32), the rest of the proof is similar to the one for case (0), with the only difference being the use of Δf_k and Ψ in the place of s_k and $\Psi - \mathcal{A}$, respectively.

- Case III: By $f(\cdot)$ being an LPV function as well as (A.32), we obtain

$$\tilde{x}_{k+1|k+1} = (I - \tilde{L}C_2)\Phi \hat{A}_k \tilde{x}_k, \tag{A.40}$$

  where $\hat{A}_k \triangleq \sum_{i=1}^{N} \lambda_{i,k}(A^i - \Psi)$. Then, the result follows directly from applying [88, Lemma 1]. □

Proof of Lemma 4.5.4

We define a similar candidate Lyapunov function V_k^{wn} as in the proof of Theorem 4.5.3, and we show that (4.18) implies $\Delta V_k^{wn} \leq 0$. First, notice that (4.18) is equivalent to $\Delta \succeq 0$ and $\begin{bmatrix} I & (I-\tilde{L}C_2)^\top P \\ P(I-\tilde{L}C_2) & P \end{bmatrix} \succeq 0$, which by pre- and post-multiplication by $\begin{bmatrix} P^{(-\frac{1}{2})} & 0 \\ 0 & P^{-1} \end{bmatrix}$ is, in turn, equivalent to

$$\begin{bmatrix} P^{-1} & P^{(-\frac{1}{2})}(I-\tilde{L}C_2)^\top \\ (I-\tilde{L}C_2)P^{(-\frac{1}{2})} & P^{-1} \end{bmatrix} \succeq 0 \text{ and } \Delta \succeq 0.$$

Applying Schur complement, we obtain equivalently that $\Delta \succeq 0$ and $P^{-1} - P^{(-\frac{1}{2})}(I - \tilde{L}C_2)^\top P(I-\tilde{L}C_2)P^{(-\frac{1}{2})} \succeq 0$. Pre- and post-multiplication by $P^{\frac{1}{2}}$ returns, equivalently,

$$\Delta \triangleq P - 2L_f^2 \lambda_{\max}(\Phi^\top \Phi) I - 2\Psi^\top \Phi^\top \Phi \Psi \succ 0, \tag{A.41}$$

$$(I-\tilde{L}C_2)^\top P (I-\tilde{L}C_2) \prec I. \tag{A.42}$$

Finally, by (A.33), (A.42), Lemma 4.1.19, Lipschitz continuity of $f(\cdot)$ and (A.41), we obtain

$$\begin{aligned} \Delta V_k^{wn} &\le \Delta f_k^\top (2\Phi^\top \Phi)\Delta f_k + \tilde{x}_{k|k}^\top (2\Psi^\top \Phi^\top \Phi \Psi - P)\tilde{x}_{k|k} \\ &\le 2\lambda_{\max}(\Phi^\top \Phi)\Delta f_k^\top \Delta f_k + \tilde{x}_{k|k}^\top (2\Psi^\top \Phi^\top \Phi \Psi - P)\tilde{x}_{k|k} \\ &\le \tilde{x}_{k|k}^\top (2L_f^2 \lambda_{\max}(\Phi^\top \Phi) I + 2\Psi^\top \Phi^\top \Phi \Psi - P)\tilde{x}_{k|k} \le 0. \end{aligned}$$

□

Proof of Lemma 4.5.5

To show that uniform detectability is sufficient for existence of an observer, notice that for a Class III function $f(\cdot)$, (4.12) can be written as

$$\tilde{x}_{k|k} = (I-\tilde{L}C_2)\overline{A}_{k-1}\tilde{x}_{k-1|k-1} + (I-\tilde{L}C_2)\tilde{w}_{k-1} - \tilde{L}\tilde{v}_{k-1}, \tag{A.43}$$

where $\tilde{w}_{k-1} \triangleq -(I-G_2M_2C_2)(G_1M_1v_{1,k-1} - w_{k-1}) - G_2M_2v_{2,k}, \overline{A}_k \triangleq \Phi(\sum_{i=1}^N \lambda_{i,k}A^i - \Psi)$ and $\tilde{v}_{k-1} \triangleq v_{2,k}$. Now, consider the following linear time-varying system without unknown inputs:

$$x_{k+1} = \overline{A}_k x_k + \tilde{w}_k, y_k = C_2 x_k + \tilde{v}_k. \tag{A.44}$$

Systems (A.43) and (A.9) are equivalent from the viewpoint of estimation, since the estimation error equations for both problems are the same, hence they both have the same objective. Therefore, the pair $(\overline{A}_k, C_2)$ needs to be uniformly detectable such that the observer is stable [10, Section 5].

Moreover, as for the necessity of the strong detectability of the constituent LTI systems, suppose for contradiction, that there exists a stable observer for system (4.3) with any sequence $\{\lambda_{i,k}\}_{k=0}^\infty$ for all $i \in \{1,2,\dots,N\}$ that satisfies $0 \le \lambda_{i,k} \le 1, \sum_{i=1}^N \lambda_{i,k} = 1, \forall k$, but one of the constituent linear time-invariant systems (e.g., (A^j, G, C, H)) is not strongly detectable. Since the observer exists for any sequence of $\lambda_{i,k}$, that means that an observer also exists when $\lambda_{j,k} = 1$ and $\lambda_{i,k} = 0, \forall i \neq j$ for all k. However, we know from [131] that strong detectability is necessary for the stability of the linear time-invariant system (A^j, G, C, H), which is a contradiction. Hence, the proof is complete. □

Proof of Theorem 4.5.8

We use a similar approach as in the proof of Theorem 4.5.3 for Class 0, I and II systems and a different approach for Class III systems. First, for Class 0, I and II systems, consider the error dynamics with bounded noise signals (4.12) and the candidate Lyapunov function $V_k^n \triangleq \tilde{x}_{k|k}^\top P \tilde{x}_{k|k}$. Observe that

$$\Delta V_k^n \triangleq V_{k+1}^n - V_k^n = \Delta V_k^{wn} + \Delta r_k, \tag{A.45}$$

where V_k^{wn} is the Lyapunov function for the error dynamics without noise signals, defined in (A.33), and

$$\Delta r_k \triangleq 2(\Delta f_k^\top - \tilde{x}_{k|k}^\top \Psi^\top)\Phi^\top (I - \tilde{L}C_2)^\top P \mathcal{W}(\tilde{L})\overline{w}_k + \overline{w}_k^\top \mathcal{W}(\tilde{L})^\top P \mathcal{W}(\tilde{L})\overline{w}_k, \tag{A.46}$$

with $\Phi, \Psi, \overline{w}_k$ and $\mathcal{W}(\tilde{L})$ defined in Lemma 4.5.1. We will show for each of the cases (0), I and II that

$$\overline{\Delta} r_k \triangleq \Delta r_k - \eta^2 \overline{w}_k^\top \overline{w}_k + \tilde{x}_k^\top \tilde{x}_k \leq 0. \tag{A.47}$$

Then, by (A.45) and (A.47) in addition to the fact that $\Delta V_k^{wn} \leq 0$ (follows from Theorem 4.5.3), we have

$$\Delta V_k^n \leq \eta^2 \overline{w}_k^\top \overline{w}_k - \tilde{x}_{k|k}^\top \tilde{x}_{k|k}. \tag{A.48}$$

Summing up both sides of (A.48) from zero to infinity, returns $V_\infty^n - V_0^n \leq \eta^2 \sum_{k=0}^\infty \overline{w}_k^\top \overline{w}_k - \sum_{k=0}^\infty \tilde{x}_{k|k}^\top \tilde{x}_{k|k} = \eta^2 \sum_{k=0}^\infty \vec{w}_i^\top \vec{w}_i - \sum_{k=0}^\infty \tilde{x}_{k|k}^\top \tilde{x}_{k|k}$, where at each time step k, $\vec{w}_k^\top = \begin{bmatrix} w_k^\top & v_k^\top \end{bmatrix}^\top$. Then, it follows from setting the initial conditions to zero that

$$\sum_{k=0}^\infty \tilde{x}_{k|k}^\top \tilde{x}_{k|k} \leq \eta^2 \sum_{k=0}^\infty \vec{w}_i^\top \vec{w}_i$$

.

Thus, it remains to show that (A.47) holds for each case (0)–II. Plugging the expression for $\mathcal{W}(\tilde{L})$ from Lemma 4.5.1 into (A.46), we obtain

$$\begin{aligned}\overline{\Delta} r_k = {} & \tilde{x}_{k|k}^\top \tilde{x}_{k|k} + 2(\Delta f_k - \Psi \tilde{x}_{k|k})^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R + C_2^\top \tilde{L}^\top P \tilde{L}\Omega)\overline{w}_k \\ & + \overline{w}_k^\top (R^\top P R - R^\top \Omega - \Omega^\top Y^\top R + \Omega^\top \Gamma \Omega - \eta^2 I)\overline{w}_k,\end{aligned}$$

which by (A.35) and Lemma 4.1.19, implies that:

$$\begin{aligned}\overline{\Delta} r_k \ \leq {} & \overline{w}_k^\top (R^\top P R - R^\top Y \Omega - \Omega^\top Y^\top R + \Omega^\top \Gamma \Omega - \eta^2 I + 2\Omega^\top \Omega)\overline{w}_k \\ & + \tilde{x}_{k|k}^\top (I + \Psi^\top \Phi^\top C_2^\top C_2 \Phi \Psi)\tilde{x}_{k|k} + 2\Delta f_k^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R)\overline{w}_k \\ & - 2\tilde{x}_{k|k}^\top \Psi^\top \Phi^\top (PR - Y\Omega - C_2^\top Y^\top R)\overline{w}_k + \Delta f_k^\top (\Phi^\top C_2^\top C_2 \Phi)\Delta f_k.\end{aligned} \tag{A.49}$$

Now, we separately consider each of the cases (0)–II:

- Case (0): It directly follows from (A.49) that $\overline{\Delta} r_k \leq -\zeta^\top \mathcal{N} \zeta \leq 0$, where $\zeta \triangleq \begin{bmatrix} \overline{w}_k^\top & \tilde{x}_k^\top & \Delta f_k^\top \end{bmatrix}^\top$ and $\mathcal{N}$ is the matrix in (4.20) with its elements defined in (4.21).

- Case I: First, notice that

  $$\begin{aligned} \Delta f_k^\top (\Phi^\top C_2^\top C_2 \Phi) \Delta f_k &\leq \lambda_{max}(\Phi^\top C_2^\top C_2 \Phi) \Delta f_k^\top \Delta f_k \\ &\leq \tilde{x}_{k|k}^\top (L_f^2 \lambda_{max}(\Phi^\top C_2^\top C_2 \Phi) I) \tilde{x}_{k|k}, \end{aligned} \tag{A.50}$$

  where the second inequality is implied by Lipschitz continuity of $f(\cdot)$. Then, it can be concluded from (A.49) and (A.50) that $\overline{\Delta} r_k \leq -\zeta^\top \mathcal{N} \zeta \leq 0$, where $\zeta \triangleq \begin{bmatrix} \overline{w}_k^\top & \tilde{x}_k^\top & \Delta f_k^\top \end{bmatrix}^\top$ and $\mathcal{N}$ is the matrix in (4.20) with its elements defined in (4.22).

- Case II: By Lemma A.0.2 and (A.46) we have

  $$\Delta r_k = \overline{w}_k^\top \mathcal{W}(\tilde{L})^\top P \mathcal{W}(\tilde{L}) \overline{w}_k + 2(\tilde{x}_{k|k}^\top (\mathcal{A} - \Psi)^\top + s_k^\top) \Phi^\top (I - \tilde{L} C_2)^\top P \mathcal{W}(\tilde{L}) \overline{w}_k, \tag{A.51}$$

  where $s_k = \Delta f_k - \mathcal{A} \tilde{x}_{k|k}$. Comparing (A.51) with (A.46), the rest of the proof is similar to the one for case (0),by replacing Δf_k and Ψ with s_k and $\Psi - \mathcal{A}$, respectively, which results in $\overline{\Delta} r_k \leq -\zeta^\top \mathcal{N} \zeta \leq 0$, where $\zeta \triangleq \begin{bmatrix} \overline{w}_k^\top & \tilde{x}_k^\top & \Delta f_k^\top \end{bmatrix}^\top$ and $\mathcal{N}$ is the matrix in (4.20) with its elements defined in (4.23).

- Case III: For this case, we consider a different approach compared to the previous cases. By $f(\cdot)$ being LPV and (4.12), we can define a system with $\tilde{x}_{k|k}$ as its state and $\tilde{z}_{k|k} = \tilde{x}_{k|k}$ as the output:

  $$\begin{aligned} \tilde{x}_{k|k} &= (I - \tilde{L} C_2) \overline{A}_{k-1} \tilde{x}_{k-1|k-1} + [(I - \tilde{L} C_2) R + \tilde{L} Q] \overline{w}_{k-1}, \\ \tilde{z}_{k|k} &= \tilde{x}_{k|k}, \end{aligned} \tag{A.52}$$

  where $\overline{A}_k \triangleq \Phi \sum_{i=1}^{N} \lambda_{i,k} (A^i - \Psi)$, each A^i is a constituent matrix of $f(\cdot)$ and Φ and Ψ are defined in Lemma 4.5.1. Now, by [38, Lemma 3], system (A.52) has an $\mathcal{H}_\infty$ performance bounded by η, if there exists a symmetric positive definite matrix S such that:

  $$\begin{gathered} \mathcal{S} \triangleq \begin{bmatrix} S & (I - \tilde{L} C_2) \overline{A^i} S & (I - \tilde{L} C_2) R + \tilde{L} Q & 0 \\ * & S & 0 & S \\ * & * & \eta I & 0 \\ * & * & * & \eta I \end{bmatrix} \succ 0, \\ \forall i \in \{1, 2, \ldots, N\}, \end{gathered} \tag{A.53}$$

  where $\overline{A}^i \triangleq \Phi(A^i - \Psi)$. Notice that the referenced lemma requires the existence of a *bounded matrix sequence*, which in our case is a sequence of time-invariant matrices (S is the same for each k), that is obviously bounded. By plugging $P = S^{-1} \succ 0$, defining $\mathcal{P} = \mathcal{P}^\top \triangleq \begin{bmatrix} 0 & P & 0 & 0 \\ P & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$ and applying some similarity transformations, we obtain

$$\mathcal{P}\mathcal{S}\mathcal{P} = \begin{bmatrix} P & \overline{A^i}^\top (I - C_2^\top \tilde{L}^\top) P & 0 & I \\ * & P & (I - \tilde{L}C_2)R + \tilde{L}Q & 0 \\ * & * & I & 0 \\ * & * & * & \eta I \end{bmatrix} \succeq 0, \forall i \in \{1, 2, \dots, N\}.$$

Setting $Y \triangleq P\tilde{L}$ completes the proof. $\square$

Proof of Theorem 4.5.9

To obtain (4.25) and (4.26), we will show that $\tilde{x}_{k|k} \leq \sqrt{\frac{\tilde{x}_{0|0}^\top P \tilde{x}_{0|o}}{\lambda_{\min}(P)}}$ and $\tilde{x}_{k|k} \leq \delta_0^x \theta^k + \overline{\eta} \sum_{i=1}^{k} \theta^{i-1}$. The former inequality follows from the non-increasing Lyapunov function defined in the proof of Theorem 4.5.3, as well as the fact that $\lambda_{\min}(A)\|x\|^2 \leq x^\top A x, \forall x \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n}$. The following shows that the latter inequality holds for each of the different system classes.

(I) If $f(\cdot)$ is a Class I function, then, the result in (4.25) with θ and $\overline{\eta}$ defined in (4.27), directly follows from Lipschitz continuity of $f(\cdot)$, as well as applying triangle and sub-multiplicative inequalities for norms on (4.12). Moreover, the result in (4.26) with β and $\overline{\alpha}$ defined in (4.27), is obtained by triangle and sub-multiplicative inequalities, (A.27) and (A.28).

(II) If $f(\cdot)$ is a Class II function, by Lemma A.0.2, (4.12), (A.27), (A.28) and triangle and sub-multiplicative inequalities, we obtain the results in (4.25) and (4.26) with $\theta, \overline{\eta}, \beta$ and $\overline{\alpha}$ defined in (4.28).

(III) If $f(\cdot)$ is a Class III function, we first need to find closed form expressions for the state and input estimation errors through the following lemma.

Lemma A.0.3. *The state and input estimation errors are*

$$\tilde{x}_{k|k} = \sum_{i=1}^{k} \prod_{j=0}^{i-2} A_{e,k-j}(\Psi \tilde{w}_{k-i} - \tilde{L}\tilde{v}_{k-i}) + \prod_{j=0}^{k-1} A_{e,k-j}\tilde{x}_{0|0},$$

$$\begin{aligned} \tilde{d}_{k-1} = & - \sum_{i=1}^{N} \lambda_{i,k-1}(V_1 M_1 C_1 + V_2 M_2 C_2 A_{e,i})\tilde{x}_{k-1|k-1} \\ & + (V_2 M_2 C_2 G_1 M_1 - V_1 M_1) T_1 v_{k-1} \\ & - V_2 M_2 C_2 w_{k-1} - V_2 M_2 T_2 v_k. \end{aligned}$$

Proof. Starting from (A.43) and applying simple induction return the results for the state errors. Then, the expression for the input errors follows from (A.27), (A.28) and (2.4). □

Now, we are ready to prove Theorem 2.3.5 for LPV functions. First, we define

$$\begin{aligned} B_{e,k} &\triangleq \prod_{j=0}^{k-1} A_{e,k-j}, \\ C_{e,k}^{i} &\triangleq \prod_{j=0}^{i-2} A_{e,k-j},\ \tilde{t}_k \triangleq \Psi\tilde{w}_k - \tilde{L}\tilde{v}_k, \end{aligned} \tag{A.54}$$

for $1 \leq i \leq k$. Then, from Lemma A.0.3, we have

$$\|\tilde{x}_{k|k}\| \leq \|B_{e,k}\|\|\tilde{x}_{0|l0}\| + \|\sum_{i=1}^{k} C_{e,k}^{i}\bar{t}_{k-i}\|, \tag{A.55}$$

by triangle inequality and submultiplicativity of norms. Moreover, by similar reasoning, we find

$$\begin{aligned} &\|B_{e,k}\|{\leq}\|\prod_{j=0}^{k-1}\sum_{i=1}^{N}\lambda_{k-j}^{i}\Psi\Phi(A^{i}-G_1M_1C_1)\|{\leq}\theta^{k}; \\ &\|\sum_{i=1}^{k}C_{e,k}^{i}\bar{t}_{k-i}\|{\leq}\sum_{i=1}^{k}\|C_{e,k}^{i}\|\|\bar{t}_{k-i}\|, \\ &\|C_{e,k}^{i}\| \leq \prod_{j=0}^{i-2}\|\sum_{s=1}^{N}\lambda_{s,k-j}A_{e,s}\| \leq \theta^{i-1}. \end{aligned} \tag{A.56}$$

Moreover, from (A.13),

$$\|\tilde{t}_{k-i}\| = \|\Re v_{k-i} + \Psi\Phi w_{k-i}\| \leq \overline{\eta}, \tag{A.57}$$

with $\Re \triangleq -(\Psi\Phi G_1M_1T_1 + \Psi G_2M_2T_2 + \tilde{L}T_2)$. Then, from (A.55)–(A.57), we obtain (4.25) with θ and $\overline{\eta}$ defined in (4.29). Furthermore, the result in (4.26) with β and $\overline{\alpha}$ defined in (4.29), follows from applying Lemma A.0.3, as well as triangle inequality, the facts that $0 \leq \lambda_{i,k} \leq 1, \sum_{i=1}^{N}\lambda_{i,k} = 1$ and submultiplicativity of matrix norms.

Finally, the steady state values are obtained by taking the limit from both sides of (4.25) and (4.26), assuming $\theta < 1$. □

Proof of Corollary 4.5.11

Clearly $\|A_{e,i}\| < 1$ implies that $\theta < 1$, which is a sufficient condition for the convergence of errors by Theorem 4.5.9. □

Proof of Proposition 5.2.3

The result follows directly from plugging the corresponding expressions into the right hand side term of Definition 5.2.2.

Proof of Theorem 5.2.4

To use contradiction, suppose that $\|r_k^q\|_2 > \delta_{r,k}^{q,*}$ and let q be the true mode, i.e., $q = q^*$ and thus, $T_2^q = T_2^{q*}$. By Proposition 5.2.3, $\Delta r_k^{q|q*} = 0$ and hence, $\|r_k^q\|_2 = \|r_k^{q|*}\|_2 \leq \delta_{r,k}^{q,*}$, which contradicts with the assumption. $\square$

Proof of Lemma 5.2.5

The first equality in (5.10) comes from Definition 5.2.2 and $z_{2,k}^q = C_2^q x_k + D_{2,k}^q u_k^q + v_{2,k}^q$ from (4.5) in Section 4.4.1, assuming that q is the true mode. To obtain the second equality, note that [118, (A.11)] returns

$$\begin{aligned}\tilde{x}_{k|k}^{\star,q} &= \Phi^q[\Delta f_{k-1}^q - G_1^q M_1^q C_1^q \tilde{x}_{k-1|k-1}^q] + \tilde{w}_k^q, \qquad (A.58)\\ \tilde{w}_k^q &\triangleq -\Phi^q(G_1^q M_1^q v_{1,k-1}^q - W^q w_{k-1}^q) - G_2^q M_2^q v_{2,k}^q.\end{aligned}$$

Now, from the first equality and (5.5), we have

$$r_k^{q|*} = C_2^q \Phi^q(\Delta f_{k-1}^q - G_1^q M_1^q C_1^q \tilde{x}_{k-1|k-1}^q) + \mathcal{Y}^q \overline{w}_{k-1}^q. \qquad (A.59)$$

On the other hand, by iteratively applying (5.7), we obtain:

$$\begin{aligned}\tilde{x}_{k|k}^q = &\sum_{i=1}^{i-1}[((I - \tilde{L}^q C_2^q)\Psi^q)^{i-1}(I - \tilde{L}^q C_2^q)\Phi^q \Delta f_{k-i}^q + (I - \tilde{L}^q C_2^q)^{i-1}\mathcal{W}^q \overline{w}_{k-i+1}^q]\\ &+ (-1)^k((I - \tilde{L}^q C_2^q)\Phi^q \Psi^q)^k \tilde{x}_{0|0}^q. \qquad (A.60)\end{aligned}$$

Combining (A.59) and (A.60) yields

$$r_k^{q|*} = A_k^q \tilde{x}_{0|0}^q + \sum_{i=0}^{k-1} F_i^q \Delta f_{k-1-i}^q + J_i^q \overline{w}_{k-i}^q,$$

which is equivalent to the second equality in (5.10).

Proof of Lemma 5.2.6

Consider the following optimization problem for $\|r_k^{q|*}\|_2$ by leveraging Lemma 5.2.5:

$$\begin{aligned}&\delta_{r,k}^q \triangleq \max_{t_k} \|\mathbb{A}_k^q t_k\|_2 \qquad (A.61)\\ &s.t.\ t_k = \begin{bmatrix}\tilde{x}_{0|0}^\top & v_0^{q\top} & \dots & v_k^{q\top} & w_0^{q\top} & \dots & w_{k-1}^{q\top} & \Delta f_0^{q\top} \dots \Delta f_{k-1}^{q\top}\end{bmatrix}^\top,\\ &\|\tilde{x}_{0|0}\|_2 \leq \delta_0^x,\ \|v_i^q\|_2 \leq \eta_v^q,\ \|w_j^q\|_2 \leq \eta_w^q,\ \|\Delta f_j^q\|_2 \leq L_f^q \overline{\delta}_j^{x,q} \leq L_f^q \overline{\delta}^{x,q},\\ &i \in \{0, ..., k\},\ j \in \{0, ..., k-1\}.\end{aligned}$$

The objective ℓ_2-norm function is continuous and the constraint set is an intersection of level sets of lower dimensional norm functions, which is closed and bounded, so is compact. Hence, by the Weierstrass Theorem [15, Proposition 2.1.1], the objective function attains its maxima on the constraint set and so a finite-valued upper bound exists. $\square$

Proof of Theorem 5.2.7

Consider the following optimization problem:

$$\begin{aligned} &\delta_{r,k}^{q,inf} \triangleq \max_{t_k} \|\mathbb{A}_k^q t_k\|_2 \qquad\qquad (A.62)\\ &s.t.\ t_k = t_k = \begin{bmatrix} \tilde{x}_{0|0}^\top & v_0^{q\top} & \dots & v_k^{q\top} & w_0^{q\top} & \dots & w_{k-1}^{q\top} & \Delta f_0^{q\top} \dots \Delta f_{k-1}^{q\top} \end{bmatrix}^\top, \\ &\quad \|\tilde{x}_{0|0}\|_\infty \le \delta_0^x,\ \|v_i^q\|_\infty \le \eta_v^q,\ \|w_j^q\|_\infty \le \eta_w^q,\ \|\Delta f_j^q\|_\infty \le L_f^q \bar{\delta}_j^{x,q} \\ &\quad \forall i \in \{0, ..., k\},\ \forall j \in \{0, ..., k-1\}. \end{aligned}$$

Comparing (A.61) with (A.62), the two problems have the same objective functions. Then, since $\|.\|_\infty \le \|.\|_2$, the constraint set for (A.61) is a subset of the one for (A.62). Hence $\delta_{r,k}^q \le \delta_{r,k}^{q,inf}$. Also, it is easy to see that $\delta_{r,k}^q \le \delta_{r,k}^{q,tri}$, which is obtained using triangle inequality and the sub-multiplicative property of norms. Moreover, (A.62) is a *maximization* of a convex objective function over a convex constraint (hypercube $\mathcal{X}_k^q$). By a famous result [101, Corollary 32.2.1], in such a problem, the objective function attains its maxima on some of the extreme points of the constraint set, which in this case are the vertices $\mathcal{T}_k$ of the hypercube $\mathcal{X}_k^q$.

Proof of Corollary 5.2.9

The result follows directly from plugging the above expressions into the right hand side term of Definition 5.2.2. $\square$

Proof of Lemma 5.3.4

To show (5.12), we first find a lower bound for $\delta_{r,k}^{q,inf}$. Then, we prove that the lower bound diverges and so does $\delta_{r,k}^{q,inf}$. Define $\tilde{t}_k^\star \triangleq \frac{t_k^\star}{\eta_k^t}$, where η_k^t is defined in Corollary 5.2.9. Now consider

$$\eta_k^t \sigma_{min}(\mathbb{A}_k^q) = \sigma_{min}(\eta_k^t \mathbb{A}_k^q) = \min_{\|t\|_2 \le 1} \|\eta_k^t \mathbb{A}_k^q t\|_2 \le \|\eta_k^t \mathbb{A}_k^q \tilde{t}_k^\star\|_2 = \|\mathbb{A}_k^q t_k^\star\|_2 \triangleq \delta_{r,k}^{q,inf},$$

where $\sigma_{min}(A)$ is the smallest non-trivial singular value of matrix A. The first equality holds since $\sigma_{\min}(.)$ is a linear operator and the second equality is a special case of the *matrix lower bound* [50] when ℓ_2-norms are considered. The inequality holds since $\|\tilde{t}_k^\star\|_2 = 1$ by Corollary 5.2.9, so $\tilde{t}_k^\star$ is a feasible point for the minimization problem (i.e., $\min_{\|t\|_2 \le 1} \|\eta_k^t \mathbb{A}_k^q t\|_2$) and the last equality holds by Theorem 5.2.7. So far we have shown that $\eta_k^t \sigma_{\min}(\mathbb{A}_k^q)$ is a lower bound for $\delta_{r,k}^{q,inf}$. Next, we will prove that $\eta_k^t \sigma_{min}(\mathbb{A}_k^q)$

is unbounded. First, it is trivial to observe that η_k^t grows unbounded by its definition in Corollary 5.2.9. Second, $\sigma_{\min}(\mathbb{A}_k^q) \leq \sigma_{\min}(\mathbb{A}_{k+1}^q)$, since the latter is an augmentation of the former with additional columns. Hence, $\eta_k^t \sigma_{\min}(\mathbb{A}_k^q)$ grows unbounded, since the product of the unbounded and positive $\sigma_{\min}(\mathbb{A}_k^q)$ and the unbounded and positive η_k^t is unbounded.

To prove (5.13), we show that $\{\delta_{r,k}^{q,tri}\}_{k=1}^{\infty}$ is a convergent sequence. Then, this fact, as well as (5.12) and the fact that $\hat{\delta}_{r,k}^q \triangleq \min\{\delta_{r,k}^{q,tri}, \delta_{r,k}^{q,inf}\}$ by Theorem 5.2.7, imply (5.13). To show the convergence of $\{\delta_{r,k}^{q,tri}\}_{k=1}^{\infty}$, starting from (5.11), we first show that $\forall q \in \mathbb{Q}$, $S_{1,k}^q \triangleq \sum_{i=0}^{k-2} L_f^q \|F_i^q\|_2 \overline{\delta}_{k-1-i}^{x,q} + \frac{1}{\sqrt{2}}\eta_v^q(\|J_i^{q,1}\|_2 + \|J_i^{q,3}\|_2) + \eta_w^q \|J_i^{q,2}\|_2$ on the right hand side of (5.11) converges to some steady state value. Note that $\|F_i^q\|_2 \leq \mathcal{R}^q \theta^{q^i}$ by the sub-multiplicative property of norms, where

$$\mathcal{R}^q \triangleq L_f^q \|C_2^q \Phi^q G_1^q M_1^q C_1^q\|_2 \|\Psi^q\|_2 \|\Phi^q\|_2$$

and θ^q is given in (5.8). Combining this and (A.79) implies that

$$\sum_{i=0}^{k-2} L_f^q \|F_i^q\|_2 \overline{\delta}_{k-1-i}^{x,q} \leq \mathcal{R}^q \left((\delta_0^x - \frac{\overline{\eta}^q}{1-\theta^q})(k-1)(\theta^q)^{k-1} + \frac{\overline{\eta}^q}{1-\theta^q} \frac{1-(\theta^q)^{k-1}}{1-\theta^q} \right),$$

and the upper bound tends to $\mathcal{R}^q \frac{\overline{\eta}^q}{(1-\theta^q)^2}$ as k tends to ∞, since $0 < \theta^q < 1$ (cf. (5.8)) and $\lim_{k\to\infty} k(\theta^q)^k = 0$ when $0 < \theta^q < 1$. Moreover, it follows from the definitions of J_i^q and θ^q (cf. Proposition 5.2.1 and Lemma 5.2.5), as well as the sub-multiplicative property of norms that:

$$\frac{1}{\sqrt{2}}\eta_v^q(\|J_i^{q,1}\|_2 + \|J_i^{q,3}\|_2) + \eta_w^q \|J_i^{q,2}\|_2 \leq \begin{cases} O^q, & i = 0, \\ S^q \theta^{q^i}, & i \geq 1, \end{cases}$$

where

$$O^q \triangleq \eta_w^q(\|C_2^q \Phi^q G_1^q M_1^q T_1^q\|_2 + \|(I - C_2^q G_2^q M_2^q) T_2^q\|_2) + \eta_v^q \|C_2^q \Phi^q W^q\|_2,$$
$$S^q \triangleq (\eta_w^q \|C_2^q \Phi^q G_1^q M_1^q C_1^q\|_2 (\|\Phi^q G_1^q M_1^q T_1^q\|_2 + \|G_2^q M_2^q T_2^q\|_2) + \eta_v^q \|\Phi^q W^q\|_2).$$

Combining these and (5.8) results in

$$\sum_{i=0}^{k-2} \frac{1}{\sqrt{2}}\eta_v^q(\|J_i^{q,1}\|_2 + \|J_i^{q,3}\|_2) + \eta_w^q \|J_i^{q,2}\|_2 \leq O^q + S^q \frac{\theta^q - \theta^{q^{k-1}}}{1-\theta^q},$$

where the upper bound tends to $\frac{S^q \theta^q}{1-\theta^q}$ as k tends to ∞. Next, it is straightforward to observe that all constitutent terms in $S_{2,k}^q \triangleq (\|A_k^q\|_2 + L_f^q \|F_{k-1}^q\|_2)\delta_0^x + \frac{1}{\sqrt{2}}\eta_v^q(\|J_{k-1}^{q,1}\|_2 + \|J_{k-1}^{q,3}\|_2) + \eta_w^q \|J_{k-1}^{q,2}\|_2$ (on the right hand side of (5.11)) are all decreasing to zero as k increases, since they are all upper bounded by some terms involving $(\theta^q)^k$ by their definitions (cf. Lemma 5.2.5) and the sub-multiplicative property. Hence, $\lim_{k\to\infty} \delta_{r,k}^{q,tri} = \lim_{k\to\infty}(S_{1,k}^q + S_{2,k}^q) = \lim_{k\to\infty} S_{1,k}^q < \infty$.

Proof of Lemma 5.3.5

Suppose, for contradiction, that none of q and q' are eliminated. Then

$$\begin{aligned}&\|C_2^q \hat{x}_{k|k}^{\star,q} + D_2^q u_k^q - C_2^{q'} \hat{x}_{k|k}^{\star,q'} - D_2^{q'} u_k^{q'}\|_2 = \|r_k^{q'} - r_k^q + z_{2,k}^q - z_{2,k}^{q'})\|_2 \\ &\leq \|r_k^{q'}\|_2 + \|r_k^q\|_2 + \|z_{2,k}^q - z_{2,k}^{q'}\|_2 \leq \delta_{r,k}^q + \delta_{r,k}^{q'} + R_y\|T_2^q - T_2^{q'}\|_2,\end{aligned}$$

where the equality holds by Definition 5.2.2, the first inequality holds by triangle inequality and the last inequality holds by the assumption that none of q and q' can be eliminated, as well as the boundedness assumption for the measurement space. This last inequality contradicts with the inequality in the lemma, thus the result holds. $\square$

Proof of Lemma 5.3.6

Recall from Proposition 5.2.3, Lemma 5.2.5 and (5.1) that:

$$r_k^q = \mathbb{A}_k^q t_k^q + (T_2^q - T_2^{q^*})(C^{q^*} x_k + H^{q^*} d_k^{q^*} + D^{q^*} u_k^{q^*} + v_k^{q^*}). \tag{A.63}$$

On the other hand, by applying Taylor series expansion to (5.1) we obtain:

$$x_k = J_{f,0}^{q^*} x_{k-1} + B^{q^*} u_{k-1}^{q^*} + G^{q^*} d_{k-1}^{q^*} + W^{q^*} w_{k-1}^{q^*} + (H.O.T)_k^{q^*}, \tag{A.64}$$

where $(H.O.T)_k^{q^*}$ is an error term that satisfies $\|(H.O.T)_k^{q^*}\|_2 \leq \frac{1}{2} H_f^{q^*}(\xi_k)$ for some $\xi_k \in X$. Then, by applying (A.64) at time steps $k, k-1, \ldots, 1$, plugging them into (A.63) and augmentating the results, we obtain (5.15). $\square$

Proof of Theorem 5.3.7

To show that (iii) is sufficient for asymptotic mode detectability, consider Lemma 5.3.5 with $\delta_{r,k}^{q,tri}$ as the upper bound. It suffices to show that $\exists K \in \mathbb{N}$, such that (5.14) holds for $k \geq K, \forall q \neq q' \in \mathbb{Q}$. Notice that by Definition 5.2.2, $C_2^q \hat{x}_{k|k}^{\star,q} = C_2^q x_k + T_2^q v_k - r_k^{q|*}$. Hence, by plugging this into (5.14), we need to show that $\exists K \in \mathbb{N}$ such that:

$$\|W^{q,q'} s_k^{q,q'}\|_2 > \delta_{r,k}^{q,tri} + \delta_{r,k}^{q',tri} + R_z^{q,q'}, \forall k \geq K, \forall q \neq q' \in \mathbb{Q}, \tag{A.65}$$

where $s_k^{q,q'} \triangleq \begin{bmatrix} x_k^\top & v_k^\top & r_k^{q|*\top} & r_k^{q'|*\top} & u_k^{q\top} & u_k^{q'\top} \end{bmatrix}^\top$. A sufficient condition to satisfy (A.65) is that $\exists K \in \mathbb{N}$ such that $\forall k \geq K$, (A.65) holds for all $s_k^{q,q'}$. Equivalently, it suffices that:

$$\underline{W}_k^{q,q'} > \delta_{r,k}^{q,tri} + \delta_{r,k}^{q',tri} + R_z^{q,q'}, \forall k \geq K, \forall q \neq q' \in \mathbb{Q},$$

where

$$\begin{aligned}\underline{W}_k^{q,q'} \triangleq & \min_{x_k, v_k, r_k^q, r_k^{q'}} \|W^{q,q'} s_k^{q,q'}\|_2 \\ & s.t. \ \|x_k\|_2 \leq R_x, \|v_k\|_2 \leq \eta_v, \|r_k^{q|*}\|_2 \leq \delta_{r,k}^{q,tri}, \|r_k^{q'|*}\|_2 \leq \delta_{r,k}^{q',tri}.\end{aligned}$$

Finally, by expanding the constraint set, it suffices to require that $\exists K \in \mathbb{N}$ such that:

$$\underline{\underline{W}}_k^{q,q'} > \delta_{r,k}^{q,tri} + \delta_{r,k}^{q',tri} + R_z^{q,q'}, \forall k \geq K, \forall q \neq q' \in \mathbb{Q},$$

where

$$\begin{aligned} \underline{\underline{W}}_k^{q,q'} \triangleq & \min_{s_k^{q,q'}} \|W^{q,q'} s_k^{q,q'}\|_2 \\ & s.t.\ \|s_k^{q,q'}\|_2^2 \leq R_x^2 + \eta_v^2 + (\delta_{r,k}^{q,tri})^2 + (\delta_{r,k}^{q',tri})^2 + (u_k^q)^2 + (u_k^{q'})^2. \end{aligned}$$

Now, by the *matrix lower bound* theorem [50] and a similar argument to the proof of Lemma 5.3.4, it is sufficient to require that $\exists K \in \mathbb{N}$ such that $\forall k \geq K, \forall q \neq q' \in \mathbb{Q}$:

$$\sigma_{min}^2(W^{q,q'}) > \frac{(\delta_{r,k}^{q,tri} + \delta_{r,k}^{q',tri} + R_z^{q,q'})^2}{R_x^2 + \eta_v^2 + (\delta_{r,k}^{q,tri})^2 + (\delta_{r,k}^{q',tri})^2 + (u_k^q)^2 + (u_k^{q'})^2}. \tag{A.66}$$

The result in (A.66) provides us a *time-dependent* sufficient condition for mode detectability. In order to find a *time-independent* sufficient condition, notice that $\frac{(\bar{\delta}_{r,k}^{q,tri} + \bar{\delta}_{r,k}^{q',tri} + R_z^{q,q'})^2}{R_x^2 + \eta_v^2}$ is an upper bound for the right hand side of (A.66), since the latter's denominator is smaller than the former's and the numerator of the latter is an upper bound signal for the former's by triangle inequality and the sub-multiplicative property of norms. So, a sufficient condition for (A.66) is that $\exists K \in \mathbb{N}$ such that $\forall k \geq K, \forall q \neq q' \in \mathbb{Q}$:

$$\sigma_{min}^2(W^{q,q'}) > \frac{(\bar{\delta}_{r,k}^{q,tri} + \bar{\delta}_{r,k}^{q',tri} + R_z^{q,q'})^2}{R_x^2 + \eta_v^2}. \tag{A.67}$$

Then, for the above to hold, it suffices that

$$\sigma_{min}^2(W^{q,q'}) > \lim_{k \to \infty} \frac{(\bar{\delta}_{r,k}^{q,tri} + \bar{\delta}_{r,k}^{q',tri} + R_z^{q,q'})^2}{R_x^2 + \eta_v^2},$$

which is equivalent to (iii) by (5.13).

As for the sufficiency of (i), we show that the sufficient conditions in (i) imply that if $q \neq q^*$, then the residual signal r_k^q grows unbounded. Then, since we showed in Lemma 5.3.4 that the computed upper bound signal $\hat{\delta}_{r,k}^q$ is bounded, so there must exist a time step K such that $r_k^q > \hat{\delta}_{r,k}^q$ for $k \geq K$, and hence, mode q will be eliminated after time step K and therefore, mode detectability holds. To do so, we show that if (i) holds, then the right hand side of (5.15) grows unbounded, and so does r_k^q. First, note that by Lemma 5.3.4, the first term in the right hand side of (5.15), i.e., $\mathbb{A}_k^q t_k^q$, is bounded. Moreover, (5.16) and the facts that the state space is bounded and $\|J_{f,0}^{q^*}\|_2 < 1$ imply that $\epsilon_k^{q^*}$, i.e., the third term in the right hand side of (5.15), is bounded.

Next, we show that the second term in the right hand side of (5.15), i.e. $\alpha_k^{q^*}$, grows unbounded. Consequently, the summation of the two bounded terms $\mathbb{A}_k^q t_k^q$ and

$\epsilon_k^{q^*}$ as well as the unbounded term $\alpha_k^{q^*}$ will be unbounded. To show that $\alpha_k^{q^*}$ grows unbounded, it suffices to show that for any $c > 0$, any specific mode q with the true mode being q^*, there exists a large enough K such that:

$$\|\alpha_K^{q^*}\|_2 = \left\| \begin{bmatrix} \mathbb{T}_K^{q,q^*} & \mathbb{C}_{u,K}^{q,q^*} & \mathbb{C}_{d,K}^{q,q^*} \end{bmatrix} \begin{bmatrix} \zeta_K^\top & u_{0:k}^{q^*\top} & d_{0:K}^{q*\top} \end{bmatrix}^\top \right\|_2 > c,$$

with $\mathbb{T}_K^{q,q^*} \triangleq (T^q - T^{q^*}) \begin{bmatrix} C_{x,K}^{q^*} & C_{\tilde{w},K}^{q^*} \end{bmatrix}$, $\mathbb{C}_{u,K}^{q,q^*} \triangleq (T^q - T^{q^*}) C_{u,K}^{q^*}$, $\mathbb{C}_{d,K}^{q,q^*} \triangleq (T^q - T^{q^*}) C_{d,K}^{q^*}$ and $\zeta_K \triangleq \begin{bmatrix} x_0^\top & \tilde{w}_{0:K}^{q^*\top} \end{bmatrix}^\top$. Since q^* is unknown, a sufficient condition to satisfy the above equality is that $\forall c > 0, \forall q' \neq q \in Q, \exists K \in \mathbb{N}$ such that:

$$\left\| \begin{bmatrix} \mathbb{T}_K^{q,q'} & \mathbb{C}_{u,K}^{q,q'} & \mathbb{C}_{d,K}^{q,q'} \end{bmatrix} \begin{bmatrix} \zeta_K^\top & u_{0:K}^{q'\top} & d_{0:k}^{q*\top} \end{bmatrix}^\top \right\|_2 > c.$$

So it suffices that $\forall c > 0, \forall q' \neq q \in Q, \exists \overline{d} \in \mathbb{R}, \exists K \in \mathbb{N}$, such that:

$$\underline{T}_k^{q,q'} > c,$$

where

$$\begin{aligned} \underline{T}_k^{q,q'} \triangleq \min_{\zeta_k'} & \left\| \begin{bmatrix} \mathbb{T}_K^{q,q'} & \mathbb{C}_{u,K}^{q,q'} & \mathbb{C}_{d,K}^{q,q'} \end{bmatrix} \zeta_K' \right\|_2 \\ s.t.\ & \zeta_K' = \begin{bmatrix} x_0^\top & \tilde{w}_{0:K}^{q^*\top} & u_{0:K}^{q'\top} & d_{0:K}^{q*\top} \end{bmatrix}^\top, \|d_{0:K}^{q*}\|_2 \geq \overline{d}, \\ & \|w_i\|_\infty \leq \eta_w,\ \|v_j\|_\infty \leq \eta_v, i \in \{0, ..., K-1\}, j \in \{0, ..., K\}. \end{aligned}$$

Once again, by the matrix lower bound theorem, a sufficient condition for the above inequality to hold is that $\exists \overline{d} \in \mathbb{R}, \exists K \in \mathbb{N}$, such that:

$$\underline{\underline{T}}_k^{q,q'} > \frac{c}{\sigma_{min}(\begin{bmatrix} \mathbb{T}_K^{q,q'} & \mathbb{C}_{u,K}^{q,q'} & \mathbb{C}_{d,K}^{q,q'} \end{bmatrix})},$$

where

$$\begin{aligned} \underline{\underline{T}}_k^{q,q'} \triangleq \min_{\tilde{w}_{0:K}^{q^*}, d_{0:K}^{q^*}} & \|\zeta_K'\|_2 \\ s.t.\ & \zeta_K' = \begin{bmatrix} x_0^\top & \tilde{w}_{0:K}^{q^*\top} & u_{0:k}^{q'\top} & d_{0:K}^{q*\top} \end{bmatrix}^\top, \|d_{0:K}^{q*}\|_2 \geq \overline{d}, \\ & \|w_i\|_\infty \leq \eta_w,\ \|v_j\|_\infty \leq \eta_v, i \in \{0, ..., K-1\}, j \in \{0, ..., K\}. \end{aligned} \tag{A.68}$$

Finally, since

$$\|\zeta_K'\|_2 = \left\| \begin{bmatrix} x_0^\top & \tilde{w}_{0:k}^{q^*\top} & u_{0:K}^{q'\top} & d_{0:K}^{q*\top} \end{bmatrix} \right\|_2 \geq \sqrt{0^2 + 0^2 + 0^2 + \|d_{0:K}^{q*\top}\|_2^2} = \|d_{0:K}^{q*\top}\|_2,$$

then a sufficient condition for (A.68) to hold is that

$$\|d_{0:K}^{q*\top}\|_2 > \frac{c}{\sigma_{min}(\ \begin{bmatrix} \mathbb{T}_K^{q,q'} & \mathbb{C}_{u,K}^{q,q'} & \mathbb{C}_{d,K}^{q,q'} \end{bmatrix}\ )}. \tag{A.69}$$

Now, suppose that $T_2^q \neq T_2^{q'}$ (otherwise the matrix in the denominator of (A.69) is zero and it never holds). Then, the right hand side of (A.69) converges asymptotically to $\tilde{\delta} \triangleq \max\{0, \frac{c}{\overline{\sigma}^{q,q'}}\}$, since the smallest singular value in the denominator either diverges, or converges to some steady value $\overline{\sigma}^{q,q'}$. So we set $\overline{d}$ to be equal to any real number that is strictly greater than $\tilde{\delta}$. By the unlimited energy assumption for the unknown input signal, at some large enough time step K, the monotonely increasing function $\|d_{0:k}^{q*}\|_2$ will exceed $\overline{d}$ and so, the system will be mode detectable. $\square$

Proof of Lemma 6.1.5

For $j \in \{1 \dots m\}$, consider the problem of $\overline{s}_j = \max\limits_{\underline{x} \leq x \leq \overline{x}} [Ax]_j$, where $[Ax]_j = \sum_{i=1}^n A_{j,i} x_i$ is the j-th component of the vector Ax. It is easy to verify that the solutions of this linear program are $x_i^* = \overline{x}_i$ if $A_{i,j} \geq 0$, and $x_i^* = -\underline{x}_i$ if $A_{i,j} < 0$, for $i \in \{1 \dots n\}$. Consequently, $\overline{s}_j = [A]_j^+ \overline{x} - [A]_j^{++} \underline{x}$, where $[A]_j$ is the j-th row of A. By similar reasoning, $\underline{s}_j = \min_{\underline{x} \leq x \leq \overline{x}} [Ax]_j = [A]_j^+ \underline{x} - [A]_j^{++} \overline{x}$. Thus, considering that $\sup_{\underline{x} \leq x \leq \overline{x}} Ax = [\overline{s}_1 \quad \dots \quad \overline{s}_m]^\top$ and $\inf_{\underline{x} \leq x \leq \overline{x}} Ax = [\underline{s}_1 \quad \dots \quad \underline{s}_m]^\top$, the proof is complete. $\square$

Proof of Lemma 6.1.10

Starting from (6.2), we obtain $f_d(\overline{x}, \underline{x}) = f(x_1) + C_f(\overline{x} - \underline{x})$ and $f_d(\underline{x}, \overline{x}) = f(x_2) + C_f(\underline{x} - \overline{x})$, which together imply

$$f_d(\overline{x}, \underline{x}) - f_d(\underline{x}, \overline{x}) = f(x_1) - f(x_2) + 2C_f(\overline{x} - \underline{x}), \tag{A.70}$$

where $\forall i \in \{1 \dots n\}$, $x_{1,i}$ and $x_{2,i}$ are either $\overline{x}_i$, or $\underline{x_i}$, depending on the case (cf. [128, Theorem 1; (10)–(13)]). Moreover, $\underline{x} \leq \overline{x}$ and $\underline{x} \leq x_1, x_2 \leq \overline{x}$. This implies that

$$-(\overline{x} - \underline{x}) \leq x_1 - x_2 \leq \overline{x} - \underline{x} \Rightarrow \|x_1 - x_2\| \leq \|\overline{x} - \underline{x}\|. \tag{A.71}$$

On the other hand, applying triangle inequality to (A.70) and by the Lipschitz continuity of f, we obtain

$$\|f_d(\overline{x}, \underline{x}) - f_d(\underline{x}, \overline{x})\| \leq L_f \|x_1 - x_2\| + 2\|C_f\| \|(\overline{x} - \underline{x})\|. \tag{A.72}$$

Combining (A.71) and (A.72) yields the result. $\square$

Proof of Lemma 6.3.3

Augmenting the state and output equations in (6.3) and from Corollary 6.1.9, we obtain

$$\underline{h}_k \leq [G^\top H^\top]^\top d_{k-1} \leq \overline{h}_k,$$

with $\underline{h}_k, \overline{h}_k$ defined in (6.11),(6.12). Then, the input framers in (6.10) can be obtained by using Propositions 6.1.2–6.1.8 and considering the fact that J is full rank. Finally, tightness is implied by Lemma 6.1.5 (where the A matrix equals J). $\square$

Proof of Theorem 6.3.4 and Corollary 6.3.5

From the state equation in 6.3, Corollary 6.1.9 and Proposition 6.1.4, we have $\underline{x}^p_k \leq x_k \leq \overline{x}^p_k$, where, $\underline{x}^p_k = \underline{f}_k + Bu_{k-1} + \underline{w} + G^+\underline{d}^p_{k-1} - G^{++}\overline{d}^p_{k-1}$, $\overline{x}^p_k = \overline{f}_k + Bu_{k-1} + \overline{w} + G^+\overline{d}^p_{k-1} - G^{++}\underline{d}^p_{k-1}$, where $\overline{d}^p_{k-1}, \underline{d}^p_{k-1}$ are the corresponding input framers, which can be obtained as affine functions of $\overline{x}^p_k, \underline{x}^p_k$ from (6.10) by Lemma 6.3.3. Doing so and plugging them back into the above expressions for $\overline{x}^p_k, \underline{x}^p_k$ yields the following linear system of equations

$$\begin{aligned} A_x \begin{bmatrix} \overline{x}^{p\top}_k & \underline{x}^{p\top}_k \end{bmatrix}^\top = & A_f \begin{bmatrix} \overline{f}^\top_k & \underline{f}^\top_k \end{bmatrix}^\top + A_g \begin{bmatrix} \overline{g}^\top_k & \underline{g}^\top_k \end{bmatrix}^\top + A_u u_{k-1} \\ & + A_w \begin{bmatrix} \overline{w}^\top & \underline{w}^\top \end{bmatrix}^\top + A_v \begin{bmatrix} \overline{v}^\top & \underline{v}^\top \end{bmatrix}^\top + A_y y_{k-1} \triangleq p_k, \end{aligned} \tag{A.73}$$

with $A_s, \forall s \in \{x, f, g, u, w, v, y\}$ given in the statement of the theorem and $\overline{q}_k, \underline{q}_k, \forall q \in \{f, g\}$ obtained from Corollary 6.1.9 with the corresponding interval $[\underline{x}_{k-1}, \overline{x}_{k-1}]$. By [53], the set of all solutions of (A.73) lies in an interval with the following maximal and minimal elements

$$\overline{x}^{p\top}_k = \overline{x}^{p,f}_k + \mu r, \quad \underline{x}^{p\top}_k = \underline{x}^{p,f}_k - \mu r, \tag{A.74}$$

where μ is a very large positive real number (infinity), $\overline{x}^{p,f}_k \triangleq (A^\dagger_x p_k)_{(1:n)}, \underline{x}^{p,f}_k \triangleq (A^\dagger_x p_k)_{(n+1:2n))}$, and $r \triangleq rowsupp(I - A^\dagger_x A_x)_{(1:n)}$, which also equals to $rowsupp(I - A^\dagger_x A_x)_{(n+1:2n)}$ by [75, Corollary 4.7] and the fact that A_x is a block real centro-Hermitian matrix by its definition. Now, the fact that $x_k \in [\underline{x}^p_k, \overline{x}^p_k]$, existence of affine parallelized abstraction matrix $A = (1/2)(\overline{A} + \underline{A})$ for $g_2(.)$ (cf. Proposition 6.1.2 and Corollary 6.1.3) and Proposition (6.1.4) imply that:

$$\underline{\alpha}_k \triangleq A^+\underline{x}^p_k - A^{++}\overline{x}^p_k \leq Ax_k \leq A^+\overline{x}^p_k - A^{++}\underline{x}^p_k \triangleq \overline{\alpha}_k. \tag{A.75}$$

Multiplying (A.75) by $A^\dagger$ and applying Proposition 6.1.4, (A.74) and [53] yield $\underline{x}^u_k \leq x_k \leq \overline{x}^u_k$, where

$$\begin{aligned} \overline{x}^u_k &= \min(\overline{x}^{p,f}_k + \mu r, A^{\dagger+}\overline{\alpha}_k - A^{\dagger++}\underline{\alpha}_k + \mu\tilde{r}), \\ \underline{x}^u_k &= \max(\underline{x}^{p,f}_k - \mu r, A^{\dagger+}\underline{\alpha}_k - A^{\dagger++}\overline{\alpha}_k - \mu\tilde{r}), \end{aligned} \tag{A.76}$$

with $\tilde{r} \triangleq rowsupp(I - A^\dagger A)$. Note that for the implementation of the update step, we iteratively find new *local* parallel abstraction slopes $A_{i,k}$ by iteratively solving the LP (6.1) for g_2 on the intervals obtained in the previous iteration, $\mathcal{B}^{*,i}_k = [\underline{x}^{*,i-1}_k, \overline{x}^{*,i-1}_k]$, to find *local* framers $\overline{x}^{*,i}_k, \underline{x}^{*,i}_k$ (cf. (6.13)–(6.16)), with additional constraints given in (6.18) in the optimization problems, which guarantees that the iteratively updated *local* intervals obtained using the local abstraction slopes are inside the global interval $[\underline{x}^u_k \; \overline{x}^u_k]$, computed in (A.76) using the *global* parallel affine abstraction slope A. This, in addition to (6.9), (6.13)–(6.14) and (A.76) ensure that

$$\begin{aligned} &\underline{x}^u_k \leq \underline{x}^{*,0} \leq \cdots \leq \underline{x}^{*,i} \leq \cdots \leq \lim_{i\to\infty} \underline{x}^{*,i} \triangleq \underline{x}_k, \\ &\overline{x}_k \triangleq \lim_{i\to\infty} \overline{x}^{*,i} \leq \overline{x}^{*,0} \leq \cdots \leq \overline{x}^{*,i} \leq \cdots \leq \overline{x}^u_k, \end{aligned}$$

$\forall i \in \{1 \dots \infty\}$, where $\overline{x}_k, \underline{x}_k$ are the returned updated state framers by the observer. Since our goal is to obtain sufficient existence conditions that can be checked *a priori* instead of for each time step k, we use (A.74) and (A.76) with the *global* interval (that includes all local intervals), which result in

$$\begin{aligned}&\overline{x}_k \le \min(\overline{x}^{p,f}_k + \mu r, \mathbb{A}_1 \overline{x}^{p,f}_k - \mathbb{A}_2 \underline{x}^{p,f}_k + ((\mathbb{A}_1 + \mathbb{A}_2)r + \mu\tilde{r})),\\ &\underline{x}_k \ge \max(\underline{x}^{p,f}_k - \mu r, \mathbb{A}_1 \underline{x}^{p,f}_k - \mathbb{A}_2 \overline{x}^{p,f}_k - ((\mathbb{A}_1 + \mathbb{A}_2)r + \mu\tilde{r})),\end{aligned} \tag{A.77}$$

where $\mathbb{A}_1 \triangleq A^{\dagger +} A^{+} + A^{\dagger ++} A^{++}$ and $\mathbb{A}_2 \triangleq A^{\dagger +} A^{++} + A^{\dagger ++} A^{+}$. Considering (A.77) and given the facts that μ is infinite and $r(j), r'(j) \in \{0, 1\}, \forall j \in \{1 \dots n\}$, where $r' \triangleq (\mathbb{A}_1 + \mathbb{A}_2)r + \tilde{r}$, it suffices for the finiteness of the right hand sides of (A.77) that $\forall j \in \{1 \dots n\} : r(j)r'(j) = 0$. This is equivalent to (6.17). Moreover, since $\{\overline{x}^{*,i}_k\}$ and $\{\underline{x}^{*,i}_k\}$ for all i are, by construction, computed with over-approximations of the observation function g_2, $\underline{x}^{*,i}_k \le x_k \le \overline{x}^{*,i}_k$ holds by (6.13)–(6.14). Further, $(\underline{x}^{*,i}_k, \overline{x}^{*,i}_k) \xrightarrow{i \to \infty} (\underline{x}_k, \overline{x}_k)$, hence correctness follows for the state framer, while correctness for the input framer holds by Lemma 6.3.3. Finally, without the update step in (6.9), (6.17) reduces to $r = rowsupp(I - A^{\dagger}_x A_x) = 0$, which is equivalent to the rank condition in Corollary 6.3.5 by [75]. ☐

Proof of Lemma 6.3.6

The bounds for $d_{1,k}$ can be obtained by applying Propositions 6.1.4 and 6.1.8 to (6.5). Moreover, since $d_{2,k}$ does not appear in (6.5) and (6.6), it cannot be estimated at the current time. ☐

Proof of Theorem 6.3.8

Let $\Delta^x_k \triangleq \overline{x}_k - \underline{x}_k$, (similarly for $\Delta x^{p,f}_k$). Then, by (A.77),

$$\Delta^x_k \le \min(\Delta x^{p,f}_k + 2\mu r, (\mathbb{A}_1 + \mathbb{A}_2)\Delta x^{p,f}_k + 2((\mathbb{A}_1 + \mathbb{A}_2)r + \mu\tilde{r}))).$$

From this and using the fact that $\min(a, b) \le \mathbf{D}a + (I - \mathbf{D})b,\ \forall a, b \in \mathbb{R}^n,\ \forall \mathbf{D} \in \mathbb{D}$, where $\mathbb{D}$ is the set of all diagonal matrices that their diagonal elements are 0 or 1, we obtain

$$\Delta^x_k \le (\mathbf{D} + (I - \mathbf{D})(\mathbb{A}_1 + \mathbb{A}_2))\Delta x^{p,f}_k + 2\mu(\mathbf{D}r + (I - \mathbf{D})r'),$$

where $r' \triangleq (\mathbb{A}_1 + \mathbb{A}_2)r + \tilde{r}$. Since (6.17) holds (equivalently $r(j)r'(j) = 0, \forall j \in \{1 \dots n\}$), choosing any $\mathbf{D} \in \mathbb{D}^* \subseteq \mathbb{D}$, with $\mathbb{D}^* = \{\mathbf{D}^* \in \mathbb{D} \mid \mathbf{D}^*_{jj} = r'(j) \text{ if } r(j) \neq r'(j), \forall j \in \{1 \dots n\}\}$ eliminates the second term on the right hand side of the above inequality and returns

$$\Delta^x_k \le (\mathbf{D} + (I - \mathbf{D})(\mathbb{A}_1 + \mathbb{A}_2))\Delta x^{p,f}_k, \quad \forall \mathbf{D} \in \mathbb{D}^*. \tag{A.78}$$

On the other hand, from (A.73), (A.74) and Corollary 6.1.9, we obtain

$$\Delta x^{p,f}_k \le \Delta \tilde{f}^x_{k-1} + \Delta z, \tag{A.79}$$

where $\Delta\tilde{f}^x_k \triangleq T_f\Delta f^x_k + T_g\Delta g^x_k$, $\Delta f^x_k \triangleq f_d(\overline{x}_k, \underline{x}_k) - f_d(\underline{x}_k, \overline{x}_k)$, $\Delta g^x_k \triangleq g_d(\overline{x}_k, \underline{x}_k) - g_d(\underline{x}_k, \overline{x}_k)$, $\Delta z \triangleq T_f\Delta w + T_g\Delta v$, $\Delta w \triangleq \overline{w} - \underline{w}$, $\Delta v \triangleq \overline{v} - \underline{v}$, $T_f \triangleq (I - K_1 - L_1)^\dagger(I - K_1 + L_1)$ and $T_g \triangleq (I - K_1 - L_1)^\dagger(K_2 + L_2)$. Next, by (A.78), (A.79), non-negativity of $\hat{\mathbf{D}} \triangleq (\mathbf{D} + (I - \mathbf{D})(\mathbb{A}_1 + \mathbb{A}_2))$ and Proposition 6.1.4, an upper bound sequence for the interval widths holds:

$$\Delta^x_k \leq \hat{\mathbf{D}}\Delta\tilde{f}^x_{k-1} + \hat{\mathbf{D}}\Delta z \quad \forall \mathbf{D} \in \mathbb{D}^*. \tag{A.80}$$

Below, we will show that either of the three conditions in the theorem implies uniform boundedness of $\{\Delta^x_k\}^\infty_{k=0}$.

Condition (i): Since Assumption 6.2.1 holds, the application of triangle inequality to (A.80) yields

$$\|\Delta^x_k\| \leq \mathcal{L}_{\mathbf{D}}\|\Delta^x_{k-1}\| + \|\hat{\mathbf{D}}\Delta z\| \quad \forall \mathbf{D} \in \mathbb{D}^*, \tag{A.81}$$

with $\mathcal{L}_{\mathbf{D}} \triangleq L_{f_d}\|\hat{\mathbf{D}}T_f\| + L_{g_d}\|\hat{\mathbf{D}}T_g\|$ and L_{f_d}, L_{g_d} obtained from Lemma 6.1.10. Since $\mathcal{L}^* \leq 1$ (by Condition (i)), the sequence $\{\|\Delta^x_k\|\}^\infty_{k=0}$ is uniformly bounded. Therefore, the interval width dynamics is stable.

Condition (ii): To show that Condition (ii) implies stability, with slightly abuse of notation, let $\mathbf{D}$ be a specific member of $\mathbb{D}^*$ and suppose we show the stability of the dynamical system $\Delta^x_{k+1} = \hat{\mathbf{D}}\Delta\tilde{f}^x_k + \hat{\mathbf{D}}_0\Delta z$, where $\hat{\mathbf{D}} \triangleq (\mathbf{D} + (I - \mathbf{D})(\mathbb{A}_1 + \mathbb{A}_2))$. Then, by *Comparison Lemma* [64], the dynamical system $\Delta^x_{k+1} \leq \hat{\mathbf{D}}\Delta\tilde{f}^x_k + \hat{\mathbf{D}}_0\Delta z$ is stable. To do so, consider a candidate Lyapunov function $V_k = \Delta^{x\top}_k\Delta^x_k$ and let $\hat{T}_f \triangleq \hat{\mathbf{D}}T_f, \hat{T}_g \triangleq \hat{\mathbf{D}}T_g$. Then, it can be shown that $\Delta V_k \triangleq V_{k+1} - V_k \leq \Delta^{\zeta\top}_k\hat{\mathcal{T}}\Delta^\zeta_k$, with $\Delta^\zeta_k \triangleq \begin{bmatrix}\Delta^{x\top}_k & \Delta v^\top & \Delta w^\top & \Delta f^{x\top}_k\end{bmatrix}^\top$ and $\hat{\mathcal{T}}$ defined in the statement of the theorem, as follows:

$$\begin{aligned}\Delta V_k &= \Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_f\,\Delta f^x_k + \Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_g\Delta g^x_k + \Delta v^\top\hat{T}^\top_g\hat{T}_g\Delta v + \Delta w^\top\hat{T}^\top_f\hat{T}_f\Delta w\\ &\quad + 2(\Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_g\Delta g^x_k + \Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_g\Delta v + \Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_f\Delta w + \Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_g\Delta v\\ &\quad + \Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_f\Delta w + \Delta v^\top\hat{T}^\top_g\hat{T}_f\Delta w) - \Delta^{x\top}_k\Delta^x_k \leq (\lambda_{\max}(\hat{T}^\top_f\hat{T}_f)L^2_{f_d}\\ &\quad + \lambda_{\max}(\hat{T}^\top_g\hat{T}_g)L^2_{g_d} - 1)\Delta^{x\top}_k\Delta^x_k + \Delta v^\top\hat{T}^\top_g\hat{T}_g\Delta v + \Delta w^\top\hat{T}^\top_f\hat{T}_f\Delta w\\ &\quad + 2(\Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_g\Delta g^x_k + \Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_g\Delta v + \Delta f^{x\top}_k\hat{T}^\top_f\hat{T}_f\Delta w + \Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_g\Delta v)\\ &\quad + 2(\Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_f\Delta w + \Delta v^\top\hat{T}^\top_g\hat{T}_f\Delta w) = \Delta^{\zeta\top}_k\hat{\mathcal{T}}\Delta^\zeta_k,\end{aligned}$$

where the first inequality holds because $\Delta f^{x\top}_k\Delta f^x_k = \|\Delta f^x_k\|^2 \leq L^2_{f_d}\|\Delta^x_k\|^2$ (and similarly for $\Delta g^{x\top}_k\Delta g^x_k$) by Lemma 6.1.10 and $\Delta g^{x\top}_k\hat{T}^\top_g\hat{T}_g\Delta g^x_k \leq \lambda_{\max}(\hat{T}^\top_g\hat{T}_g)\Delta g^{x\top}_k\Delta g^x_k = \lambda_{\max}(\hat{T}^\top_g\hat{T}_g)\|\Delta g^x_k\|^2 \leq L^2_{g_d}\lambda_{\max}(\hat{T}^\top_g\hat{T}_g)\|\Delta^x_k\|^2$ by using the *Rayleigh Quotient* and Lemma 6.1.10. Now, by the Lyapunov Theorem, stability is satisfied if $\hat{\mathcal{T}} \preceq 0$ or equivalently $\lambda_{max}(\hat{\mathcal{T}}) \leq 0$ and hence $\Delta V_k \leq \Delta^{\zeta\top}_k\hat{\mathcal{T}}\Delta^\zeta_k \leq 0$. This, and given that in system (A.80), $\mathbf{D}$ can be *any* member of $\mathbb{D}^*$ (not a specific member), it suffices for stability that $\exists\mathbf{D} \in \mathbb{D}^*$ such that $\lambda_{max}(\hat{\mathcal{T}}) \leq 0$, i.e., Condition (ii) should hold.

Condition (iii): Similarly, we consider a candidate Lyapunov function $V_k = \Delta_k^{x\top} P \Delta_k^x$, where $P \succ 0$, which can be shown to satisfy $\Delta V_k \triangleq V_{k+1} - V_k \leq 0$ under Condition (iii). To show this, let $\hat{\Delta}\tilde{f}_k^{x\top} \triangleq \hat{\mathbf{D}}\Delta\tilde{f}_k^{x\top}$, $\hat{\Delta}z \triangleq \hat{\mathbf{D}}\Delta z$, $\hat{\Delta}\zeta_k \triangleq \begin{bmatrix}\hat{\Delta}\tilde{f}_k^{x\top} & \Delta_k^{x\top} & \hat{\Delta}z^\top\end{bmatrix}^\top$ and note that $\hat{\Delta}\tilde{f}_k^{x\top}\Lambda\hat{\Delta}\tilde{f}_k^x \leq \hat{\Delta}\tilde{f}_k^{x\top}\hat{\Delta}\tilde{f}_k^x \leq \mathcal{L}_{\mathbf{D}}^2\hat{\Delta}_k^{x\top}\hat{\Delta}_k^x$, where the inequalities hold by choosing Γ such that $\Gamma \triangleq I - \Lambda \succeq 0$ and Lemma 6.1.10, respectively. Consequently, $\mathcal{L}_{\mathbf{D}}^2\Delta_k^{x\top}\Delta_k^x - \hat{\Delta}\tilde{f}_k^{x\top}\Lambda\hat{\Delta}\tilde{f}_k^x \geq 0$. Then, inspired by a simplifying trick used in [39, Proof of Theorem 1] to satisfy $\Delta V_k \leq 0$, it suffices to guarantee that $\tilde{V}_k \triangleq \Delta V_k + \mathcal{L}_{\mathbf{D}}^2\Delta_k^{x\top}\Delta_k^x - \hat{\Delta}\tilde{f}_k^{x\top}\Lambda\hat{\Delta}\tilde{f}_k^x = \Delta V_k + \mathcal{L}_{\mathbf{D}}^2\Delta_k^{x\top}\Delta_k^x - \hat{\Delta}\tilde{f}_k^{x\top}(I-\Gamma)\hat{\Delta}\tilde{f}_k^x \leq 0$, where

$$\begin{aligned}\tilde{V}_k &= \hat{\Delta}\tilde{f}_k^{x\top}P\hat{\Delta}\tilde{f}_k^x + \hat{\Delta}z^\top P\hat{\Delta}z + 2\hat{\Delta}z^\top P\hat{\Delta}\tilde{f}_k^x - \Delta_k^{x\top}P\Delta_k^x + \mathcal{L}_{\mathbf{D}}^2\Delta_k^{x\top}\Delta_k^x - \hat{\Delta}\tilde{f}_k^{x\top}(I-\Gamma)\hat{\Delta}\tilde{f}_k^x \\ &= \hat{\Delta}\tilde{f}_k^{x\top}(P+\Gamma-I)\hat{\Delta}\tilde{f}_k^x + \Delta_k^{x\top}(\mathcal{L}_{\mathbf{D}}^2 I - P)\Delta_k^x + \hat{\Delta}z^\top P\hat{\Delta}z + 2\hat{\Delta}z^\top P\Delta\tilde{f}_k^x \\ &= \hat{\Delta}\zeta_k^\top\mathcal{P}_{\mathbf{D}}\hat{\Delta}\zeta_k \leq 0,\end{aligned}$$

with $\mathcal{P}_{\mathbf{D}}$ given in the statement of the theorem. This, along with $\Gamma \succeq 0$, is equivalent to Condition (iii). ☐

Proof of Lemma 6.3.10

Applying (A.81) repeatedly, for all $\mathbf{D} \in \mathbb{D}^{**}$, we have

$$\|\Delta_k^x\| \leq \hat{\mathcal{L}}^k\|\Delta_0^x\| + \textstyle\sum_{i=0}^{k-1}\hat{\mathcal{L}}^{k-i}\|\hat{\Delta}z\| = \hat{\mathcal{L}}^k\delta_0^x + \|\hat{\Delta}z\|\frac{1-\hat{\mathcal{L}}^k}{1-\hat{\mathcal{L}}}.$$

Further, from (6.10)–(6.12) we obtain $\Delta_{k-1}^d \leq \hat{J}_1(\Delta_k^x + \Delta f_k^x) + \hat{J}_2\Delta g_k^x + \hat{J}_1\Delta w + \hat{J}_2\Delta v$, where $\hat{J} \triangleq \begin{bmatrix}\hat{J}_1 & \hat{J}_2\end{bmatrix} \triangleq J^+ + J^{++}$. Applying Lemma 6.1.10 and triangle inequality returns the upper bound for $\|\Delta_{k-1}^d\|$, while taking the limit of $k \to \infty$ results in the steady-state values. The rest of the results follow from the non-increasing Lyapunov functions defined in the proof of Theorem 6.3.8 and the use of the Rayleigh Quotient. ☐

Proof of Proposition 6.1.2

Consider the case when the global affine abstraction matrices are unknown. Then, by setting $\mathcal{B} = \mathbb{X}$, $A_{\mathcal{B}}^q = \mathbb{A}^q$ and $\theta_{\mathcal{B}}^q$, constraint (6.18) is redundant and so, the LP (6.1) boils down to a special case of the LP in [108, (16)], with only one considered partition. Then, (7.12) follows from [108, Theorem 1]. Moreover, given the global affine abstractions, solving the LP in (6.1) is equivalent to solving the the LP in [108, (16)] on the corresponding interval (set) of $\mathcal{B}$, with the extra (non-trivial) constraint (6.18). This constraint along with the result in [108, Theorem 1] lead to (7.13). ☐

Proof of Theorem 7.2.2

We will prove this by induction. For the base case, by assumption, $\underline{z}_0 \leq z_0 \leq \overline{z}_0$ holds. Now, for the induction step, suppose that $\underline{z}_{k-1} \leq z_{k-1} \leq \overline{z}_{k-1}$. Then, Propositions 6.1.4–6.1.2 as well as (6.3),(7.6)–(7.3c) and [59, Theorem 1] imply that $\underline{z}_k^p \leq z_k \leq \overline{z}_k^p$. Given this, iteratively obtaining upper and lower abstraction matrices for the observation function $g(.)$ based on Proposition 6.1.2 and applying Proposition 6.1.4, we have

$$\underline{\alpha}_{i,k} \leq A^g_{i,k} z_k \leq \overline{\alpha}_{i,k}, \tag{A.82}$$

where $\underline{\alpha}_{i,k}, \overline{\alpha}_{i,k}$ are given in (7.10) and $A^g_{i,k}$ is a solution of the LP in (6.1), i.e., the parallel abstraction slope for function $g(.)$ at iteration i in the corresponding compatible interval $[\underline{z}^u_{i-1,k}, \underline{z}^u_{i-1,k}]$. Then, multiplying (A.82) by $A^{g\dagger}_{i,k}$, Proposition 6.1.4 and using the fact that $\underline{z}^u_{i-1,k}, \overline{z}^u_{i-1,k}$ are framers for the augmented state z_k at time k and [53], we obtain $\underline{z}^u_{i,k} \leq z_k \leq \overline{z}^u_{i,k}$, with $\underline{z}^u_{i,k}, \overline{z}^u_{i,k}$ given in (7.8). Now, note that by construction, the sequences of updated upper and lower framers, $\{\overline{z}^u_{i,k}\}^{\infty}_{i=0}$ and $\{\underline{z}^u_{i,k}\}^{\infty}_{i=0}$ with $\overline{z}^u_{0,k} = \overline{z}^p_k$ and $\overline{z}^u_{0,k} = \overline{z}^p_k$, are monotonically decreasing and increasing, respectively, and hence are convergent by the *monotone convergence theorem*. Consequently, their limits $\overline{z}_k, \underline{z}_k$ are the tightest possible framers, i.e., $\forall i \in \{1 \dots \infty\}$:

$$\begin{array}{c}\underline{z}^u_{0,k} \leq \cdots \leq \underline{z}^u_{i,k} \leq \cdots \leq \lim_{i \to \infty} \underline{z}^u_{i,k} \triangleq \underline{z}_k, \\ \overline{z}_k \triangleq \lim_{i \to \infty} \overline{z}^u_{i,k} \leq \cdots \leq \overline{z}^u_{i,k} \leq \cdots \leq \overline{z}^u_{0,k},\end{array}$$

where $\overline{z}_k, \underline{z}_k$ are the returned updated augmented state framers by the observer. This completes the proof. □

Proof of Lemma 7.2.3

It directly follows from [59, Theorem 1] and Theorem 7.2.2 that the model estimates are correct, i.e, $\forall k \in \{0 \dots \infty\} : \underline{h}_k(\zeta_k) \leq h(\zeta_k) \leq \overline{h}_k(\zeta_k)$. Moreover, considering the data-driven abstraction procedure in the model learning step, note that by construction, the data set used at time step k is a subset of the one used at time $k+1$. Hence, by [59, Proposition 2] the abstraction model satisfies *monotonicity*, i.e., (7.16) holds. □

Proof of Lemma 6.1.10

Starting from (6.2), it is not hard to verify that

$$\Delta q_\zeta = q(\zeta_1) - q(\zeta_2) + 2C^q \Delta\zeta, \tag{A.83}$$

for some ζ_1, ζ_2 that satisfy $\underline{\zeta} \leq \zeta_1, \zeta_2 \leq \overline{\zeta}$. On the other hand, by Proposition 6.1.2 in addition to Proposition 6.1.4, $\forall j \in \{1, 2\}$:

$$\mathbb{A}^{q+}\underline{\zeta} - \mathbb{A}^{q++}\overline{\zeta} + \underline{e}^q \leq q(\zeta_j) \leq \mathbb{A}^{q+}\overline{\zeta} - \mathbb{A}^{q++}\underline{\zeta} + \overline{e}^q,$$

which implies that $q(\zeta_1) - q(\zeta_2) \leq |\mathbb{A}^q|\Delta q_\zeta + \Delta e^q$. Combining this and (A.83) yields the result. □

Proof of Theorem 7.2.6

Note that our goal is to obtain sufficient stability conditions that can be checked *a priori* instead of for each time step k. On the other hand, for the implementation of the update step, we iteratively find new *local* parallel abstraction slopes $A^g_{i,k}$ by iteratively solving the LP (6.1) for g on the intervals obtained in the previous iteration, $\mathcal{B}^u_{i,k} = [\underline{z}^u_{i-1,k}, \overline{z}^u_{i-1,k}]$, to find *local* framers $\overline{z}^u_{i,k}, \underline{z}^u_{i,k}$ (cf. (7.7)–(7.10)), with additional constraints given in (6.18) in the optimization problems, which guarantees that the iteratively updated *local* intervals obtained using the local abstraction slopes are inside the global interval, i.e.,

$$\begin{array}{c}\underline{z}^u_k \leq \underline{z}^u_{0,k} \leq \cdots \leq \underline{z}^u_{i,k} \leq \cdots \leq \lim_{i\to\infty} \underline{z}^u_{i,k} \triangleq \underline{z}_k,\\ \overline{z}_k \triangleq \lim_{i\to\infty} \overline{z}^u_{i,k} \leq \cdots \leq \overline{z}^u_{i,k} \leq \cdots \leq \overline{z}^u_{0,k} \leq \overline{z}^u_k,\end{array}$$

where we apply (7.8) for just one iteration (dropping index i) with $\overline{z}^u_{k,0} = \overline{z}^p_k, \underline{z}^u_{k,0} = \underline{z}^p_k$ to obtain:

$$\begin{bmatrix}\overline{z}^u_k\\ \underline{z}^u_k\end{bmatrix} = \begin{bmatrix}\min(A^{g\dagger+}\overline{\alpha}_k - A^{g\dagger++}\underline{\alpha}_k + \omega, \overline{z}^p_k)\\ \max(A^{g\dagger+}\underline{\alpha}_k - A^{g\dagger++}\overline{\alpha}_k - \omega, \underline{z}^p_k)\end{bmatrix}, \tag{A.84}$$

This allows us to use the *global* parallel affine abstraction slope A^g for the stability analysis as follows. Dropping index i in (7.9)–(7.10) and defining $\Delta^z_k \triangleq \overline{z}_k - \overline{z}_k$ (and similarly for $\Delta^{z^p}_k, \Delta^g_e, \Delta^f_e, \Delta^h_e, \Delta^\alpha_k, \Delta^t_k$), (7.8) implies that $\forall D_1 \in \mathbb{D}_{n+p}$

$$\Delta^z_k \leq \min(|A^{g\dagger}|\Delta^\alpha_k + 2\kappa\mathbf{r}, \Delta^{z^p}_k) \leq D_1(|A^{g\dagger}|\Delta^\alpha_k + 2\kappa\mathbf{r}) + (I - D_1)\Delta^{z^p}_k, \tag{A.85}$$

where the second inequality follows from generalization of the fact that $\min(a, b) \leq \lambda a + (1-\lambda)b, \forall a, b \in \mathbb{R}, \lambda \in [0, 1]$. Moreover, from (7.9)–(7.10) and a similar reasoning, we observe that $\forall D_2 \in \mathbb{D}_l$:

$$\Delta^\alpha_k \leq \min(|W^g|\Delta v + \Delta^g_e, |A^g|\Delta^{z^p}_k) \leq D_2(|W^g|\Delta v + \Delta^g_e) + (I - D_2)|A^g|\Delta^{z^p}_k. \tag{A.86}$$

On the other hand, by similar arguments, it follows from (7.3a)–(7.3c) that $\forall D_3 \in \mathbb{D}_n$,

$$\Delta^{z^p}_k \leq \begin{bmatrix} D_3(|A^f|\Delta^z_{k-1} + |W^f|\Delta w + \Delta^f_e) + (I - D_3)\Delta^f_{k-1}\\ |A^h|\Delta^z_{k-1} + |W^h|\Delta w + \Delta^h_e\end{bmatrix}, \tag{A.87}$$

where $\Delta^f_{k-1} \triangleq f_d(\overline{\zeta}_{k-1}, \underline{\zeta}_{k-1}) - f_d(\underline{\zeta}_{k-1}, \overline{\zeta}_{k-1})$. Furthermore, by Lemma 6.1.10, $\Delta^f_{k-1} \leq (|A^f| + 2C^f_z)\Delta^z_{k-1} + (|W^f| + 2C^f_w)\Delta w + \Delta^f_e$, with $C^f = \begin{bmatrix} C^f_z & C^f_u & C^f_w\end{bmatrix}$ given in (6.2). This, in addition to (A.85)–(A.87), Proposition 6.1.4 and non-negativity of both sides of all the inequalities, lead to:

$$\Delta^z_k \leq \mathcal{A}^g(D_1, D_2)\mathcal{A}^{f,h}(D_3)\Delta^z_{k-1} + \Delta^g(D_1, D_2) + \mathcal{A}^g(D_1, D_2)\Delta^{f,h}(D_3) + 2\kappa D_1\mathbf{r}, \tag{A.88}$$

for $(D_1, D_2, D_3) \in \mathbb{D}_{n+p} \times \mathbb{D}_l \times \mathbb{D}_n$, where we define

$$\begin{aligned}\mathcal{A}^g(D_1, D_2) &\triangleq D_1|A^{g\dagger}|D_2|A^g| + (I - D_1),\\ \mathcal{A}^{f,h}(D_3) &\triangleq \begin{bmatrix}(|A^f| + 2(I - D_3)C^f_z)^\top & |A^h|^\top\end{bmatrix}^\top,\\ \Delta^g(D_1, D_2) &\triangleq D_1|A^{g\dagger}|D_2(|W^g|\Delta v + \Delta^g_e),\\ \Delta^{f,h}(D_3) &\triangleq \begin{bmatrix}((|W^f| + 2(I - D_3)C^f_w)\Delta w + \Delta^f_e)^\top & (|W^h|\Delta w + \Delta^h_e)^\top\end{bmatrix}^\top.\end{aligned}$$

Since κ can be infinitely large, in order to make the right hand side of (A.88) finite in finite time, we choose $D_1 \in \mathbb{D}_{n+p}$ such that $D_1\mathbf{r} = 0$, i.e., $D_{1,i,i} = 0$ if $r(i) = 1, \forall i \in \{1 \ldots n+p\}$. Then, by the *Comparison Lemma* [64], it suffices for uniform boundedness of $\{\Delta_k^z\}_{k=0}^{\infty}$ that the following dynamic system be stable:

$$\Delta_k^z = \mathcal{A}^g(D_1, D_2)\mathcal{A}^{f,h}(D_3)\Delta_{k-1}^z + \tilde{\Delta}(D_1, D_2), \tag{A.89}$$

where the error term $\tilde{\Delta}(D_1, D_2) \triangleq \Delta^g(D_1, D_2) + \mathcal{A}^g(D_1, D_2)\Delta^{f,h}(D_3)$ is a bounded disturbance. This implies that the system (A.89) is stable (in the sense of uniform stability of the interval sequnces) if and only if the matrix $\mathcal{A}(D_1, D_2, D_3) \triangleq \mathcal{A}^g(D_1, D_2)\mathcal{A}^{f,h}(D_3)$ is (non-strictly) stable for at least one choice of (D_1, D_2, D_3), equivalently (7.17) should hold. $\square$

Proof of Lemma 6.3.10

The proof is straightforward by applying Proposition 6.1.4, computing (A.88) iteratively, using the fact that by Theorem 7.2.6, $\mathcal{A}(D_1, D_2, D_3)$ is a stable matrix for the tuple of (D_1, D_2, D_3) that is a solution of (7.17), and from triangle inequality. $\square$

Proof of Theorem 8.3.1

We need to show that for all $j \in \mathcal{J}$, $f_d^j(\zeta, \hat{\zeta}; \mathbf{m}^j)$ given in (8.10), satisfies all the conditions in Definition 8.1.5. Starting from (8.10), $f_d^j(\zeta, \hat{\zeta}; \mathbf{m}^j) = h_j(z_{\mathbf{m}^j}^c(\hat{\zeta}, \zeta)) + f_j(z_{\mathbf{m}^j}^c(\zeta, \hat{\zeta})) - h_j(z_{\mathbf{m}^j}^c(\zeta, \hat{\zeta}))$, with $z_{\mathbf{m}^j}^c(\zeta, \hat{\zeta})$ given in (8.14). Then defining $g_j(\cdot) \triangleq f_j(\cdot) - h_j(\cdot)$, we have $f_d^j(\zeta, \hat{\zeta}; \mathbf{m}^j) = h_j(z_{\mathbf{m}^j}^c(\hat{\zeta}, \zeta)) + g_j(z_{\mathbf{m}^j}^c(\zeta, \hat{\zeta}))$.

We will show that $h_j(x_{\mathbf{m}^j}^c(\hat{\zeta}, \zeta))$ and $g_j(x_{\mathbf{m}^j}^c(\zeta, \hat{\zeta}))$ are both non-decreasing in ζ and non-increasing in $\hat{\zeta}$ and so is $f_d^j(\zeta, \hat{\zeta}; \mathbf{m}^j)$. First, since $\mathbf{m}^j \in \mathbf{M}^j$, by construction of $\mathbf{M}^j$, (cf. (8.12)–(8.13)), $h_j(\cdot) \in \mathcal{H}_{\mathbf{M}^j}$ is a Jacobian sign-stable function. Next, in order to proceed, we need the following proposition.

Proposition A.0.4. *The Jacobian sign-stable functions $h_j(\cdot) \in \mathcal{H}_{\mathbf{M}^j}$ (as introduced in the statement of Theorem 8.3.1) and $g_j(\cdot) \triangleq f_j(\cdot) - h_j(\cdot)$ are $(h_j(\cdot), -g_j(\cdot))$ aligned, i.e., given $i \in \{1, \ldots, n_z\}$:*

(8.2) *holds for $g_j \Leftrightarrow$* (8.3) *holds for h_j, and equivalently,* (8.3) *holds for $g_j \Leftrightarrow$* (8.2) *holds for h_j.*

Proof: First, note that $g_j(\cdot) = f_j(\cdot) - h_j(\cdot)$ and (8.9) imply that $\forall z \in \mathcal{Z}$ and $\forall \nu^g \in \partial^o g_j(z)$, there exists $\xi^h \in \partial h_j(z)$ such that:

$$\underline{a}^{f_j} - \xi^{h_j} \leq \nu^{g_j} \leq \overline{a}_{f_j} - \xi^{h_j}. \tag{A.90}$$

Now, $\forall i \in \{1, \ldots, n_z\}$, consider the two following possible cases:

- $\mathbf{m}_i^j \geq \max(\overline{a}_i^{f_j}, 0) \to h_j(\cdot)$ is Jacobian positive sign-stable in the i'th dimension. Moreover, since $h_j(\cdot) \in \mathcal{H}_{\mathbf{M}^j}$, then $\forall z' \in \mathcal{Z}, \partial^o h_j(z') \subseteq \mathbf{M}^j$ by (8.11). Particularly, considering $z' = z$ and given the fact that in the i'th dimension

$\mathbf{m}_i^j \geq \max(\overline{a}_i^{f_j}, 0)$, we conclude that $\forall \xi^{h_j} \in \partial^o h_j(z), \overline{a}_i^{f_j} - \xi_i^{h_j} \leq 0$. This and (A.90) imply that $\nu_i^{g_j} \leq 0, \forall \nu^{g_j} \in \partial^o g_j(z)$, which means that $g_j(\cdot)$ is Jacobian negative sign-stable in the ith dimension, since z has been taken arbitrarily in the domain of $g_j(\cdot)$.

- $\mathbf{m}_i^j \leq \min(\underline{a}_i^{f_j}, 0) \rightarrow h_j(\cdot)$ is Jacobian negative sign-stable in the ith dimension. Moreover, since $h_j(\cdot) \in \mathcal{H}_{\mathbf{M}^j}$, then $\forall z' \in \mathcal{Z}, \partial^o h_j(z') \subseteq \mathbf{M}^j$ by (8.11). Particularly, considering $z' = z$ and given the fact that in the i'th dimension $\mathbf{m}_i^j \leq \min(\underline{a}_i^{f_j}, 0)$, we conclude $\forall \xi^{h_j} \in \partial^o h_j(z), \underline{a}_i^{f_j} - \xi_i^{h_j} \geq 0$. This and (A.90) imply that $\nu_i^{g_j} \geq 0, \forall \nu^{g_j} \in \partial^o g_j(z)$ which means that $g_j(\cdot)$ is Jacobian positive sign-stable in the i'th dimension, since z has been taken arbitrarily in the domain of $g_j(\cdot)$.

$\square$

Now consider $z_1 \geq z_2, z_0$ all in $\mathcal{Z}$ and hence $\forall s \in \{1, 2\}$:

$$\begin{aligned} z^c_{\mathbf{m}^j}(z_s, z_0) &= D^{\mathbf{m}^j} z_s + (I_{n_z} - D^{\mathbf{m}^j}) z_0, \\ z^c_{\mathbf{m}^j}(z_0, z_s) &= D^{\mathbf{m}^j} z_0 + (I_{n_z} - D^{\mathbf{m}^j}) z_s, \end{aligned} \tag{A.91}$$

by (8.14). So, for each $i \in \{1, \ldots, n_z\}$, we can consider two cases:

- $\mathbf{m}_i^j \geq \max(\overline{a}_i^{f_j}, 0) \rightarrow h_j(\cdot)$ is Jacobian positive sign-stable in the i'th dimension and also $D^{\mathbf{m}^j}_{i,i} = 0 \leftrightarrow g_j(\cdot)$ is Jacobian negative sign-stable in the i'th dimension by Proposition A.0.4 and $z^c_{\mathbf{m}^j,i}(z_1, z_0) \leq z^c_{\mathbf{m}^j,i}(z_2, z_0)$, $z^c_{\mathbf{m}^j,i}(z_0, z_1) \geq z^c_{\mathbf{m}^j,i}(z_0, z_2)$, by (A.91) and the fact that $D^{\mathbf{m}^j}_{i,i} = 0$.
- $\mathbf{m}_i^j \leq \min(\underline{a}_i^{f_j}, 0) \rightarrow h_j(\cdot)$ is Jacobian negative sign-stable in the i'th dimension and also $D^{\mathbf{m}^j}_{i,i} = 1 \leftrightarrow g_j(\cdot)$ is Jacobian positive sign-stable in the ith dimension by Proposition A.0.4 and $z^c_{\mathbf{m}^j,i}(z_1, z_0) \geq z^c_{\mathbf{m}^j,i}(z_2, z_0), z^c_{\mathbf{m}^j,i}(z_0, z_1) \leq z^c_{\mathbf{m}^j,i}(z_0, z_2)$.

Considering these two cases together, we observe that whenever

$$z^c_{\mathbf{m}^j,i}(z_1, z_0) \leq z^c_{\mathbf{m}^j,i}(z_2, z_0)$$

holds, then $h_j(\cdot)$ is Jacobian positive sign-stable in dimension i and whenever

$$z^c_{\mathbf{m}^j,i}(z_1, z_0) \geq z^c_{\mathbf{m}^j,i}(z_2, z_0)$$

holds, $h_j(\cdot)$ is Jacobian negative sign-stable in dimension i. Similarly, whenever $z^c_{\mathbf{m}^j,i}(z_0, z_1) \leq z^c_{\mathbf{m}^j,i}(z_0, z_2)$, $g_j(\cdot)$ is Jacobian positive sign-stable in dimension i and whenever $z^c_{\mathbf{m}^j,i}(z_0, z_1) \geq z^c_{\mathbf{m}^j,i}(z_0, z_2)$, $g_j(\cdot)$ is Jacobian negative sign-stable in dimension i. These facts imply that $h_j(z^c_{\mathbf{m}^j}(z_1, z_0)) \leq h_j(z^c_{\mathbf{m}^j}(z_2, z_0))$ and $g_j(z^c_{\mathbf{m}^j}(z_0, z_1)) \leq g_j(z^c_{\mathbf{m}^j}(z_0, z_2))$, which means that $h(x^c_{\mathbf{m}^j}(\hat{\zeta}, \zeta))$ and $g_j(z^c_{\mathbf{m}^j}(\zeta, \hat{\zeta}))$ are non-decreasing in ζ. Precisely similar arguments imply that $h_j(z^c_{\mathbf{m}^j}(\hat{\zeta}, \zeta))$ and $g_j(z^c_{\mathbf{m}^j}(\zeta, \hat{\zeta}))$ are non-increasing in $\hat{\zeta}$. Finally, it is easy to observe that if $\zeta = \hat{\zeta}$, $z^c_m(\zeta, \zeta) = \zeta$ and so $f^j_d(\zeta, \zeta; \mathbf{m}^j) = f_j(\zeta)$. $\square$

Proof of Lemma 8.3.2

The first equalities in (8.15) and (8.16) are implications of sequentially applying Corollary 8.1.10 on all the decomposition functions in the family (8.10). As for the second inequalities in (8.15) and (8.16), consider $m^j \in \mathbf{M}^j$ and construct $\tilde{m}^j \in \mathbf{M}_c^j$ as follows. $\tilde{m}_i^j = \max(\overline{a}_i^{f_j}, 0)$ if $m_i^j \geq \max(\overline{a}_i^{f_j}, 0)$ and $\tilde{m}_i^j = \min(\underline{a}_i^{f_j}, 0)$ if $m_i^j \leq \min(\underline{a}_i^{f_j}, 0)$. Then, it can be easily verified from (8.14) that $z_{m^j}^c(z_1, z_2) = z_{\tilde{m}^j}^c(z_1, z_2)$ and so by (8.10), $f_d(z_1, z_2; m^j\}_{j=1}^{|\mathcal{J}|}) = f_d(z_1, z_2; \tilde{m}^j\}_{j=1}^{|\mathcal{J}|})$. Hence, for any $m^j \in \mathbf{M}^j, \exists \tilde{m}^j \in \mathbf{M}_c^j$ that admits an equivalent decomposition function as the one that m^j admits. So, the serch, i.e., the minimization and maximization operations can be equivalently done over $\mathbf{M}_c^j$, instead of over $\mathbf{M}^j$. This fact, results in the second equalities in (8.15) and (8.16) and so completes the proof. $\square$

Proof of Theorem 8.3.3

First, note that since for any $\{h_j(\cdot)\}_{j=1}^{|\mathcal{J}|} \in \{\mathcal{H}_{\mathbf{M}^j}\}_{j=1}^{|\mathcal{J}|}$ and any pair of corresponding $\mathbb{M}^l, \mathbb{M}^u \in \{\mathbf{M}_j\}_{j=1}^{|\mathcal{J}|}$, $W_{f_j}^{\mathbb{M}_l, \mathbb{M}_u}(\mathcal{Z}) \supseteq V_{f_j}(\mathcal{Z}), \forall j \in \mathcal{J}$, then (8.20) and (8.5) imply that:

$$\begin{aligned} q(W_{f_j}^{\mathbb{M}_l, \mathbb{M}_u}(\mathcal{Z}), V_{f_j}(\mathcal{Z})) &= \max\{l_1(m_u^j) - \overline{f}_j^{\text{true}}, \underline{f}_j^{\text{true}} + l_2(m_l^j)\} \\ &\geq \tfrac{1}{2}(l_1(m_u^j) + l_2(m_l^j) + \Delta f_j^{true}) \end{aligned} \tag{A.92}$$

with $l_1(m_u^j) \triangleq h_j(\overline{z}_{m_u^j}^c) - h_j(\underline{z}_{m_u^j}^c) + f_j(\underline{z}_{m_u^j}^c)$, $l_2(m_l^j) \triangleq h_j(\overline{z}_{m_l^j}^c) - h_j(\underline{z}_{m_l^j}^c) - f_j(\overline{z}_{m_l^j}^c)$, $\overline{z}_{m_u^j}^c \triangleq z_{m_u^j}^c(\overline{z}, \underline{z}), \underline{z}_{m_l^j}^c \triangleq z_{m_l^j}^c(\underline{z}, \overline{z})$, $\Delta f_j^{true} \triangleq \underline{f}_j^{\text{true}} - \overline{f}_j^{\text{true}}$ and the inequality in (A.92) holds by $\max(a, b) \geq \frac{1}{2}(a + b)$. Next, consider the following proposition.

Proposition A.0.5. *The following programs are equivalent, in the sense that they attain equal sets of solutions and optimal values:*

$$\min_{m \in \mathbb{M}_j} l_s(m) \equiv \min_{m \in \mathbb{M}_j^c} l_s(m), \ \forall s \in \{1, 2\},$$
$$\min_{m \in \mathbb{M}_j} [l_1(m) + l_2(m)] \equiv \min_{m \in \mathbb{M}_j^c} [l_1(m) + l_2(m)],$$

where $\mathbf{M}_j \triangleq \{\mathbf{m} \in \mathbb{R}^{n_z} | \mathbf{m}_i \leq \min(\underline{a}_i^{f_j}, 0)$ *or* $\mathbf{m}_i \geq \max(\overline{a}_i^{f_j}, 0)\}$ *and* $\mathbf{M}_j^c \triangleq \{\mathbf{m} \in \mathbf{M}_j | \mathbf{m}_i = \min(\underline{a}_i^{f_j}, 0)$ *or* $\mathbf{m}_i = \max(\overline{a}_i^{f_j}, 0)\}$.

Proof: We will show that for each $m' \in \mathbf{M}_j$, there exists an $m^0 \in \mathbf{M}_j^c$ such that $l_s(m') = l_s(m^0), \ \forall s \in \{1, 2\}$. To do this, consider $m' \in \mathbf{M}_j$ and construct m_0 as follows: $m_i^0 = \min(\underline{a}_i^{f_j}, 0)$ if $m_i' \leq \underline{a}_i^{f_j}$ and $m_i^0 = \max(\overline{a}_i^{f_j}, 0)$ if $m_i' \geq \overline{a}_i^{f_j}$. Clearly $m^0 \in \mathbf{M}_j^c$. Moreover, if $m_i' \geq \overline{a}_i^{f_j}$, then by (8.14), $D_{i,i}^{m'} = \text{sgn}(\min(\underline{a}_i^{f_j}, 0) - m_i') = 0 = \text{sgn}(0) = \text{sgn}(\min(\underline{a}_i^{f_j}, 0) - m_i^0) = D_{i,i}^{m^0}$ and if $m_i' \leq \underline{a}_i^{f_j}$, then again by (8.14), $D_{i,i}^{m'} = \text{sgn}(\min(\underline{a}_i^{f_j}, 0) - m_i') = 1 = \text{sgn}(\min(\underline{a}_i^{f_j}, 0) - m_i^0) = D_{i,i}^{m^0}$. Hence, $D^{m'} = D^{m^0}$ and by (8.14), $z_{m'}^c(z_1, z_2) = z_{m^0}^c(z_1, z_2), \forall z_1, z_2 \in \mathcal{Z}$. This implies that $\overline{z}_{m'}^c = \overline{z}_{m^0}^c$ and $\underline{z}_{m'}^c = \underline{z}_{m^0}^c$, which result in $l_s(m') = l_s(m^0), \ \forall s \in \{1, 2\}$. $\square$

Now, (A.92) and Proposition A.0.5 imply (8.21). Moreover, the inequalities in (8.22) are direct implications of [35, Theorem 4-(b)], where in the second inequality, only the function $h_j(\cdot)$ has been considered as the remainder function, but in the third inequality, both of the $h_j(\cdot)$ and $g_j(\cdot) \triangleq f_j(\cdot) - h_j(\cdot)$ have been considered as remainders, separately, and the minimum of the obtained values has been taken. □

Proof of Lemma 8.3.4

Consider $l_1^j(m) \triangleq \Delta h^c_{j;m} + f_j(\underline{z}^c_{j;m})$ in the expression for the error lower bound $\underline{q}_j(W, V; h_j) \triangleq \frac{1}{2}[\min_{m \in \mathbf{M}_j^c} l_1^j(m) + \min_{m \in \mathbf{M}_j^c} l_2^j(m) + \Delta f_j^{\text{true}}]$, in Theorem 8.20, with $h_j(\cdot) \in \mathcal{H}_{\mathbf{M}_j}$ being an arbitrary remainder function with the corresponding set of supporting vectors $\mathbf{M}_j^c$. Now, applying the Mean Value Theorem, $\forall m \in \mathbf{M}_j^c, \exists \zeta \in \mathcal{Z}$ such that the following holds:

$$\Delta h^c_{j;m} = h_j(\overline{z}^c_{j;m}) - h_j(\underline{z}^c_{j;m}) = \nabla^\top_{h_j}(\zeta)(\overline{z}^c_{j;m} - \underline{z}^c_{j;m}). \tag{A.93}$$

Recall that $\overline{z}^c_{j;m} \triangleq z^c_m(\overline{z}, \underline{z}) = D^m \overline{z} + (I_{n_z} - D^m)\underline{z}, \underline{z}^c_{m;j} \triangleq z^c_m(\underline{z}, \overline{z}) = D^m \underline{z} + (I_{n_z} - D^m)\overline{z}, D^m_{i,i} = \text{sgn}(\min(\underline{a}_i^{f_j}, 0) - m_i), \forall i \in \{1, \ldots, n_z\}$ and $\nabla^\top_{h_j}(\zeta) = [\frac{\partial h_j}{\partial z_1} \cdots \frac{\partial h_j}{\partial z_{n_z}}]$. Given these, for $i = 1, \ldots, n_z$, we can consider two cases (similar to the cases in the proof of Proposition A.0.4):

- $m_i \leq \min(\underline{a}_i^{f_j}, 0)$ and hence, $m_i \leq \frac{\partial h_j}{\partial z_i} \leq 0$ and also $D^m_{i,i} = 1$ and so $\overline{z}^c_{j;m}(i) = \overline{z}_i \geq \underline{z}_i = \underline{z}^c_{j;m}(i)$, that implies $\overline{z}^c_{j;m}(i) - \underline{z}^c_{j;m}(i) = \overline{z}_i - \underline{z}_i \geq 0$. Therefor,
$$m_i(\overline{z}_i - \underline{z}_i) \leq \frac{\partial h_j}{\partial z_i}(\overline{z}^c_{j;m}(i) - \underline{z}^c_{j;m}(i)).$$
- $m_i \geq \max(\overline{a}_i^{f_j}, 0)$, and hence, $m_i \geq \frac{\partial h_j}{\partial z_i} \geq 0$ and also $D^m_{i,i} = 0$ and so $\overline{z}^c_{j;m}(i) = \underline{z}_i \leq \overline{z}_i = \underline{z}^c_{j;m}(i)$, that implies $\overline{z}^c_{j;m}(i) - \underline{z}^c_{j;m}(i) = \underline{z}_i - \overline{z}_i \leq 0$. Therefor,
$$m_i(\overline{z}_i - \underline{z}_i) \leq \frac{\partial h_j}{\partial z_i}(\overline{z}^c_{j;m}(i) - \underline{z}^c_{j;m}(i)).$$

Considering the above two cases for $i = 1, \ldots, n_z$, we conclude that $l_1^j(m) \geq m^\top(\overline{z} - \underline{z}) + f_j(\underline{z}^c_{j;m})$, where the right hand side of the inequality, is the error that is achieved using the linear remainder $\tilde{h}_j(z) = m^\top(\overline{z} - \underline{z})$. Hence, given m, no remainder function can attain smaller value for $l_1^j(m)$ than a linear one. Similar reasoning shows the same conclusion for $l_2^j(m) \triangleq \Delta h^c_{j;m} - f_j(\overline{z}^c_{j;m})$, which concludes the proof. □

Proof of Theorem 8.3.6

The result in (8.27) is obtained by (8.24) and the fact that the remainder function $h_j(\cdot)$ is locally Lipschitz, i.e, is bounded gradient. Particularly, for linear remainder $h(z) = Mz$, with $M^\top_{(j,:)} = m^j$, it can be easily verified that with choosing $\beta^f_R = \max_{j \in \mathcal{J}} \|m^j\|_\infty$, (8.27) holds. The proof of (8.28) goes along the lines of the proof of [98, Theorem 4.1]. □

Proof of Lemma 8.3.7

Obviously $\mathcal{Z}^* \subseteq \mathcal{Z}^p$, since by construction (cf. Algorithm 8), $\underline{z}^p \leq \underline{z}^*$ and $\overline{z}^p \geq \overline{z}^*$. We are required to show that $\mathcal{Z}^* \supseteq \hat{\mathcal{Z}} \triangleq \{z \in \mathcal{Z}^p | \underline{y} \leq \mu(z) \leq \overline{y}\}$. To use contradiction, suppose not. Then, $\exists \zeta \in \hat{\mathcal{Z}}$ such that $\zeta \notin \mathcal{Z}^*$, i.e., $\exists i \in \{1, \ldots, n_z\}$ such that $\zeta_i > \overline{z}^*_i$ or $\zeta_i < \underline{z}^*_i$. Without loss of generality, suppose the first case holds, i.e., $\zeta_i > \overline{z}^*_i$ (the reasoning for the other case would be similar). Then, $\zeta \in [\underline{z}^m, \overline{z}^p]$, where $\underline{z}^m_i = \overline{z}^*_i$ and $\underline{z}^m_s = \underline{z}^p_s, \forall s \neq i$. Hence,

$$\underline{\mu}_{d,R}(\underline{z}^m, \overline{z}^p) \leq \mu(\zeta) \leq \overline{\mu}_{d,R}(\overline{z}^p, \underline{z}^m), \tag{A.94}$$

where $\overline{\mu}_{d,R}(\cdot,\cdot)$ and $\underline{\mu}_{d,R}(\cdot,\cdot)$ are the proposed upper and lower remainder-form decomposition functions in Algorithm 7. On the other hand, note that $\mathcal{Z}^* \cap [\underline{z}^m, \overline{z}^p] = \emptyset$, hence the interval $[\underline{z}^m, \overline{z}^p]$ has been "ruled out" by the Algorithm 8. In other words, one of the "or" conditions in the line 5 of the Algorithm 8 must hold for this interval, i.e.,

$$\overline{\mu}_{d,R}(\overline{z}^p, \underline{z}^m) < \underline{y} \ \text{ or } \ \underline{\mu}_{d,R}(\underline{z}^m, \overline{z}^p) < \overline{y}. \tag{A.95}$$

From (A.94) and (A.95) we have $\mu(\zeta) < \underline{y}$ or $\mu(\zeta) > \overline{y}$. This contradicts $\zeta \in \hat{\mathcal{Z}} \Leftrightarrow \underline{y} \leq \mu(\zeta) \leq \overline{y}$. ☐

Proof of Theorem 8.4.1

(i) By choosing $h_j(z) = \mathbf{m}^{j\top} z$ and

$$\mathbf{m}^j_i = \begin{cases} \min(\underline{a}^{f_j}_i, 0) \text{ if } |\underline{a}^{f_j}_i| \leq |\overline{a}^{f_j}_i| \\ \max(\overline{a}^{f_j}_i, 0) \text{ if } |\underline{a}^{f_j}_i| > |\overline{a}^{f_j}_i| \end{cases}, \forall i = 1, \ldots, n_z, \tag{A.96}$$

one can observe that ζ and $(\alpha_j - \beta_j)^\top (z - \hat{z})$ in (8.4) coincide with $z^c_{\mathbf{m}^j}(z, \hat{z})$ and $h_j(z^c_{\mathbf{m}^j}(\hat{z}, z)) - h_j(z^c_{\mathbf{m}^j}(z, \hat{z})) = \mathbf{m}^{j\top}(z^c_{\mathbf{m}^j}(\hat{z}, z) - z^c_{\mathbf{m}^j}(z, \hat{z}))$ in (8.10) and hence the decomposition function in Proposition 8.1.13, coincides with one of the decomposition functions in the family (8.10).

(ii) The proof goes along the lines of the proof of Lemma 8.3.4, showing that $\Delta h^c_{j;m} = h_j(\overline{z}^c_{j;\mathbf{m}}) - h_j(\underline{z}^c_{j;\mathbf{m}}) \geq \mathbf{m}^{j\top}(z^c_{\mathbf{m}^j}(\hat{z}, z) - z^c_{\mathbf{m}^j}(z, \hat{z}))$, when $\mathbf{m}^j$ is chosen as in (A.96).

(iii) The result is a direct implication of (i) and Lemma 8.3.4. ☐

Proof of Proposition 8.5.1

The results directly follow form combining corollaries 8.1.6–8.1.9, Theorem 8.3.1, Lemmas 8.3.4–8.3.7 and applying algorithms 7–8. ☐

Proof of Proposition 9.1.2

To prove (9.1), consider $z \in \mathbb{IZ} \Leftrightarrow \underline{z} \leq z \leq \overline{z} \Leftrightarrow \underline{z} - \mathrm{mid}(\mathbb{IZ}) \leq z - \mathrm{mid}(\mathbb{IZ}) \leq \overline{z} - \mathrm{mid}(\mathbb{IZ}) \Leftrightarrow -\frac{1}{2}\mathrm{diam}(\mathbb{IZ}) \leq z - \mathrm{mid}(\mathbb{IZ}) \leq \frac{1}{2}\mathrm{diam}(\mathbb{IZ}) \Leftrightarrow \mathrm{mid}(\mathbb{IZ}) - \frac{1}{2}\mathrm{diam}(\mathbb{IZ}) \leq z \leq \frac{1}{2}\mathrm{diam}(\mathbb{IZ}) + \mathrm{mid}(\mathbb{IZ}) \Leftrightarrow \exists \xi \in \mathbb{B}_\infty^n, s.t.\ z = \mathrm{mid}(\mathbb{IZ}) + \frac{1}{2}\mathrm{diag}(\mathrm{diam}(\mathbb{IZ}))\xi \Leftrightarrow z \in \mathrm{mid}(\mathbb{IZ}) \oplus \frac{1}{2}\mathrm{diag}(\mathrm{diam}(\mathbb{IZ}))\mathbb{B}_\infty^n$. The result in (9.2) is a straightforward extension of (9.1). □

Proof of Corollary 9.1.7

The proof follows the lines of the proof of [63, Lemma 1, Proposition 10 and Corollary 2]. □

Proof of Lemma 9.3.1

To show (9.13), $\forall s \in \mathbb{N}_S$, consider the zonotope $\mathcal{Z}_s \triangleq \{G_s, c_s\}_Z \triangleq \{z = G_s\xi + c_s | \xi \in \mathbb{B}_\infty^{n_s}\}$ and let us define $\tilde{f}_s(\xi) : \mathbb{B}_\infty^{n_s} \to \mathbb{R}^{n_x} \triangleq f(G_s\xi + c_s)$ which implies that

$$f(\mathcal{Z}_s) \subseteq \tilde{f}_s(\mathbb{B}_\infty^{n_s}), \forall s \in \mathbb{N}_S. \tag{A.97}$$

On the other hand, note that by Corollary 9.1.7, $\forall H_s \in \mathbf{H}_{\tilde{f}_s}$, $\tilde{f}_s(\cdot)$ can be decomposed as

$$\tilde{f}_s(\xi) = g_s^{H_s}(\xi) + H_s\xi, \forall s \in \mathbb{N}_S, \forall \xi \in \mathbb{B}_\infty^{n_s}, \forall H_s \in \mathbf{H}_{\tilde{f}_s} \tag{A.98}$$

where $g_s^{H_s}(\xi)$ is a JSS function in $\mathbb{B}_\infty^{n_s}$ and $\mathbf{H}_{\tilde{f}_s}$ can be computed from (9.9), with the corresponding function being $\tilde{f}_s$. Now (A.97) and (A.98) together imply:

$$f(\mathcal{Z}_s) \subseteq g_s^{H_s}(\mathbb{B}_\infty^{n_s}) \oplus H_s\mathbb{B}_\infty^{n_s}, \forall s \in \mathbb{N}_S, \forall H_s \in \mathbf{H}_{\tilde{f}_s}. \tag{A.99}$$

Again, it follows from Corollary 9.1.7 and the fact that $g_s^{H_s}(\xi)$ is a JSS function that in each dimension $i \in \mathbb{N}_{n_x}$, $g_{s,i}^{H_s}(\xi)$ can be tightly bounded as $\underline{g}_{s,i}^{H_s} \leq g_{s,i}^{H_s}(\xi) \leq \overline{g}_{s,i}^{H_s}, \forall \xi \in \mathbb{B}_\infty^{n_s}, \forall H_s \in \mathbf{H}_{\tilde{f}_s}$, with $\overline{g}_{s,i}^{H_s}, \underline{g}_{s,i}^{H_s}$ given in (9.14) and (9.15), respectively. Augmenting all these $\mathbb{N}_{nx}$ one-dimensional inequalities yields the following set inclusion for all $s \in \mathbb{N}_S$ and all $H_s \in \mathbf{H}_{\tilde{f}_s}$: $g_s^{H_s}(\mathbb{B}_\infty^{n_s}) \subseteq [\underline{g}_s^{H_s}, \overline{g}_s^{H_s}] = \frac{1}{2}((\underline{g}_s^{H_s} + \overline{g}_s^{H_s}) \oplus \mathrm{diag}(\overline{g}_s^{H_s} - \underline{g}_s^{H_s})\mathbb{B}_\infty^{n_x})$, where the last equality follows from Proposition 9.1.2. Combining this, (A.99) and the fact that the inclusion in (A.99) holds for all $s \in \mathbb{N}_S$ and all $H_s \in \mathbf{H}_{\tilde{f}_s}$ and hence for the intersection of all of them, we obtain (9.13). □

Proof of Lemma 9.3.2

To prove the inclusion in (9.16), consider the constrained zonotope representation of the set $\mathcal{Z}$, i.e., $\mathcal{Z} \triangleq \{\tilde{G}, \tilde{c}, \tilde{A}, \tilde{b}\}_{CZ} \triangleq \{z = \tilde{G}\xi + c | \xi \in \mathbb{B}_\infty^{n_g}, \tilde{A}\xi = \tilde{b}\}$. Using similar notation as in the proof of Lemma 9.3.1, let us define $\tilde{f}(\xi) : \mathbb{B}_\infty^{n_g} \to \mathbb{R}^{n_x} \triangleq f(\tilde{G}\xi + \tilde{c})$ that consequently returns

$$f(\mathcal{Z}) \subseteq \{\tilde{f}(\xi) \mid \xi \in \mathbb{B}_\infty^{n_g}, \tilde{A}\xi = \tilde{b}\}. \tag{A.100}$$

Note that by [53, Theorem 2], $\tilde{A}\xi = \tilde{b} \Rightarrow \xi \in \mathbb{I}\Xi \triangleq [\tilde{A}^{\dagger}\tilde{b} - \kappa \mathbf{r}_{n_g}, \tilde{A}^{\dagger}\tilde{b} - \kappa \mathbf{r}_{n_g}]$, where $\mathbf{r}_{n_g} \triangleq \text{rowsupp}(I_{n_g} - \tilde{A}^{\dagger}\tilde{A})$ and κ is a very large positive real number. Combining this with the fact that $\xi \in \mathbb{B}_{\infty}^{n_g}$ (cf. (A.100)), imply that $\xi \in \mathbb{I}\tilde{\Xi} \triangleq \mathbb{I}\Xi \cap \mathbb{B}_{\infty}^{n_g} = [\underline{\mathbf{l}}_{n_g}, \bar{\mathbf{l}}_{n_g}]$, where $\underline{\mathbf{l}}_{n_g}, \bar{\mathbf{l}}_{n_g}$ are defined below (9.18). On the other hand, similar to the proof of Lemma 9.3.1, we conclude by Corollary 9.1.7 that $\forall H \in \mathbf{H}_{\tilde{f}}$, $\tilde{f}(\cdot)$ can be decomposed as

$$\begin{aligned} \tilde{f}(\xi) \quad &= \tilde{g}^H(\xi) + H\xi, \quad \forall H \in \mathbf{H}_{\tilde{f}}, \forall \xi \in \mathbb{I}\tilde{\Xi}, \\ \Rightarrow \tilde{f}(\mathbb{I}\tilde{\Xi}) \quad &\subseteq \tilde{g}^H(\mathbb{I}\tilde{\Xi}) \oplus H\mathbb{I}\tilde{\Xi}, \forall H \in \mathbf{H}_{\tilde{f}}, \end{aligned} \tag{A.101}$$

where $\tilde{g}^H(\xi)$ is a JSS function in $\mathbb{I}\tilde{\Xi}$ and $\mathbf{H}_{\tilde{f}}$ is given in (9.9). By Corollary 9.1.7, in each dimension $i \in \mathbb{N}_{n_x}$, $\tilde{g}_i^H(\xi)$ can be tightly bounded as $\underline{g}_i^H \leq \tilde{g}_i^H(\xi) \leq \overline{g}_i^H, \forall \xi \in \mathbb{I}\tilde{\Xi}, \forall H \in \mathbf{H}_{\tilde{f}}$, with $\overline{g}_i^H, \underline{g}_i^H$ given in (9.17) and (9.18), respectively. Augmenting all these $\mathbb{N}_{nx}$ one-dimensional inequalities and applying Proposition 9.1.2 yield the following set inclusion:

$$\forall H \in \mathbf{H}_{\tilde{f}},\ \tilde{g}^H(\mathbb{I}\tilde{\Xi}) \subseteq [\underline{g}^H, \overline{g}^H] = \frac{1}{2}((\underline{g}^H + \overline{g}^H) \oplus \text{diag}(\overline{g}^H - \underline{g}^H)\mathbb{B}_{\infty}^{n_x}).$$

Combining this, (A.100), (A.101) and the fact that the inclusion in (A.101) holds for all $H \in \mathbf{H}_{\tilde{f}}$ and hence for the intersection of all of them, we obtain $f(\mathcal{Z}) \subseteq \{H\xi + \text{diag}(\overline{g}^H - \underline{g}^H)\theta + \frac{1}{2}((\underline{g}^H + \overline{g}^H) \mid \xi \in \mathbb{B}_{\infty}^{n_g}, \theta \in \mathbb{B}_{\infty}^{n_x}, \tilde{A}\xi = \tilde{b}\}, \forall H \in \mathbf{H}_{\tilde{f}}$, where the set on the right hand side of the inclusion is equivalent to the intersection of the CZs on the right hand side of (9.16). □

Proof of Theorem 9.3.3

It follows from Lemmas 9.3.1 and 9.3.2 that $f(\mathcal{Z}) \subseteq \mathcal{ZB}_f$ and $f(\mathcal{Z}) \subseteq \mathcal{CZ}_f$, and so $f(\mathcal{Z}) \subseteq \mathcal{ZB}_f \cap \mathcal{CZ}_f$. □

Proof of Lemma 9.3.4

Suppose $z \in \mathcal{ZB}_f \cap_{\mu} \mathcal{ZB}_{\mu}$. Then, by definition of the operator $\cap_{\mu}$ (cf. (9.12)), $z \in \mathcal{ZB}_f$ and $\mu(z) \in \mathcal{ZB}_{\mu}$. The former implies that $\forall r \in \mathbb{N}_R, \exists \alpha \in \mathbb{B}_{\infty}^{n_r}$ such that $z = G_f^r\alpha + c_f^r$, while it follows from the latter that $\mu(z) = \mu(G_f^r\alpha + c_f^r) \triangleq \tilde{\mu}_r(\alpha) \in \mathcal{ZB}_{\mu} \Rightarrow \forall t \in \mathbb{N}_T, \exists \zeta \in \mathbb{B}_{\infty}^{n_t}$, such that $\tilde{\mu}_r(\alpha) = c_{\mu}^t + G_{\mu}^t\zeta$. Putting these two results in a set representation form, we obtain:

$$z \in \bigcap_{r=1}^{R}\bigcap_{t=1}^{T}\{G_f^r\alpha + c_f^r | \tilde{\mu}_r(\alpha) = c_{\mu}^t + G_{\mu}^t\zeta, \alpha \in \mathbb{B}_{\infty}^{n_r}, \zeta \in \mathbb{B}_{\infty}^{n_t}\}. \tag{A.102}$$

On the other hand, using Corollary 9.1.7, $\tilde{\mu}_r(\cdot)$ can be decomposed into a JSS and a linear mapping as follows: $\forall r \in \mathbb{N}_R, \forall Q_r \in \mathbf{Q}_{\tilde{\mu}_r}, \forall \alpha \in \mathbb{B}_{\infty}^{n_r}$:

$$\tilde{\mu}_r(\alpha) = p_r^{Q_r}(\alpha) + Q_r\alpha. \tag{A.103}$$

Moreover, by the same corollary, the JSS component $p_r^{Q_r}(\cdot)$ is tightly bounded as follows: $\forall i \in \mathbb{N}_{n_\mu}, \forall Q_r \in \mathbf{Q}_{\tilde{\mu}_r}, \underline{p}_{r,i}^{Q_r} \leq p_{r.i}^{Q_r}(\alpha) \leq \underline{p}_{r,i}^{Q_r}, \forall \alpha \in \mathbb{B}_\infty^{n_r}$, with $\underline{p}_{r,i}^{Q_r}, \overline{p}_{r,i}^{Q_r}$ given in (9.20) and (9.21), respectively. Combining this, as well as (A.103) and Proposition 9.1.2 results in: $\forall r \in \mathbb{N}_R, \forall Q_r \in \mathbf{Q}_{\tilde{\mu}_r}, \forall \alpha \in \mathbb{B}_\infty^{n_r}, \exists \theta \in \mathbb{B}_\infty^{n_\mu}$ such that $\tilde{\mu}_r(\alpha) = \frac{1}{2}(\underline{p}_{r,i}^{Q_r} + \overline{p}_{r,i}^{Q_r}) + \frac{1}{2}\text{diag}(\overline{p}_{r,i}^{Q_r} - \underline{p}_{r,i}^{Q_r})\theta + Q_r\alpha$. Further, putting this together with (A.102) returns $z \in \bigcap_{r=1}^{R} \bigcap_{t=1}^{T} \bigcap_{Q_t \in \mathbf{Q}_{\tilde{\mu}_t}} \{G_f^r\alpha + c_f^r | \frac{1}{2}(\underline{p}_{r,i}^{Q_r} + \overline{p}_{r,i}^{Q_r}) + \frac{1}{2}\text{diag}(\overline{p}_{r,i}^{Q_r} - \underline{p}_{r,i}^{Q_r})\theta + Q_r\alpha = c_\mu^t + G_\mu^t\zeta, \alpha \in \mathbb{B}_\infty^{n_r}, \zeta \in \mathbb{B}_\infty^{n_t}, \theta \in \mathbb{B}_\infty^{n_\mu}\}$, where the set on the right hand side is equivalent to the one on the right hand side of (9.19). □

Proof of Lemma 9.3.5

Suppose $z \in \mathcal{CZ}_f \cap_\mu \mathcal{CZ}_\mu$. Then, by definition of the operator $\cap_\mu$ (cf. (9.12)), $z \in \mathcal{CZ}_f$ and $\mu(z) \in \mathcal{CZ}_\mu$. The former implies that $\exists \beta \in \mathbb{B}_\infty^{n_c}$ such that $\tilde{A}_f\beta = \tilde{b}_f \wedge z = \tilde{G}_f\beta + \tilde{c}_f$, while it follows from the latter that $\mu(z) = \mu(\tilde{G}_f\beta + \tilde{c}_f) \triangleq \lambda(\beta) \in \mathcal{CZ}_\mu \Rightarrow$ $, \exists \gamma \in \mathbb{B}_\infty^{n_\tau}$, such that $\tilde{A}_\mu\gamma = \tilde{b}_\mu$ and $\lambda(\beta) = \tilde{c}_\mu + \tilde{G}_\mu\gamma$. Putting these two results into a set representation form, we obtain:

$$z \in \{\tilde{G}_f\beta + \tilde{c}_f | \lambda(\beta) = \tilde{c}_\mu + tildeG_\mu\gamma, \tilde{A}_f\beta = \tilde{b}_f, \tilde{A}_\mu\gamma = \tilde{b}_\mu, \beta \in \mathbb{B}_\infty^{n_c}, \gamma \in \mathbb{B}_\infty^{n_\tau}\} \tag{A.104}$$

On the other hand, using Corollary 9.1.7, $\lambda(\cdot)$ can be decomposed into a JSS and a linear mapping as follows:

$$\forall \Omega \in \mathbf{\Omega}_\lambda, \forall \beta \in \mathbb{B}_\infty^{n_c} : \lambda(\beta) = \nu^\Omega(\beta) + \Omega\beta. \tag{A.105}$$

Further, note that by [53, Theorem 2], $\tilde{A}_f\beta = \tilde{b}_f \Rightarrow \beta \in \mathbb{IB} \triangleq [\tilde{A}_f^\dagger\tilde{b}_f - \kappa\mathbf{r}_{n_c}, \tilde{A}_f^\dagger\tilde{b}_f - \kappa\mathbf{r}_{n_c}]$, where $\mathbf{r}_{n_c} \triangleq \text{rowsupp}(I_{n_c} - \tilde{A}_f^\dagger\tilde{A}_f)$ and κ is a very large positive real number. Then, since $\beta \in \mathbb{B}_\infty^{n_c}$, we have $\beta \in \mathbb{IB} \cap \mathbb{B}_\infty^{n_c} = [\underline{\mathbf{l}}_{n_c}, \bar{\mathbf{l}}_{n_c}]$. Putting this and Corollary 9.1.7 together results in the JSS component $\nu^\Omega(\cdot)$ being tightly bounded, i.e., $\forall i \in \mathbb{N}_{n_\mu}, \forall \Omega \in \mathbf{\Omega}_\lambda, \underline{\nu}_i^\Omega \leq \nu_i^\Omega(\beta) \leq \underline{\nu}_i^\Omega, \forall \beta \in \mathbb{B}_\infty^{n_c}$, with $\underline{\nu}_i^\Omega, \overline{\nu}_i^\Omega$ given in (9.23) and (9.24), respectively. Combining this, (A.105) and Proposition 9.1.2 leads to: $\forall \Omega \in \mathbf{\Omega}_\lambda, \forall \beta \in \mathbb{B}_\infty^{n_c}, \exists \rho \in \mathbb{B}_\infty^{n_\mu}$ such that $\lambda(\beta) = \frac{1}{2}(\underline{\nu}^\Omega + \overline{\nu}^\Omega) + \frac{1}{2}\text{diag}(\overline{\nu}^\Omega - \underline{\nu}^\Omega)\rho + \Omega\alpha$, which along with (A.104) returns $z \in \bigcap_{\Omega \in \mathbf{\Omega}_\lambda} \{\tilde{G}_f\beta + \tilde{c}_f | \frac{1}{2}(\underline{\nu}^\Omega + \overline{\nu}^\Omega) + \frac{1}{2}\text{diag}(\overline{\nu}^\Omega - \underline{\nu}^\Omega)\rho + \Omega\beta = \tilde{c}_\mu + \tilde{G}_\mu\gamma, \tilde{A}_f\beta = \tilde{b}_f, \tilde{A}_\mu\gamma = \tilde{b}_\mu, \beta \in \mathbb{B}_\infty^{n_c}, \gamma \in \mathbb{B}_\infty^{n_\tau}, \rho \in \mathbb{B}_\infty^{n_\mu}\}$, where the set on the right hand side is equivalent to the one on the right hand side of (9.19). □

Proof of Theorem 9.3.6

By Lemmas 9.3.4 and 9.3.5: $\mathcal{Z}_f \cap_\mu \mathcal{Z}_\mu \subseteq \mathcal{ZB}_u$ and $\mathcal{Z}_f \cap_\mu \mathcal{Z}_\mu \subseteq \mathcal{CZ}_u$, and hence $\mathcal{Z}_f \cap_\mu \mathcal{Z}_\mu \subseteq \mathcal{ZB}_u \cap \mathcal{CZ}_u$. □

Proof of Proposition 9.3.7

This directly follows from Proposition 9.1.6. □

Proof of Lemma 9.3.8

Let $z \in \mathcal{CZ}_f \cap_\mu \mathcal{CZ}_\mu$. Then by definition of the operator $\cap_\mu$ (cf. (9.12)), $z \in \mathcal{CZ}_f$ and $\mu(z) \in \mathcal{CZ}_\mu$. Further, by Proposition 9.1.2 and the mean value theorem, $z \in \mathcal{CZ}_f$ implies that $\mu(z) \in \mu(\mathcal{CZ}_f) \subseteq \mu(x_0) \oplus \mathbb{J}^\mu(\mathcal{CZ}_f \ominus x_0)$, where

$$\begin{aligned}
&\mu(x_0) \oplus \mathbb{J}^\mu(\mathcal{CZ}_f \ominus x_0) \\
&= \mu(x_0) \oplus (\mathrm{mid}(\mathbb{J}^\mu) + \mathbb{J}^\mu_\Delta)(\mathcal{CZ}_f \ominus x_0) \\
&= (\mu(x_0) - \mathrm{mid}(\mathbb{J}^\mu)x_0) \oplus \mathrm{mid}(\mathbb{J}^\mu)\mathcal{CZ}_f \oplus \mathbb{J}^\mu_\Delta(\mathcal{CZ}_f \ominus x_0).
\end{aligned} \tag{A.106}$$

On the other hand, by Proposition 9.1.3 :

$$\mathbb{J}^\mu_\Delta(\mathcal{CZ}_f \ominus x_0) \subseteq \mathcal{CZ}_R \triangleq \{G_R, c_R, A_R, b_R\}_{CZ}, \tag{A.107}$$

with G_R, c_R, A_R, b_R given in (9.25) and (9.27) (note that $\mathrm{mid}(\mathbb{J}^\mu_\Delta) = 0$ by its definition) and where $\mathcal{CZ}_R$ has n_R generators. Then, the facts that $z \in \mathcal{CZ}_f \triangleq \{\tilde{G}_f\beta + \tilde{c}_f | \tilde{A}_f\beta = \tilde{b}_f, \beta \in \mathbb{B}^{n_c}_\infty\}$, $\mu(z) \in \mathcal{CZ}_\mu \triangleq \{\tilde{G}_\mu\gamma + \tilde{c}_\mu | \tilde{A}_\mu\gamma = \tilde{b}_\mu, \gamma \in \mathbb{B}^{n_\tau}_\infty\}$, along with (A.106) and (A.107) imply that $z \in \{\tilde{G}_f\beta + \tilde{c}_f | \tilde{c}_\mu + \tilde{G}_\mu\gamma = \mu(x_0) + \mathrm{mid}(\mathbb{J}^\mu)(\tilde{c}_f - x_0) + \mathrm{mid}(\mathbb{J}^\mu)\beta + C_R + G_R\xi_R, \tilde{A}_f\beta = \tilde{b}_f, \tilde{A}_\mu\gamma = \tilde{b}_\mu, A_R\xi_R = b_R, \beta \in \mathbb{B}^{n_c}_\infty, \gamma \in \mathbb{B}^{n_\tau}_\infty, \xi_R \in \mathbb{B}^{n_R}_\infty\}$, where the set on the right hand side is equivalent to the CZ on the right hand side of (9.26). $\square$