Set-Theoretic Frameworks for Online Optimization, Estimation, and Control

by

DIGANTA BHATTACHARJEE

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
August 2021

Dedicated to the memory of my little sister Moumi

Acknowledgements

This dissertation would have been inconceivable without support, generosity, and guidance from many people. First of all, I would like to express my sincere appreciation and heartfelt gratitude towards my advisor Dr. Kamesh Subbarao for his continuous support, mentorship, and unlimited patience with me over the last four years. His insightful comments during our many discussions on various aspects of academic research were instrumental in my gradual development and enabled me to broaden my general understanding of academia. He has had a profound academic influence on me, and I am extremely grateful for everything he has taught me.

I wholeheartedly thank Dr. Animesh Chakravarthy, Dr. Ramtin Madani, Dr. Alan Bowling, Dr. Ashfaq Adnan for all their support and for taking the time to serve on my doctoral dissertation committee. I am deeply indebted to Dr. Chakravarthy for the opportunity to work on one of his projects across a couple of semesters and for kindly funding my research during that period as well. I have collaborated with him on a key component of this dissertation, and I am grateful for the freedom he allowed me to express my ideas. Furthermore, my formal introduction to convex optimization, a topic that went on to become an integral part of my doctoral research, was through the class I took with Dr. Madani, and I am grateful to him for the lessons. My doctoral research has been funded in part by the Office of Naval Research (award no. N00014-18-1-2215), the National Science Foundation (award no. IIS-1851817), and the US Coast Guard (award no. MSRDC #D01_W911SR-14-2-0001-(OR #0042)), and I appreciate the support. In this regard, I wish to extend my sincere gratitude to

Dr. Kiran Bhaganagar (the University of Texas at San Antonio) for the opportunity to work on the US Coast Guard project.

Many thanks to the entire administrative staff at the Department of Mechanical and Aerospace Engineering for making my stay a pleasant one. Especially, I would like to extend my sincere thanks to Lanie Gordon, Ayesha Fatima, and Wendy Ryan for all their help.

I am thankful to my past and present lab mates at the Aerospace Systems Laboratory for their friendship, words of advice, constant support and encouragement. Especially, Rajnish and Ameya, thank you for your invaluable friendship and, even at times, for putting up with me.

My family has always been supportive of my decisions in life, and I am thankful for that. It gives me great pleasure to express my gratitude towards my dear mother, Mira Bhattacharjee, whose constant love and care kept me going through some tough times, and towards my dear father, Rathindra Nath Bhattacharjee, whom I have always regarded as my 'friend, philosopher, and guide' in life, for encouraging me to think critically and instilling in me the curiosity of science. Finally, I am fortunate enough to have been able to share this entire journey with Sarmistha. I want to thank her from the bottom of my heart for cheering me up whenever I felt low, for rejoicing in my success as her own, and above all, for her constant love and understanding throughout my stay at UTA.

July 30, 2021

Abstract

Set-Theoretic Frameworks for Online Optimization, Estimation, and Control

Diganta Bhattacharjee, Ph.D.

The University of Texas at Arlington, 2021

Supervising Professor: Kamesh Subbarao

This research is primarily focused on developing online frameworks, which are suited to real-time implementation, for performance optimization, estimation, and control of dynamical systems using set-theoretic concepts and/or having set-theoretic interpretations. First, two perturbation-based extremum seeking control schemes, based on the classical setup and equipped with novel adaptation laws for the perturbation signal amplitudes, are proposed for general single input single output nonlinear systems. The proposed schemes are able to extremize steady-state system output in a practical asymptotic sense, i.e., the system is driven to an arbitrarily small set centered at the true steady-state optimal operating point. The next development involves a semi-analytical model for avian-scale (or bird-scale) forward flapping flight. Results generated through this model indicate optimal characteristics of force generation in the unique range of Strouhal numbers used by birds for cruising. Then, constructive arguments are provided leading up to a hypothesis which postulates that birds use some form of online optimization for converging to this unique range during a flight. The hypothesis is investigated using one of the proposed extremum seeking control schemes as the optimization framework.

Furthermore, a novel set-membership state estimation algorithm using state dependent coefficient parameterization for discrete-time nonlinear systems is developed, and it requires solutions to two semi-definite programs. A linear variant of this estimator is considered in the context of leader-follower multi-agent synchronization. A distributed protocol design is proposed to make the agents closely follow a leader's trajectory. Finally, model predictive control is applied to synthesize lateral acceleration commands of missiles for planar engagements. The guidance problem is converted into a recursive algorithm that does not require target acceleration information and involves solving for strictly convex quadratic programs. Detailed simulation results are used to both illustrate all the theoretical results and verify the hypothesis.

Table of Contents

Appendix

List of Illustrations

List of Tables

List of Notations

$\mathbb{R}_{\geq 0}$	set of non-negative real numbers
$\mathbb{Z}_{\star}$	set of non-negative integers
$\boldsymbol{X} > 0$, $\boldsymbol{X} \geq 0$	square symmetric matrix $\boldsymbol{X}$ that is positive definite and positive semi-definite, respectively
$\boldsymbol{X} < 0$, $\boldsymbol{X} \leq 0$	square symmetric matrix $\boldsymbol{X}$ that is negative definite and negative semi-definite, respectively
$\sigma_{\max}(\boldsymbol{Y})$	maximum singular value of any matrix $\boldsymbol{Y}$
$C(a, b)$	open circle of radius b in the complex plane, centered at $a \in \mathbb{R}$
$\boldsymbol{I}_n$, $\boldsymbol{O}_n$, $\boldsymbol{O}_{m \times n}$	$n \times n$ identity matrix, $n \times n$ null matrix, and $m \times n$ null matrix, respectively
$\mathbf{1}_n$, $\mathbf{0}_n$	vector of ones of dimension n and vector of zeros of dimension n, respectively
diag$(\cdot)$	block-diagonal matrices
trace$(\cdot)$, rank$(\cdot)$	trace and rank of a matrix, respectively
$(\cdot)^{\mathrm{T}}$	vector or matrix transpose
$(\cdot)_0$	initial conditions or values
$\mathcal{E}(\boldsymbol{c}, \boldsymbol{P})$	ellipsoidal set $\{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \boldsymbol{c})^{\mathrm{T}} \boldsymbol{P}^{-1} (\boldsymbol{x} - \boldsymbol{c}) \leq 1\}$ with $\boldsymbol{c} \in \mathbb{R}^n$ as the center of the ellipsoid and $\boldsymbol{P} > 0$ as the *shape matrix* that characterizes the orientation and size of the ellipsoid in $\mathbb{R}^n$

$\mathrm{col}(\cdot)$	vector concatenation operator, i.e., given vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$, $\mathrm{col}[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M] = [\boldsymbol{x}_1^{\mathrm{T}} \; \boldsymbol{x}_2^{\mathrm{T}} \; \ldots \; \boldsymbol{x}_M^{\mathrm{T}}]^{\mathrm{T}}$
$\otimes$	Kronecker product
$(f \circ g)$	composition of functions f and g
C^t, C^0	class of functions that are continuously differentiable t times and class of continuous functions, respectively

List of Acronyms

ESC	extremum seeking control
GPC	generalized predictive control
ISS	input-to-state stable
QP	quadratic program
MPC	model predictive control
NMPC	nonlinear model predictive control/controller
CC-NMPC	collision cone-based NMPC
PM-NMPC	point mass-based NMPC
SCQP	strictly convex quadratic program
SDC	state dependent coefficient
SDP	semi-definite program
SMF	set-membership filter
SDC-SMF	state dependent coefficient-based SMF

Executive Summary

This dissertation essentially deals with problems related to performance optimization, estimation, and control of dynamical systems. Solutions to these problems are provided in the form of online frameworks suitable for real-time implementation. The frameworks are based on set-theoretic concepts and/or have set-theoretic interpretations. Collectively, these are referred to as set-theoretic frameworks/methods in the dissertation. In this context, a set-theoretic framework for online optimization of a system's performance ensures that the system is driven to a (small) neighborhood of its optimal operating point by prescribing the control actions (or decision variables) online. Set-theoretic control, on the other hand, attempts to steer the states of the system to a target set while *explicitly* accounting for the state and control input constraints (i.e., the states and control inputs are restricted to some known sets). Similarly, set-theoretic estimation aims to construct set-based estimates for an estimation problem with all the uncertainties belonging to some known sets (i.e., the uncertainties are unknown but bounded). Furthermore, principles of online optimization (e.g., minimization of a suitable cost function online) can be integrated into set-theoretic control and estimation frameworks. Thus, if one takes this route, these frameworks will essentially become recursive algorithms. All of the above-mentioned aspects are studied in this research, borrowing concepts from the theories of nonlinear control and estimation, and convex optimization.

First, extremum seeking control (ESC), a data-driven (or model-free) online optimization strategy for optimizing steady-state system performance, has been studied. Two perturbation-based ESC schemes are proposed for general single input single output nonlinear dynamical systems, having structures similar to that of the classical ESC scheme. Novel adaptation laws are proposed for the excitation signal amplitudes in each scheme. These laws drive the amplitudes to zero, and the rates of decay for both the laws are governed by the gradient measures of the unknown reference-to-output equilibrium map which is the function to be optimized in this case. It is shown that the proposed ESC schemes achieve practical asymptotic convergence to the extremum with a proper tuning of the parameters in the schemes. Thus, the system converges arbitrarily close to the true optimal operating point. As the extremum is reached, and the magnitudes of the gradient measures become small, the excitation signal amplitudes converge to zero. Thus, the proposed schemes ensure that the excitation signal is driven to zero as the system output converges to a neighborhood of the extremum and the steady-state oscillations about the extremum, typically observed for the classical ESC schemes, are attenuated. Simulation examples are included to illustrate the capability of the proposed schemes to converge to the global extremum, bypassing local extrema in the process.

Next, optimal force generation in avian-scale (or bird-scale) flapping flight is investigated, largely focusing on the unique range of Strouhal numbers utilized by birds while cruising. A semi-analytical model for avian-scale forward flapping flight is constructed using the tools of quasi-steady aerodynamics and flight mechanics. Analytical expressions for cycle-averaged forces are derived, which reveal important insight into the parameters influencing force generation. Also, a detailed simulation study is conducted, the results of which reveal that cycle-averaged net thrust, lift and lift-to-drag ratio are optimal for cruising in the unique Strouhal number range. Furthermore, it is hypothesized that the in-flight mechanism employed by birds to achieve this optimization is a data-driven (or model-free) one. Another simulation study, which utilizes the semi-analytical model to calculate cycle-averaged lift-to-drag ratio (considered to be the objective function of the optimization) numerically for two different species of birds and implements one of the proposed extremum seeking schemes as the optimization framework, is performed. Results of this study illustrate that convergence to the unique Strouhal number range is possible by properly modulating the flapping frequency, thus verifying the hypothesis.

In terms of set-theoretic frameworks for estimation, the ellipsoidal state estimation problem for discrete-time nonlinear systems is considered. The nonlinear system is represented in a pseudo-linear form using the state dependent coefficient parameterization. Matrix Taylor expansions are utilized to expand the state dependent matrices about the state estimates. Upper bounds on the norms of remainders in the matrix Taylor expansions are calculated online using a non-adaptive random search algorithm at each time step. Utilizing these upper bounds and the ellipsoidal set description of the uncertainties, a two-step filter (or state estimator) utilizing the 'correction-prediction' structure of the standard Kalman Filter variants is derived. At each time step, the optimal (minimum 'size') correction and prediction ellipsoids are constructed that contain the true state of the system by solving the corresponding semi-definite programs, which are convex. Simulation results are included to illustrate performance of the proposed filter for a two-dimensional nonlinear system governed by the Van der Pol equation.

Subsequently, the above-mentioned filter is incorporated into a leader-follower synchronization protocol design for high-order discrete-time linear multi-agent systems, with the aim of making the agents synchronize with a leader. The agents are subject to unknown but bounded uncertainties. Each agent is considered to be equipped with the filter that deals with the uncertainties and estimates the state of the agent. It is assumed that the agents are able to share the state estimate information with the neighbors locally. This information is utilized in the local control law design for synchronization. Under appropriate conditions, the global disagreement error between the agents and the leader is shown to be bounded. An upper bound on the norm of the global disagreement error is calculated and shown to be monotonically decreasing. Detailed simulation results are included to illustrate various aspects of the proposed synchronization protocol.

A hybrid missile guidance algorithm is developed next for planar engagement scenarios. The guidance algorithm implements a set-theoretic control strategy, namely nonlinear model predictive control (NMPC). Also, the NMPC is combined with the collision cone approach which is a mathematical tool for guidance. This enables the guidance algorithm to achieve interception while ensuring that the impact angle lies in a predefined range. The guidance scheme comprises two components: (i) point mass-based NMPC and (ii) collision cone-based NMPC. The point mass-based NMPC is employed when the distance between the missile and the target is large, and when this distance falls below a threshold, the algorithm switches to the collision cone-based NMPC. Connections between the impact angle and the collision cone are theoretically established, and these are utilized in the collision cone-based NMPC. The NMPC problems are formulated as quadratic programs (QPs) that include appropriate constraints on the states and inputs, while not requiring target acceleration information. These QPs are shown to be strictly convex. Detailed simulation results are included to demonstrate successful target capture for a variety of initial engagement geometries and target acceleration profiles.

Chapter 1

Introduction

"Generally speaking, in the control theory context, all the techniques which are theoretically based on some properties of subsets of the state-space could be referred to as set-theoretic methods."

– F. Blanchini and S. Miani [1]

Set-theoretic techniques provide an array of mathematical tools appropriate for design and analysis tasks pertaining to dynamical systems [1]. These tasks include performance optimization, control synthesis, and state estimation of a system. This research considers carrying out these tasks in the form of *dynamic processes* that are suitable for real-time implementation. In other words, we are interested in developing *online frameworks*. It is worth pointing out that similar online frameworks in the existing literature generally require sufficient computational resources for real-time implementation. Now, in the last two decades, there has been a significant reduction in the cost of computation, and it has brought about a realistic opportunity to implement these computationally demanding frameworks in real-time. This serves as an inspiration for our research which aims to construct online frameworks for performance optimization, control, and state estimation of a single-agent system using set-theoretic tools and/or having set-theoretic implications. These are referred to as set-theoretic frameworks/methods in this dissertation. Moreover, multi-agent systems have received a great deal of attention in recent times due to their diverse range of applications and advantages over single-agent systems. This has inspired us to explore possible extensions of the set-theoretic frameworks to a group of systems (multi-agent system) that coordinate and cooperate with each other to achieve a collective goal.

In the context of this dissertation, set-theoretic frameworks for online optimization of a system's performance entail extremizing (i.e., minimizing or maximizing) the associated performance measure (cost or objective function) and often involve control actions such that the system is steered to a neighborhood of the optimal operating point (i.e., to a set centered at the optimal operating point). Set-theoretic frameworks for control are suitable if the design requirements involve steering the system's states to a target set while explicitly accounting for the state and control input constraints (a possible means to restrict these quantities to some known sets). Likewise, utilization of set-theoretic techniques would be natural for an estimation problem where all the uncertainties, although inherently unknown, belong to some known sets (i.e., unknown but bounded uncertainties). This approach of estimation, as one can imagine, would result in set-based estimates. Furthermore, online minimization of suitable cost functions can be integrated into set-theoretic control and estimation frameworks. Adopting this approach and using the convex optimization theory, one can convert the original control and estimation problems at every step into respective well-posed convex optimization problems. Thus, the control and estimation frameworks would eventually take the form of recursive algorithms. All of these aspects constitute the scope of this research.

1.1 Goals and Objectives

The goals and objectives of this research work are summarized in the following list:

- The first goal is to develop an online strategy for optimizing system performance: one that is applicable to a large class of dynamical systems and does not require the knowledge of an intricate mathematical model of the system. In order to achieve this goal, extremum seeking control, which is a data-driven (or model-free) online optimization strategy that optimizes steady-state system performance, has been studied. Additionally, the objective is to attenuate the steady-state oscillations typically observed for the classical ESC schemes in the literature.

- Our next goal is to investigate the unique range of Strouhal numbers chosen by avian fliers for forward cruising flight. It seems reasonable that this range has some correlation with optimal flight (cruising) performance as natural selection would likely favor that. Our objective is to construct a mathematical model for avian-scale (or bird-scale) forward flapping flight, capturing the flow-field physics sufficiently. Then, the objective is to examine optimal force generation and performance characteristics in avian-scale cruising flight.
- Next, the goal is to construct a set-theoretic state estimation framework that would provide set-based optimal estimates for the states of dynamical systems. To this end, the objective is to consider the ellipsoidal state estimation problem for discrete-time nonlinear systems and construct a state estimator that is structurally similar to the celebrated Kalman (or Kalman-Bucy) Filter.
- Our next goal involves utilizing the above-mentioned state estimator for synchronization of multi-agent systems subject to uncertainties. Specifically, the objective is to develop a leader-follower synchronization protocol for discrete-time high-order linear multi-agent systems, utilizing local state estimate information provided by the state estimator.
- Subsequently, we intend to develop a missile guidance strategy that takes into account the limits on available missile acceleration, does not require target acceleration information, and achieves interception with a prescribed range of

interception angles. To this end, the objective is to study the application of model predictive control to the missile guidance problem.

1.2 Background and Motivation

The background and motivation of all the above-mentioned aspects studied in this research are described in the following subsections.

1.2.1 Extremum Seeking Control

Extremum seeking control (ESC) is a data-driven adaptive control technique [2–4] for optimizing the steady-state performance of a system [5]. ESC drives the steady-state system output to its extremum by systematically tuning the reference control input parameters to the system [5,6], given there exists an extremum in the reference-to-output map. ESC is often called a model-free optimization technique as it does not rely upon the explicit knowledge of the reference-to-output map and relies instead on the measured output values. Despite being studied as early as in 1950s and 1960s, the first rigorous stability proof of an ESC scheme was provided by Krstic and Wang [5] in 2000. This development has reinvigorated the interest in ESC among control theorists and practitioners alike. Especially, due to the model-free nature of the approach, ESC has been applied to systems that are hard to model accurately or to problems for which the input-to-output mapping is completely or partially unknown. Some such applications of ESC include antilock braking systems [7–9], power reduction, induced drag minimization, or lift maximization of the wingman aircraft in formation flight [10–12], stirred-tank reactors [13, 14], electromagnetic actuators [15], real-time optimization over a network of dynamic agents [16, 17], and source localization using unmanned aerial vehicles and mobile robots [18, 19].

The focus of this research is on the classical ESC schemes in Krstic and Wang [5] and Tan et al. [6] that deal with finding the extremum of the reference-to-output equilibrium map where the relationship between the reference and the steady-state output is characterized by a nonlinear, static mapping. It is also worth mentioning that recently a different class of extremum seeking controllers has emerged in the literature that utilizes the so-called Lie bracket system (see, for example, [20–23]).

The utilization of a periodic perturbation signal (called excitation or dither signal) to extract gradient information of the unknown reference-to-output map is the cornerstone of ESC. To this end, the classical ESC schemes synthesize a measure of the gradient using high-pass and low-pass filters [5, 6, 24, 25], whereas other ESC schemes employ estimators or observers for estimating the gradient directly [3, 4, 11]. The gradient information is then utilized to drive the reference so that the steady-state output is extremized. In the presence of multiple extrema, this approach might make the output converge to a local extremum instead of the global extremum. This issue can be resolved by introducing an adaptation law for the excitation signal amplitude, with sufficiently large initial amplitude and a sufficiently slow decay rate [26]. However, the scheme in [26] exhibits slow convergence and steady-state oscillations. Note that steady-state oscillations persist inadvertently in all the classical ESC schemes, a feature that is undesirable and might not even be permissible in certain applications [9].

The issue of steady-state oscillations in the classical ESC schemes has been scrutinized over the last decade. Wang et al. [9] proposed an ESC scheme that drives amplitude of excitation signal to zero based on the error in the extremum estimation and leads to small steady-state oscillations. The current study, however, is focussed on attenuating the steady-state oscillations for the classical ESC framework proposed by Krstic and Wang [5]. This can be achieved by driving the excitation signal amplitude to zero as the extremum is reached. In this regard, instead of using the approach provided in [9], it would be more prudent to set the decay rate in the amplitude according to a measure of the derivative or gradient of the function to be optimized or extremized. This idea, one similar to which has already been used in the ESC context by Moase et al. [27], serves as a motivation for the current study.

1.2.2 Optimality in Avian-Scale Flapping Flight

Over the years, research conducted by people from different scientific disciplines have established a strong correlation between optimal flapping locomotion and the non-dimensional parameter Strouhal number [28–34]. Among these, the fluid dynamics arguments behind this correlation are primarily based on wake dynamics, and vortex development, growth, interaction and shedding [29–31]. These are grounded in the fact that Strouhal number signifies the time scale of forward movement with respect to that of the flapping motion (and vice versa), thus influencing the wake dynamics and above-mentioned vortex characteristics. However, Strouhal number is a function of the flapping kinematics [29, 35] which play a major role in force generation as it dictates the local angle of attack [34, 36]. Therefore, it is worthwhile to study the effects of flapping kinematics on flight performance and optimality of force generation by flapping.

Taylor et al. [28] showed that the Strouhal number range between 0.2 and 0.4 is utilized by several flying and swimming creatures for cruising (see figure 2 in Taylor et al. [28]). We focus on the cruising flight of birds, which is largely restricted to Strouhal numbers between 0.1 and 0.3 [28]. We are interested in studying this remarkable observation in the following two steps:

i. First, derive an analytical or semi-analytical model, manageable in terms of the mathematical and numerical complexity, for the bird-scale forward flapping flight. This model will help understand the interplay between force generation and flapping kinematics. Further, the model can be simulated to explore variations in force generation with changes in the associated parameters for Strouhal numbers between 0.1 and 0.3, and identify optimal characteristics suited to cruising.

ii. Once such optimal characteristics are identified using the model, examine possible in-flight mechanisms that birds might employ to optimize their flapping kinematics during cruising and converge to the aforementioned unique Strouhal number range (between 0.1 and 0.3), without having a priori knowledge of the same.

Models of flapping flight go as far back as 1930s, with Theodorsen [37] and Garrick [38] analytically calculating cycle-averaged aerodynamic quantities (forces and moments) for a flapping airfoil using the potential flow theory and Kutta conditions [37, 38]. A vortex-based theory (or model) for the hovering flight (of birds or insects) and forward flight (of birds) was put forward by Rayner [39–41]. DeLaurier [42] subsequently proposed a more comprehensive model that accounted for the effects of unsteady wake and dynamic stall. More recent examples for analytical and semi-analytical models of flapping can be found in [34, 36].

Although there is an abundance of computational fluid dynamics (CFD) results for insect-scale flapping (see, for example, [43] and references therein), the opposite is true for avian-scale (or bird-scale) flapping. This has been highlighted in a recent review article by Chin and Lentink [44] as well (see figure 7 in [44]). Note that this precludes the possibility of comparing results of our research with CFD results from the existing literature.

1.2.3 Set-Membership Filtering

A broad class of state estimation and filtering approaches assume the uncertainties associated with the system, namely initial condition uncertainty, process noises or input disturbances corrupting the evolution of the states, and measurement noises or output disturbances corrupting the measured outputs, to be stochastic. Most popular approach in this class is the Kalman Filter wherein the uncertainties are assumed to be Gaussian [45] with known statistical properties. The Kalman (or Kalman-Bucy) Filter was proposed in the early 1960s [46,47]. An alternative method for the state estimation of discrete-time systems, with the uncertainties considered as unknown but bounded, was introduced around this time as well [48–50]. This method generates estimates of the true state in the form of sets by making use of the uncertainty bounds, system model, and available measurements. Also, these estimated sets guarantee to contain the true state with absolute (100%) certainty, an attribute that is more suitable for several practical applications [51, 52]. Over the years, this method has become known as set-membership, set-valued, guaranteed state estimation or filtering, with studies involving ellipsoids [51–56], polytopes [57, 58], zonotopes (a special type of polytopes) [59], and constrained zonotopes (a recently introduced class of sets in [60]) [60–62]. Note that the method is suitable for parameter estimation [55, 63], and there exists a variant of this method for continuous-time systems called interval observer which aims at synthesizing upper and lower bounding trajectories for the true state [64]. However, the focus here is on the ellipsoidal state estimation problem for discrete-time systems, and the terminology set-membership filter (SMF) is adopted.

Set-membership filtering for linear systems is well-developed (see, for example, [52, 54, 56] and the references therein) and various approaches have been outlined in the existing literature, out of which the technique that utilizes online optimization principles is of particular interest to this research. This technique involves converting the state estimation process into a recursive algorithm that generally requires solution to a semi-definite program (SDP) (see, for example, [52, 54, 65]). Several extensions of this technique have been proposed in recent literature (see, for instance, [66–68]).

In contrast to the various set-membership filtering techniques available for linear systems, SMFs for discrete-time nonlinear systems are largely based on the principles of the Extended Kalman Filter (EKF). This means that the SMFs for nonlinear systems are designed based on the linearized dynamics about the state estimate trajectory, and the residual terms (remainders of linearization) are bounded by some known sets [51, 67, 69, 70]. The state dependent coefficient (SDC) parameterization [71, 72] offers alternative to this EKF-like strategy. The SDC parameterization of a nonlinear system provides a pseudo-linear description of the original nonlinear system. Further, there are stochastic filters designed based on the SDC parameterization of discrete-time nonlinear systems (see [73, 74]). However, set-membership filtering using the SDC parameterization has not been addressed in the existing open literature to the best of our knowledge. It has motivated us to explore this avenue in this research.

1.2.4 Multi-Agent Synchronization and Set-Membership Filtering

Cooperative control of multi-agent systems, which has been studied quite extensively in the last few decades, involves some degree of cooperation (and/or synchronization) among the agents, and it can be applied to distributed task assignment and consensus problems, formation flight of spacecrafts and aerial vehicles, distributed estimation problems and more [75–80]. In the existing literature, different variants of multi-agent synchronization (or consensus) have been studied. Some examples of these are as follows: (a) synchronization without a leader [81, 82], (b) leader-follower synchronization [83–86], (c) average consensus [87,88], and (d) bipartite consensus [89]. The focus of this research is on leader-follower synchronization in the presence of a leader that pins to a group of agents, all having high-order discrete-time linear (time-invariant) dynamics.

Two of the most common assumptions encountered in the existing literature on multi-agent synchronization are as follows: (a) perfect system model [82,84–86,90,91] and (b) full-state feedback available for synchronization protocol design [82, 84, 90]. Both of these are inconsistent with real-world problems where the system often involves different kinds of uncertainties (for example, input disturbances, parametric uncertainties, unmodeled dynamics). In this regard, we focus on input disturbances for this research. Further, the full-state feedback assumption is impractical for systems that can only access measured outputs (some combination or function of the states) corrupted with output disturbances. Observer-based approaches, without considering output disturbances, have been investigated in the literature [78, 83, 91]. However, it seems that a state estimation or filtering-based approach would be more suitable to address the effects of both input and output disturbances in the synchronization problem (see, for example, [79]). This has served as the motivation for the current study.

With the above discussion in mind, we are interested in applying the set-membership filtering technique developed as a part of this research for the synchronization problem. There are a few studies that have utilized set-theoretic or set-valued concepts for synchronization [92–95]. However, the application of set-membership filtering to the multi-agent synchronization problem has been limited [87, 88], despite the practical significance of this class of estimators/filters. A recent study reported in [96] has considered the leader-follower synchronization using set-membership estimation techniques, wherein the synchronization objective was to construct ellipsoids that are centered at the leader's state trajectory and contain states of the agents.

This, however, is different from the concept of conventional leader-follower synchronization where the objective is to make the states of the agents converge to the leader's state trajectory. To the best of our knowledge, set-membership estimation techniques have not been employed for the conventional leader-follower synchronization problem in the existing literature.

1.2.5 Model Predictive Control and Missile Guidance

Model Predictive Control (MPC) is an optimal control strategy wherein the control input is synthesized based on the solution to a finite horizon optimal control problem. By using MPC, a user has the capability to directly incorporate constraints on the inputs, outputs, and states, while minimizing a cost function. MPC has seen successful applications in several engineering disciplines and is widely used in the process industry (see, for example, [97, 98] and references therein). However, the application of MPC to the problem of missile guidance has been somewhat limited, possibly due to the computational cost associated with the control synthesis. The computational cost is typically high due to the fast sampling rates that are required for this problem [99], and due to this, online implementation of an MPC-based guidance scheme might not have been feasible in the 20th century. However, with the emergence of inexpensive computation and the advent of efficient solvers in the last two decades, the application of MPC to this problem seems feasible. This is evidenced by the increased number of publications in the literature in recent years that have proposed missile guidance schemes based on generalized predictive control (GPC) or MPC (see, for example, [99–104]).

Proportional navigation (PN) is one of the most widely studied missile guidance laws. Essentially, the PN guidance law is synthesized based on the rate of rotation of the line-of-sight (LOS) and the law is easily implementable for practical applications. Several variants of PN guidance laws have appeared in the existing literature (see, for example, [105] and references therein). Although easily implementable, PN laws [106–108] do not explicitly account for pointwise-in-time hard constraints on the lateral accelerations (latax) of the missile. Thus, this might lead to engagement scenarios where the commanded lateral acceleration violates the constraint. This serves as a motivation for implementing MPC for designing guidance laws for missiles. The equivalence between PN guidance laws and solutions to unconstrained minimum energy optimal control problems has been shown in [109], [110]. In that respect, PN can be thought of as a special case of MPC, that minimizes the time integral of the square of the latax, without any explicit hard constraints on the magnitude of the latax pointwise-in-time.

Predictive control strategies based on generalized predictive control (GPC) have been implemented recently for missile guidance [102, 103]. GPC is computationally cheaper than MPC and an explicit solution can be derived for a given nonlinear system [111]. He and Lin [102] proposed a composite guidance law, based on GPC and a target maneuver estimator using a continuous second-order sliding mode technique, for planar engagement scenarios. Wang and He [103] derived robust missile guidance laws based on GPC and integral sliding mode for intercepting maneuvering targets with desired terminal LOS angle constraint. While it is possible to account for control input constraints in the GPC framework [111, 112], this was not incorporated in the guidance law designs in [102, 103].

There have been some applications of MPC to the missile guidance problem. Li et al. [101] implemented a nonlinear model predictive control (NMPC)-based guidance scheme for a planar case and treated the target acceleration components as unknown bounded disturbances. This approach, if feasible, would result in a guidance scheme robust with respect to the target maneuvers. A quadratic program (QP) was formulated for the NMPC and a neural network-based approach for online implementation was shown [101]. Bachtiar et al. [99] proposed integrated control and guidance of missiles based on NMPC, and also discussed the issues of implementation cost and computational capacity required to implement the NMPC controller. A multiobjective offline tuning framework was introduced that balances the trade-off between the performance of the scheme and the computational cost associated with the implementation. More recently, Kang et al. [104] proposed an MPC-based cooperative guidance scheme to perform salvo attacks against stationary targets, which guaranteed that multiple missiles hit the target simultaneously. The engagement kinematics was formulated in a state dependent linear form and a time-to-go estimate was also utilized. Some of the recent advancements in optimal control theory-based missile guidance can be found in [113–117].

In the missile guidance literature, the concept of terminal impact or intercept angle has been studied extensively (see, for example, [118–123]). By striking a target at a desired impact angle, or within a range of impact angles, the target can be attacked from the (set of) directions that it has less protection and is therefore more vulnerable. Oza and Padhi [118] presented an impact angle constrained guidance law for three dimensional engagement geometries based on model predictive static programming which utilizes closed-form solutions of a constrained *static* optimization problem. In contrast, MPC techniques typically require solutions to constrained *dynamic* optimization problems [97, 98]. Ratnoo and Ghose proposed an impact angle constrained PN guidance law for stationary targets in [119] and that framework was extended for non-stationary, non-maneuvering targets in [120]. Shaferman and Shima derived optimal guidance laws for achieving desired terminal impact angles for a single missile in [121] and for a group of cooperating missiles in [122].

The missile guidance literature predominantly assumes that the missile and the target can be modeled as point objects. However, this assumption does not necessarily hold in the terminal guidance phase, during which the target can be modeled as a circle whose radius is equal to the blast radius of the warhead carried by the missile. During this phase, a collision cone-based approach [124–127], which has been employed to design guidance laws to achieve or avoid collision when one or more of the objects have finite dimension, can be utilized. In this research, we utilize a collision cone-based approach to satisfy the impact angle requirement. The collision cone approach is employed during the latter phase of the engagement to steer the missile's velocity vector appropriately so that at the time the missile arrives at a pre-defined distance (which is the blast radius) to the target, the impact angle satisfies a pre-defined constraint.

1.3 Contributions

Contributions of the research work are as outlined in the following subsections.

1.3.1 Extremum Seeking Control

Two perturbation-based ESC schemes for general single input single output nonlinear systems, based on the classical ESC setup in [5], have been proposed. The proposed schemes involve novel adaptation laws for the respective excitation signal amplitudes. These laws are designed such that the amplitudes are asymptotically driven to zero once the system reaches the extremum. The rates of decay are governed by the gradient measures of the unknown reference-to-output equilibrium map, which is the function extremized in this case. Our approach leads to attenuated steady-state oscillations. Also, it is shown that the proposed schemes are able to converge to the extremum in a practical asymptotic manner. In other words, the proposed ESC schemes are able to bring the system arbitrarily close to its true optimal operating point.

1.3.2 Optimality in Avian-Scale Flapping Flight

A semi-analytical model capable of adequately capturing the forward flapping flight of birds is developed. The model involves strip theory-based quasi-steady aerodynamics for the local forces and utilizes concepts from flight mechanics to calculate forces for finite flapping wings. Using results generated through this model, it is shown that, in the range of Strouhal numbers between 0.1 and 0.3 (the unique range used by birds while cruising), cycle-averaged net thrust, lift, and lift-to-drag ratio are optimal (for cruising flight) for a given flow pattern over the upper surfaces of the wings. Furthermore, a hypothesis is presented for the in-flight mechanism employed by birds to converge to the aforementioned Strouhal number range, and it is postulated that birds use some kind of online optimization to achieve this. This hypothesis is verified in a simulation study where the above-mentioned model is used to compute the cycle-averaged lift-to-drag ratio (regarded as the optimization objective) and one of the proposed ESC schemes is employed as the optimization framework.

1.3.3 Set-Membership Filtering

A novel set-membership filtering method for nonlinear systems is proposed using SDC parameterization as a key tool. Application of SDC parameterization renders the original nonlinear systems into a pseudo linear form with state dependent system matrices, and this form is utilized to construct a nonlinear set-membership filter which has a two-step correction-prediction structure similar to a Kalman Filter. The filter is termed SDC-SMF. The state estimation problem for SDC-SMF is ultimately converted into a recursive algorithm that requires solutions to two SDPs. It is shown (in simulations) that the SDC-SMF has better overall estimation performance compared to another existing set-membership filter, despite the latter being computationally costlier than the former. Also, theoretical conditions for stability and convergence of SDC-SMF are derived.

1.3.4 Multi-Agent Synchronization Using Set-Membership Filtering

A linear version of the proposed SDC-SMF is considered in a multi-agent scenario wherein the objective is to synchronize a group of follower agents to a leader, all of which are governed by linear high-order discrete-time dynamics. The follower agents are subject to system uncertainties and only have access to measurements corrupted with disturbances. Each of the agents are equipped with a set-membership filter that estimates its state. Local state estimate sharing is allowed among the neighbors. In this setting, a distributed synchronization protocol, which utilizes an H_2 type Riccati-based approach [90] for the local controller of each agent, is proposed. The protocol, under appropriate conditions, renders the global error system input-to-state stable (ISS) with respect to the input disturbances and estimation errors.

1.3.5 Missile Guidance Using Model Predictive Control

A novel missile guidance algorithm is developed using NMPC and collision cone theory. The latax synthesis is carried out through an NMPC setup that takes into account explicit constraints on the latax magnitude and rate of change (increments or decrements in the latax). Additionally, the impact angle requirements are converted into appropriate constraints to be used in the NMPC setup. Two different components, to be used at different stages of an engagement, are derived, with each requiring to solve a strictly convex quadratic program (QP). Moreover, the proposed method is suitable for practical engagement scenarios as no target acceleration information is required.

List of Published Works

Refereed Journal Publications

1. D. Bhattacharjee and K. Subbarao, “Extremum Seeking Control with Attenuated Steady-State Oscillations,” *Automatica*, Vol. 125, p. 109432, 2021. DOI: 10.1016/j.automatica.2020.109432 (reference [128])
2. D. Bhattacharjee and K. Subbarao, “A Flight Mechanics-Based Justification of the Unique Range of Strouhal Numbers for Avian Cruising Flight,” *Proc IMechE Part G: Journal of Aerospace Engineering*, p. 0954410020976597, 2020. DOI: 10.1177/0954410020976597 (reference [129])
3. D. Bhattacharjee and K. Subbarao, “Do Birds Employ Online Optimization for Cruising Flight?,” Submitted to *Scientific Reports-Nature*, *Under Review.*
4. D. Bhattacharjee and K. Subbarao, “Set-Membership Filter for Discrete-Time Nonlinear Systems Using State Dependent Coefficient Parameterization,” *IEEE Transactions on Automatic Control*, Early access, May 2021. DOI: 10.1109/TAC.2021.3082504 (reference [130])
5. D. Bhattacharjee and K. Subbarao, “Set-Membership Filtering-Based Leader-Follower Synchronization of Discrete-Time Linear Multi-Agent Systems,” *ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 143, Issue 6, p. 064502, 2021. DOI: 10.1115/1.4049553 (reference [131])
6. D. Bhattacharjee, A. Chakravarthy, and K. Subbarao, “Nonlinear Model Predictive Control and Collision Cone-Based Missile Guidance Algorithm,” *AIAA Journal of Guidance, Control, and Dynamics*, pp. 1-17, 2021. DOI: 10.2514/1.G005879 (reference [132])

Refereed Conference Publications

1. D. Bhattacharjee, K. Subbarao, and K. Bhaganagar, "Extremum Seeking and Adaptive Sampling Approaches for Plume Source Estimation using Unmanned Aerial Vehicles," In *AIAA SciTech 2019 Forum*, Jan. 2019, AIAA 2019-1565. DOI: 10.2514/6.2019-1565 (reference [19])
2. D. Bhattacharjee and K. Subbarao, "Closed-Form Expressions for Cycle-Averaged Aerodynamic Quantities at an Airfoil Section of an Avian Flapping Wing," In *AIAA SciTech 2019 Forum*, Jan. 2019, AIAA 2019-0565. DOI: 10.2514/6.2019-0565 (reference [36])
3. D. Bhattacharjee, K. Subbarao, and A. Chakravarthy, "Set-Membership Filtering-Based Pure Proportional Navigation," In *AIAA SciTech 2021 Forum*, Jan. 2021, AIAA 2021-1567. DOI: 10.2514/6.2021-1567 (reference [133])
4. D. Bhattacharjee, K. Subbarao, and K. Bhaganagar, "Reachable Set Estimation for Discrete-Time Nonlinear Systems Using Ellipsoidal Set-Membership Frameworks," In *AIAA SciTech 2021 Forum*, Jan. 2021, AIAA 2021-1459. DOI: 10.2514/6.2021-1459 (reference [134])
5. D. Bhattacharjee and K. Subbarao, "Nonlinear Set-Membership Filtering-Based Orbit Estimation," In *31st AAS/AIAA Space Flight Mechanics Meeting*, Feb. 2021, AAS 21-314.
6. D. Bhattacharjee and K. Subbarao, "Nonlinear Set-Membership Filtering-Based State Estimation of Reentry Vehicles," In *31st AAS/AIAA Space Flight Mechanics Meeting*, Feb. 2021, AAS 21-306.
7. D. Bhattacharjee, A. Chakravarthy, and K. Subbarao, "Nonlinear Model Predictive Control based Missile Guidance for Target Interception," In *AIAA SciTech 2020 Forum*, Jan. 2020, AIAA 2020-0865. DOI: 10.2514/6.2020-0865 (reference [135])

1.4 Dissertation Outline

The interconnections between chapters in the main body of the dissertation are shown in Fig. 1.1. The proposed extremum seeking control schemes are detailed in Chapter 2. Chapter 3 describes the mathematical modeling of flapping, and Chapter 4 details the hypothesis regarding possible in-flight mechanism used by birds for optimal cruising performance. The set-membership filter for discrete-time nonlinear systems is discussed in Chapter 5. Subsequently, Chapter 6 provides details on the synchronization protocol design. Then, the application of NMPC for missile guidance is discussed in Chapter 7. Finally, the concluding remarks and possible future directions of the research are provided in Chapter 8.

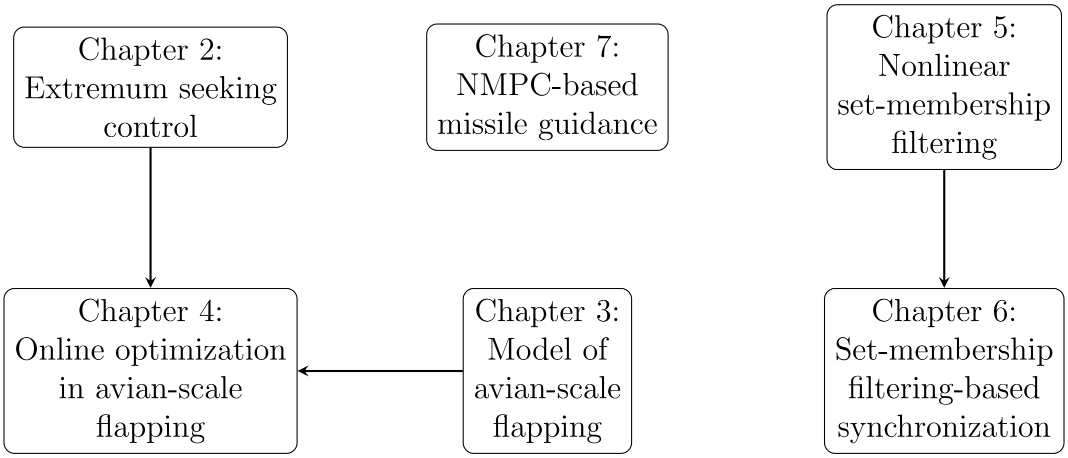

Chapter 2: Extremum seeking control
Chapter 7: NMPC-based missile guidance
Chapter 5: Nonlinear set-membership filtering
Chapter 4: Online optimization in avian-scale flapping
Chapter 3: Model of avian-scale flapping
Chapter 6: Set-membership filtering-based synchronization

Figure 1.1: Interconnections between the chapters

Chapter 2

Extremum Seeking Control With Attenuated Steady-State Oscillations*

The extremum seeking control designs are described in this chapter. The assumptions and problem formulation are described in Section 2.1. Section 2.2 elaborates the main results and illustrative simulation examples are given in Section 2.3. Finally, Section 2.4 summarizes the findings of this chapter.

2.1 Preliminaries and Problem Formulation

To be consistent with the existing results in the literature, we adopt the same notations and a similar problem formulation as given in Krstic and Wang [5] and Tan et al. [6, 26]. For the sake of completeness, we describe the problem formulation and the list of assumptions in this section. Consider a general single input and single output (SISO) nonlinear dynamical model given by

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u), \quad y = h(\boldsymbol{x}), \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}$ is the control input, $y \in \mathbb{R}$ is the measured output, and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}$ are smooth. Suppose there exists a family of smooth feedback control laws of the form

$$u = \alpha(\boldsymbol{x}, \theta), \tag{2.2}$$

*The materials of this chapter have been published in D. Bhattacharjee and K. Subbarao, "Extremum Seeking Control with Attenuated Steady-State Oscillations," *Automatica*, Vol. 125, p. 109432, 2021, DOI: 10.1016/j.automatica.2020.109432 (reference [128]), and these are included here under the author rights granted by Elsevier (Online, accessed in June, 2021: `https://www.elsevier.com/about/policies/copyright`).

parameterized by the scalar parameter θ. Then, the closed-loop system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \alpha(\boldsymbol{x}, \theta)) \tag{2.3}$$

has equilibria parameterized by θ. In this chapter, we are interested in investigating the extremum seeking schemes shown in Figs. 2.1 and 2.2. Note that one can retrieve the classical ESC scheme given in Krstic and Wang [5] by (a) removing the input to the shaded blocks, (b) replacing the shaded blocks in the proposed schemes with a constant amplitude a for the excitation signal, and (c) multiplying the demodulation signal with the same amplitude a. Moreover, $\hat{\theta}$ in these schemes can be considered to be the nominal part of the reference or the current estimate of the extremum $\theta^\star$ and $a \sin \omega t$ to be the excitation signal with the amplitude $a = a(t)$ as a function of time (Krstic and Wang [5], Haring and Johansen [3], [4]). We make same assumptions for the closed-loop system as Krstic and Wang [5]. These are given as follows.

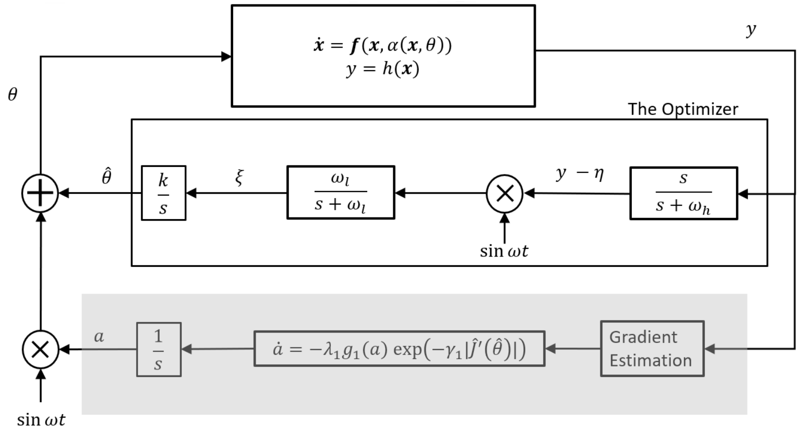


Figure 2.1: Proposed ESC scheme-1

Assumption 2.1.1. *There exists a smooth function* $\boldsymbol{l} : \mathbb{R} \to \mathbb{R}^n$ *such that* $\boldsymbol{f}(\boldsymbol{x}, \alpha(\boldsymbol{x}, \theta)) = 0$*, if and only if* $\boldsymbol{x} = \boldsymbol{l}(\theta)$*.*

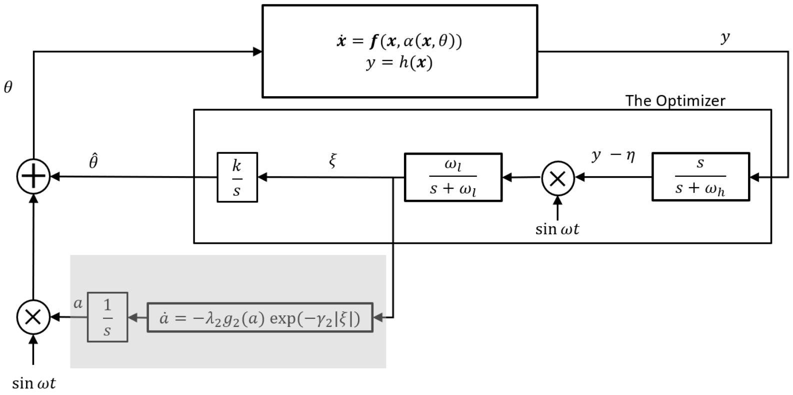


Figure 2.2: Proposed ESC scheme-2

Assumption 2.1.2. *For each $\theta \in \mathbb{R}$, the equilibrium $\boldsymbol{x} = \boldsymbol{l}(\theta)$ of system* (2.3) *is locally exponentially stable, uniformly in θ.*

This assumption means that we have a control law that would stabilize the system locally, irrespective of the modeling knowledge of either $\boldsymbol{f}(\boldsymbol{x}, u)$ or $\boldsymbol{l}(\theta)$. Without loss of generality, we consider the problem of maximizing the steady-state output by finding the maximum in the output equilibrium map $y = h(\boldsymbol{l}(\theta))$. The case for the minimization problem can be treated similarly by replacing y with $-y$. Since we are interested in finding the maximum in the output equilibrium map $y = h(\boldsymbol{l}(\theta))$, let us denote $J(\theta) = h(\boldsymbol{l}(\theta)) = (h \circ \boldsymbol{l})(\theta)$ as the objective function for the extremum seeking problem. Similarly, for the minimization problem, $-h(\boldsymbol{l}(\theta))$ can be treated as a cost function that needs to be minimized.

Assumption 2.1.3. *There exists $\theta^\star \in \mathbb{R}$ such that*

$$\begin{aligned} J'(\theta^\star) &= 0, \\ J''(\theta^\star) &< 0. \end{aligned} \tag{2.4}$$

This last assumption implies that the objective function $J(\theta)$ has a maximum at $\theta = \theta^\star$. As shown in Fig. 2.1, the gradient of the objective function has to be estimated. For that, we have adapted the Kalman Filter-based gradient estimation scheme from Chichka et al. [11] and we denote the gradient estimate at $\hat{\theta}$ as $\hat{J}'(\hat{\theta})$. The Kalman Filter 'truth' model for this case is as following.

$$\dot{\boldsymbol{\psi}} = \begin{bmatrix} 0 & \omega & 0 \\ -\omega & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{\psi} + \boldsymbol{w},\ \tilde{y} = \begin{bmatrix} a & 0 & 1 \end{bmatrix} \boldsymbol{\psi} + v, \tag{2.5}$$

where $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3) = (J'(\hat{\theta}) \sin \omega t, J'(\hat{\theta}) \cos \omega t, J(\hat{\theta}))$, ω is the excitation signal frequency, a is the excitation signal amplitude, $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q})$ and $v \sim \mathcal{N}(0, r)$ are zero-mean Gaussian white-noise terms with covariances $\boldsymbol{Q}$ and r, respectively, and $\tilde{y}$ is a measurement of the objective function at $\theta = \hat{\theta} + a \sin \omega t$. Furthermore, $\boldsymbol{w}$ and v are uncorrelated. The estimate of $\boldsymbol{\psi}$ is denoted by $\hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3)$. Therefore, the estimated gradient magnitude is given by $|\hat{J}'(\hat{\theta})| = \sqrt{\hat{\psi}_1^2 + \hat{\psi}_2^2}$. Note that we do not specify a sign to the gradient estimate and only utilize the estimated gradient magnitude (cf. (2.6)).

Assumption 2.1.4. *There exists a positive constant ϵ_0 such that the estimates satisfy $|\hat{J}'(\hat{\theta})| \leq \epsilon_0$ for all $t \geq t_0 \geq 0$. Moreover, there exists a small positive constant ϵ_J and a time interval $\Delta T > 0$ such that $|\ |J'(\hat{\theta})| - |\hat{J}'(\hat{\theta})|\ | \leq \epsilon_J$ for all $t \geq t_0 + \Delta T$.*

Assumption 2.1.4 implies that the Kalman Filter is performing adequately after some non-zero time interval ΔT from time t_0, when the scheme was initialized.

2.2 Main Results

In this section, we propose two adaptation laws for the amplitude of the excitation signal so that the excitation signal converges to zero as the extremum is reached. These are given as following.

- Adaptation law (scheme-1):

$$\dot{a} = -\lambda_1\ g_1(a)\ \exp(-\gamma_1 |\hat{J}'(\hat{\theta})|),\ \ a(t_0) = a_0 > 0, \tag{2.6}$$

- Adaptation law (scheme-2):

$$\dot{a} = -\lambda_2 \; g_2(a) \; \exp(-\gamma_2|\xi|), \;\; a(t_0) = a_0 > 0, \tag{2.7}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are design parameters, γ_1 and γ_2 are $O(1)$ positive scaling parameters, ξ is as shown in Fig. 2.2, and $g_1(a), g_2(a)$ are locally Lipschitz functions that are zero at zero and positive otherwise. The parameters γ_1, γ_2, and a_0 are to be selected based on the problem at hand and are left to the choice of the designer.

Remark 2.2.1. *The choice of the adaptation laws in* (2.6) *and* (2.7) *is motivated by the adaptation law proposed in Tan et al. [26]. We remark that the proposed laws are equivalent to the one in Tan et al. [26] when the gradient measures are sufficiently small. However, since the standing assumption in ESC is that the extremum is unknown, one cannot guarantee that the gradient measures are small when an ESC scheme is initialized. To this end, the scaling parameters γ_1 and γ_2 for scheme-1 and scheme-2, respectively, can be chosen sufficiently large so that the rates of decay in the amplitudes are initially governed by the exponents in* (2.6) *and* (2.7)*. This means that the decay in the amplitudes is seized during the initialization period of the proposed schemes and is similar, in spirit, to the requirement of a sufficiently large a_0 for the scheme in Tan et al. [26]. In fact, for the proposed schemes, a_0 has to be sufficiently small (cf. Theorems 2.2.5 and 2.2.6).*

Remark 2.2.2. *Utilzing the discussion in Remark 2.2.1, we choose γ_1 and γ_2 sufficiently large to arrest the rates of decay in the amplitudes during the initialization period of the proposed schemes and allow the optimizer to make proper corrections (see Section 2.3). Then, as the system approaches the extremum and the gradient measures become sufficiently small, the rates of decay are approximately governed by the adaptation law in Tan et al. [26], i.e., $\dot{a} \approx -\lambda_i \; g_i(a)$ $(i = 1, 2)$. Thus, for the* *proposed schemes, we would like the decay in the excitation signal amplitudes to occur in two phases, with γ_1 and γ_2 properly chosen (see Section 2.3).*

Remark 2.2.3. *A switching control-based strategy, with an adaptation law for the decay in excitation signal amplitude similar to that in Tan et al. [26], was given in Moura and Chang [136]. However, note that the switching requires a Lyapunov function that utilizes the knowledge of an accurate enough 'nominal' extremum and the corresponding numerical values of the objective function derivatives as well as a switching threshold. No such prior knowledge is required for our proposed schemes.*

Next, we elaborate the stability analysis of the proposed ESC schemes shown in Figs. 1 and 2. Letting $\tilde{\theta} = \hat{\theta} - \theta^\star$, $\tilde{\eta} = \eta - J(\theta^\star)$, and substituting $y = h(\boldsymbol{x})$, the closed-loop systems can be expressed as following.

Scheme-1:

$$\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{f}\left(\boldsymbol{x}, \alpha(\boldsymbol{x}, \tilde{\theta} + \theta^\star + a\sin\omega t)\right),\\
\dot{\tilde{\theta}} &= k\xi,\\
\dot{\xi} &= -\omega_l\xi + \omega_l(h(\boldsymbol{x}) - \tilde{\eta} - J(\theta^\star))\sin\omega t,\\
\dot{\tilde{\eta}} &= -\omega_h\tilde{\eta} + \omega_h(h(\boldsymbol{x}) - J(\theta^\star)),\\
\dot{a} &= -\lambda_1\ g_1(a)\ \exp(-\gamma_1|\hat{J}'(\tilde{\theta} + \theta^\star)|).
\end{aligned} \tag{2.8}$$

Scheme-2:

$$\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{f}\left(\boldsymbol{x}, \alpha(\boldsymbol{x}, \tilde{\theta} + \theta^\star + a\sin\omega t)\right),\\
\dot{\tilde{\theta}} &= k\xi,\\
\dot{\xi} &= -\omega_l\xi + \omega_l(h(\boldsymbol{x}) - \tilde{\eta} - J(\theta^\star))\sin\omega t,\\
\dot{\tilde{\eta}} &= -\omega_h\tilde{\eta} + \omega_h(h(\boldsymbol{x}) - J(\theta^\star)),\\
\dot{a} &= -\lambda_2\ g_2(a)\ \exp(-\gamma_2|\xi|).
\end{aligned} \tag{2.9}$$

We introduce the following representation for the parameters in (2.8) and (2.9).

$$\begin{aligned}
\omega_h &= \omega\omega_H = \omega\delta\omega'_H = O(\omega\delta), \\
\omega_l &= \omega\omega_L = \omega\delta\omega'_L = O(\omega\delta), \\
k &= \omega K = \omega\delta K' = O(\omega\delta), \\
\lambda_1 &= \omega\lambda_{1_1} = \omega\delta\epsilon\lambda'_1 = O(\omega\delta\epsilon), \\
\lambda_2 &= \omega\lambda_{2_1} = \omega\delta\epsilon\lambda'_2 = O(\omega\delta\epsilon),
\end{aligned}$$

where ω, δ, and ϵ are small positive constants and ω'_H, ω'_L, K', λ'_1, and λ'_2 are $O(1)$ positive constants. From the above representation of the parameters involved, we can conclude that the closed-loop systems of the proposed schemes should exhibit four time scales. These requirements on the time scale properties of the proposed schemes are similar to the ones introduced in Krstic and Wang [5] and Haring and Johansen [3], [4]. These time scales are given by:

- fast - the system with the controller
- medium fast - the periodic perturbations
- medium slow - the filters in the proposed schemes
- slow - the adaptation in the excitation signal amplitude

The system is required to be fast compared to the rest of the components of the schemes so that the difference between true output of the system ($y = h(\boldsymbol{x})$) and the steady-state output corresponding to the objective function ($J(\theta) = h(\boldsymbol{l}(\theta))$) remains small. The filters are required to be slower compared to the periodic perturbations as that would allow the filters to accurately estimate the nominal part of the reference ($\hat{\theta}$). Also, adaptation in the excitation signal amplitude is required to be sufficiently slow so that the optimality of the current estimate $\hat{\theta}$ is checked and appropriate corrections are made by the optimizer (see Remark 2.2.4).

Remark 2.2.4. *It follows from the analysis in Krstic and Wang [5] that ω and δ should be sufficiently small for the proposed schemes. Moreover, ϵ has to be small, as in Tan et al. [26]. An analysis similar to the one recently given in Atta and Guay [137] can be carried out to establish the existence of equilibrium manifolds for the systems* (2.8) *and* (2.9)*. On these manifolds, we have $a = 0$ and $\tilde{\theta} = \theta_c$ where θ_c is a constant, not necessarily zero. As the system approaches the extremum and the exponents in* (2.6) *and* (2.7) *are approximately equal to one, the rates of decay in the amplitudes are governed by $-\lambda_i \ g_i(a)$ $(i = 1, 2)$. By making ϵ small and reducing the rates of decay of the amplitudes while the system approaches the extremum, we allow the optimizer to make corrections so that we have $\tilde{\theta} \to \theta_c$ with θ_c sufficiently small. In this way, the proposed ESC schemes are able to achieve practical asymptotic convergence to the extremum (see Theorems 2.2.5 and 2.2.6).*

Next, we summarize the main results in the following Theorems.

Theorem 2.2.5. *Consider the closed-loop system* (2.8) *under the Assumptions 2.1.1, 2.1.2, 2.1.3, and 2.1.4. Then, there exist class $\mathcal{KL}$ function β_{a_1} with $\beta_{a_1}(s, 0) = s$ and positive constants $\bar{a}_1,\ \Delta_1,\ k_1,\ k_2,\ \alpha_1,\ \alpha_2$ such that, for a choice of ω'_H, ω'_L, K', λ'_1 and for each $\rho > 0$, there exist positive constants $\bar{\epsilon}_1,\ \bar{\delta}_1,\ \bar{\omega}_1,\ \bar{\gamma}_1$ such that, for all $\delta \in (0, \bar{\delta}_1), \omega \in (0, \bar{\omega}_1), \epsilon \in (0, \bar{\epsilon}_1)$, $\gamma_1 \in (0, \bar{\gamma}_1)$ and for all initial conditions satisfying $a_0 \in (0, \bar{a}_1)$ and $|\,(\tilde{\boldsymbol{x}}(t_0), \tilde{\boldsymbol{z}}(t_0))\,| \leq \Delta_1$, the solutions of the system* (2.8) *satisfy for all $t \geq t_0 \geq 0$*

$$|\tilde{\boldsymbol{x}}(t)| \quad \leq \quad k_1 \exp(-\alpha_1(t - t_0))\ |\tilde{\boldsymbol{x}}(t_0)| + \rho, \tag{2.10}$$

$$|\tilde{\boldsymbol{z}}(t)| \quad \leq \quad k_2 \exp(-\alpha_2 \omega\delta(t - t_0))\ |\tilde{\boldsymbol{z}}(t_0)| + \rho, \tag{2.11}$$

$$|a(t)| \quad \leq \quad \beta_{a_1}(a_0,\ \omega\delta\epsilon(t - t_0)), \tag{2.12}$$

with $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \boldsymbol{l}(\tilde{\theta}(t) + \theta^\star + a(t)\sin\omega t)$*,* $\tilde{\boldsymbol{z}}(t) = \boldsymbol{z}(t) - \boldsymbol{z}_1^p(t, a(t))$ *where* $\boldsymbol{z}(t) = \Big(\tilde{\theta}(t), \xi(t), \tilde{\eta}(t)\Big)$ *and* $\boldsymbol{z}_1^p(t, a(t)) = \Big(\tilde{\theta}^p(t), \xi^p(t), \tilde{\eta}^p(t)\Big)$ *is a unique* $\left(\frac{2\pi}{\omega}\right)$*-periodic solution, characterized by* $a(t)$*.*

Proof. A sketch of the proof is provided in Appendix B. □

Theorem 2.2.6. *Consider the closed-loop system* (2.9) *under the Assumptions 2.1.1, 2.1.2, and 2.1.3. Then, there exist class* $\mathcal{KL}$ *function* β_{a_2} *with* $\beta_{a_2}(s, 0) = s$ *and positive constants* $\bar{a}_2$*,* Δ_2*,* k_3*,* k_4*,* α_3*,* α_4 *such that, for a choice of* ω'_H*,* ω'_L*,* K'*,* λ'_2 *and for each* $\rho > 0$*, there exist positive constants* $\bar{\epsilon}_2$*,* $\bar{\delta}_2$*,* $\bar{\omega}_2$*,* $\bar{\gamma}_2$ *such that, for all* $\delta \in (0, \bar{\delta}_2), \omega \in (0, \bar{\omega}_2), \epsilon \in (0, \bar{\epsilon}_2)$*,* $\gamma_2 \in (0, \bar{\gamma}_2)$ *and for all initial conditions satisfying* $a_0 \in (0, \bar{a}_2)$ *and* $|\,(\tilde{\boldsymbol{x}}(t_0), \tilde{\boldsymbol{z}}(t_0))\,| \leq \Delta_2$*, the solutions of the system* (2.9) *satisfy for all* $t \geq t_0 \geq 0$

$$|\tilde{\boldsymbol{x}}(t)| \quad \leq \quad k_3 \exp(-\alpha_3(t - t_0))\ |\tilde{\boldsymbol{x}}(t_0)| + \rho, \tag{2.13}$$

$$|\tilde{\boldsymbol{z}}(t)| \quad \leq \quad k_4 \exp(-\alpha_4 \omega\delta(t - t_0))\ |\tilde{\boldsymbol{z}}(t_0)| + \rho, \tag{2.14}$$

$$|a(t)| \quad \leq \quad \beta_{a_2}(a_0,\ \omega\delta\epsilon(t - t_0)), \tag{2.15}$$

with $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \boldsymbol{l}(\tilde{\theta}(t) + \theta^\star + a(t)\sin\omega t)$*,* $\tilde{\boldsymbol{z}}(t) = \boldsymbol{z}(t) - \boldsymbol{z}_2^p(t, a(t))$ *where* $\boldsymbol{z}(t) = \Big(\tilde{\theta}(t), \xi(t), \tilde{\eta}(t)\Big)$ *and* $\boldsymbol{z}_2^p(t, a(t)) = \Big(\tilde{\theta}^p(t), \xi^p(t), \tilde{\eta}^p(t)\Big)$ *is a unique* $\left(\frac{2\pi}{\omega}\right)$*-periodic solution, characterized by* $a(t)$*.*

Proof. The proof follows from that of Theorem 2.2.5 by making straightforward modifications and has been omitted. □

Remark 2.2.7. *Theorems 2.2.5 and 2.2.6 state that it is possible to achieve practical asymptotic convergence to the extremum by proper tuning of the parameters in the proposed ESC schemes. Adopting the terminology introduced in Tan et al. [6] and utilized in Tan et al. [26], the main results stated in Theorems 2.2.5 and 2.2.6 can be interpreted as follows: the solutions first converge to a small neighborhood of the set*

$\{(\boldsymbol{x}, \theta) : \boldsymbol{x} - \boldsymbol{l}(\theta) = 0\}$, *then with the speed proportional to* $\omega\delta$ *to a neighborhood of the sets* $\{(\tilde{\theta}, \xi, \tilde{\eta}) : \boldsymbol{z} - \boldsymbol{z}_i^p = 0\} (i = 1, 2)$, *and finally, with the speed proportional to* $\omega\delta\epsilon$ *to a neighborhood of the point* $(\tilde{\theta}, \xi, \tilde{\eta}, a) = (0, 0, 0, 0)$. *Thus,* (θ, ξ, η, a) *converge to a neighborhood of* $(\theta^\star, 0, J(\theta^\star), 0)$ *and* y *converges to a neighborhood of the extremum* $J(\theta^\star)$.

Remark 2.2.8. *It is not possible to analytically calculate* $\bar{\epsilon}_i$, $\bar{a}_i$, $\bar{\delta}_i$, $\bar{\omega}_i$, Δ_i, $\bar{\gamma}_i$ $(i = 1, 2)$. *However, it is possible to get conservative estimates of these quantities by carrying out experiments. These remarks are similar to the ones in Tan et al. [26].*

Remark 2.2.9. *The* $\hat{\theta}$ *dynamics for the ESC scheme in Tan et al. [26] is given by* $\dot{\hat{\theta}} = \omega \ \delta \ h(\boldsymbol{x}) \sin \omega t$. *Clearly, as* $y = h(\boldsymbol{x})$ *converges to a neighborhood of* $J(\theta^\star)$, $\hat{\theta}$ *would (sinusoidally) oscillate with an amplitude approximately equal to* $\delta J(\theta^\star)$. *Thus, a sufficiently large* $J(\theta^\star)$ *would make* $\hat{\theta}$ *oscillate with a large amplitude. Although this amplitude could be reduced by selecting a small enough* δ, *the resulting convergence speed to the extremum would reduce and additional tuning would be required. The oscillation in* $\hat{\theta}$ *makes* $\theta = \hat{\theta} + a \sin \omega t$ *oscillatory, even when* $a \to 0$. *Also, without loss of generality, we can deduce that the control input* $u = \alpha(\boldsymbol{x}, \theta)$, *the equilibria* $\boldsymbol{x} = \boldsymbol{l}(\theta)$, *and* y *would be oscillatory. On the other hand, consider the proposed schemes with* y *in a neighborhood of* $J(\theta^\star)$ *and* ξ *in a neighborhood of 0 (see Remark 2.2.7). For the proposed schemes, we have* $\dot{\hat{\theta}} = k\xi$ *with* k *sufficiently small. Therefore, the oscillations in* $\hat{\theta}$, θ, $\boldsymbol{x} = \boldsymbol{l}(\theta)$, $y = h(\boldsymbol{x})$, *and* $u = \alpha(\boldsymbol{x}, \theta)$ *for the proposed schemes would be attenuated as the extremum is reached. Essentially, these differences in the steady-state oscillations are due to the difference in the loop structures between the proposed schemes and the scheme in Tan et al. [26] (cf. our Figs. 2.1, 2.2 and Fig. 2 in Tan et al. [26]). Due to the attenuated steady-state oscillations, the proposed schemes would be more favorable, compared to the one in Tan et al. [26], for applications where steady-state oscillations are not desirable and/or not permitted. In*

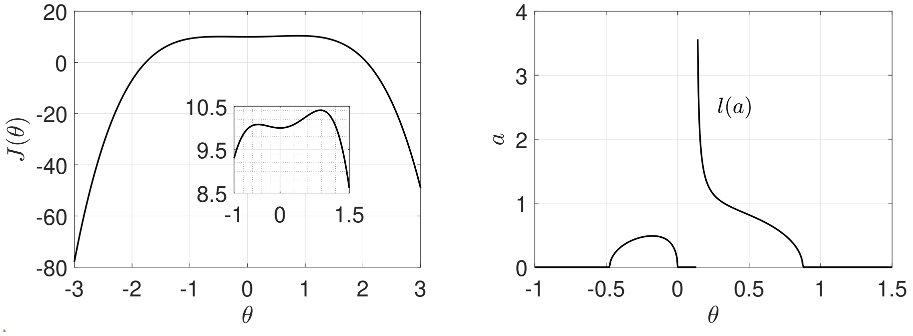


Figure 2.3: The objective function in (2.17) and the corresponding bifurcation diagram which is required for the scheme in Tan et al. [26].

addition to that, proposed ESC schemes offer the following advantages over the scheme in Tan et al. [26]: (i) λ_1 and λ_2 allow the user more control over the rate of decay in the excitation signal amplitude after the extremum is reached; (ii) the user can select the gain k properly to improve the convergence speeds (see, e.g., Krstic [138]).

2.3 Illustrative Examples

2.3.1 Example-1

We adopt the example given in Tan et al. [26] to illustrate the performance of the proposed schemes. Consider the following SISO system

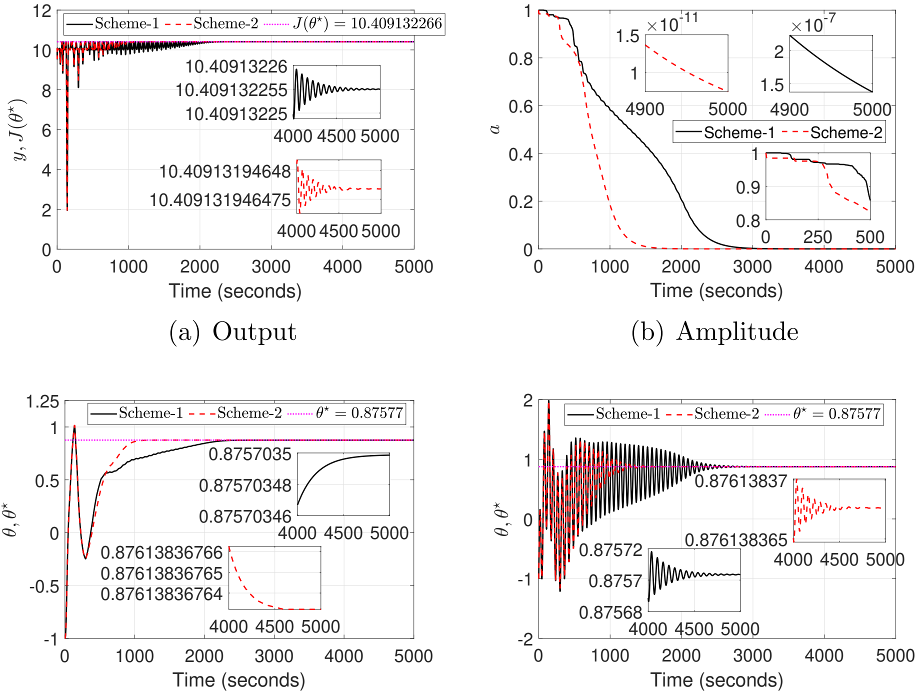
$$\dot{x}_1 = -x_1 + x_2, \quad \dot{x}_2 = x_2 + u, \quad y = h(\boldsymbol{x}), \tag{2.16}$$

where $h(\boldsymbol{x}) = -(x_1 + 3x_2)^4 + \frac{8}{15}(x_1 + 3x_2)^3 + \frac{5}{6}(x_1 + 3x_2)^2 + 10$ and the control input is chosen as $u = -x_1 - 4x_2 + \theta$. Moreover, we have the objective function given by

$$J(\theta) = -\theta^4 + \frac{8}{15}\theta^3 + \frac{5}{6}\theta^2 + 10. \tag{2.17}$$

This function has a global maximum at $\theta^\star = 0.87577$, a local minimum at $\theta^\star = 0$, and a local maximum at $\theta^\star = -0.47577$, as shown in Fig. 2.3 (left). The global maximum value $J(\theta^\star)$ is 10.409132266. The simulation results for the proposed ESC schemes, with $t_0 = 0$, $g_1(a) = g_2(a) = a$, are shown in Figs. 2.4, 2.5. We choose $\hat{\theta}(t_0) = -1$ for both the schemes in order to check if the proposed ESC schemes are able to achieve global maximum despite the presence of a local maximum at $\theta = -0.47577$. The parameter values utilized for the results are as follows: (i) for both the schemes, we utilize $a_0 = 1$, $\omega'_H = K' = 15$, $\omega'_L = 5$, $\lambda'_1 = \lambda'_2 = 30$, $\omega = \epsilon = 0.1$, $\delta = 0.02$; (ii) for scheme-1, we take $\gamma_1 = 11$, $\boldsymbol{Q} = 0.01\boldsymbol{I}$, $r = 0.01$; (iii) for scheme-2, we take $\gamma_2 = 25$. Also, for both the schemes, we choose $x_1(t_0) = x_2(t_0) = 0$. As shown in Fig. 2.4, both the schemes are successful in converging to the global maximum, bypassing the local extrema at $\theta = 0$ and $\theta = -0.47577$. Note that the zoomed-in plots in Figs. 2.4(a), 2.4(c), 2.4(d) illustrate the attenuation in the oscillations or variations of the respective quantities after convergence to the global maximum. Fig. 2.4(a) shows that the system outputs converge to a small neighborhood of the global maximum. Fig. 2.4(b) depicts the desired two phase decay of the excitation signal amplitudes (see Remark 2.2.2). In addition, $\hat{\theta}$s for both the schemes converge (approximately) within 0.04 % of the true global maximum (Fig. 2.4(c)). $\hat{\theta}$s converge to and remain in a small neighborhood of the global maximum starting from 2000 and 1000 seconds (approximately) for scheme-1 and scheme-2, respectively (Fig. 2.4(c)). Also, the excitation signal amplitudes are attenuated after converging to a neighborhood of the global maximum and θs converge to a neighborhood of $\hat{\theta}$s for both the schemes (Fig. 2.4(d)).

Furthermore, Fig. 2.5 shows the error associated with the estimated gradient magnitude (Kalman Filter). It is observed from Fig. 2.5 that the error goes down after the initial transients. This essentially verifies the Assumption 2.1.4. Tuning the parameters possibly would result in an improved Kalman Filter performance. But,



(a) Output

(b) Amplitude

(c) Estimated extremum $\hat{\theta}$

(d) Reference θ

Figure 2.4: Simulation results for the proposed ESC schemes (Example-1).

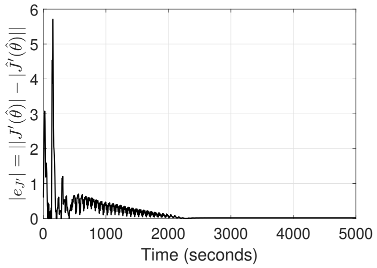
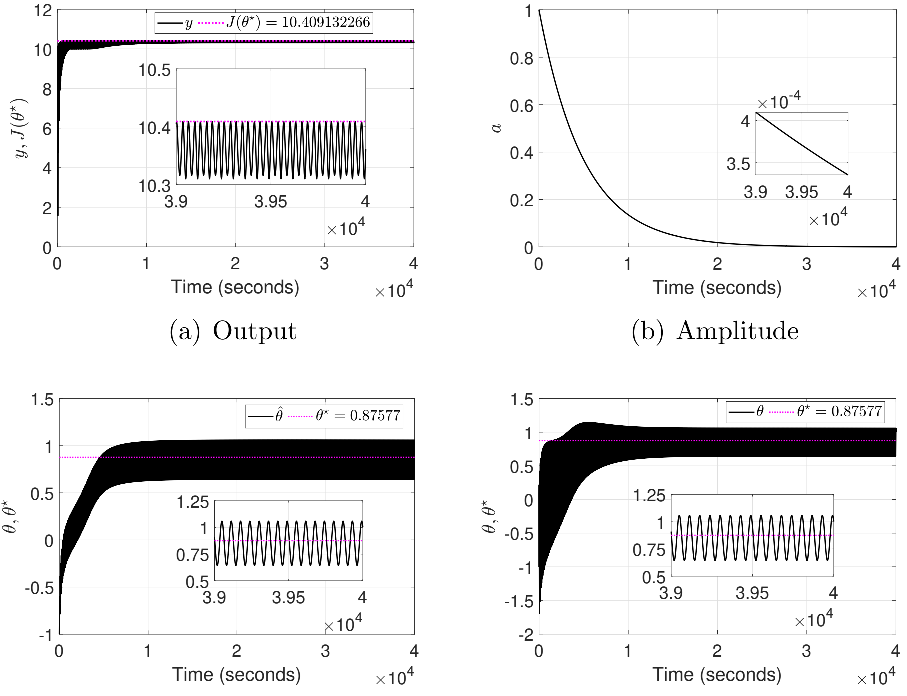


Figure 2.5: Error in the estimated gradient magnitude for the proposed ESC scheme-1 (Example-1).

for the proposed scheme-1, that is not required and selecting γ_1, λ_1' properly would suffice for convergence to the extremum, as shown in the results.

For the sake of comparison, simulation results corresponding to the scheme in Tan et al. [26] are depicted in Fig. 2.6 where we have utilized $\omega = \epsilon = 0.1$, $\delta = 0.02$, $g(a) = a$, $\hat{\theta}(t_0) = -1$, and $x_1(t_0) = x_2(t_0) = 0$. For these results, we have chosen $a_0 = 1$ which is sufficiently large to satisfy Assumption 4 in Tan et al. [26] (see Fig. 2.3 (right)). It can be observed that convergence to the global extremum has been achieved and the amplitude of the excitation signal decays to a small magnitude. However, $\hat{\theta}$ and θ keep oscillating about the extremum value (Figs. 2.6(c) and 2.6(d)). The amplitudes of these oscillations are approximately 0.2, which is roughly equal to $\delta J(\theta^\star)$ (see Remark 2.2.9). As a result of the oscillations in θ, the output of the system keeps oscillating about the extremum value. Overall, the steady-state oscillations shown in Fig. 2.4 are significantly smaller compared to the results shown in Fig. 2.6. Also, $\hat{\theta}$ converges to and remains in a neighborhood of the global maximum starting from 5000 seconds (approximately). Thus, the proposed



(a) Output

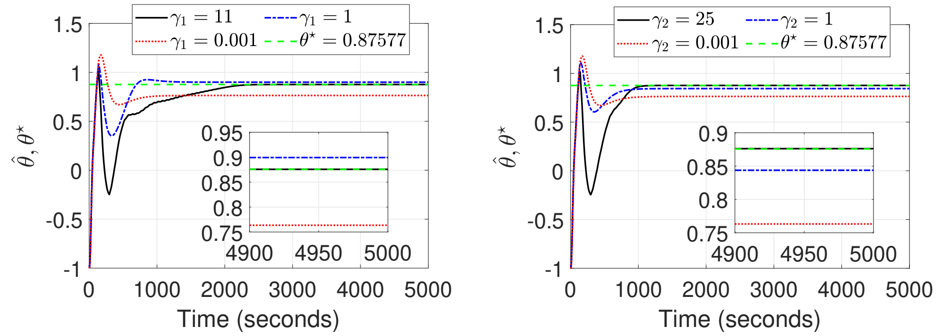
(b) Amplitude

(c) Estimated extremum $\hat{\theta}$

(d) Reference θ

Figure 2.6: Simulation results for the ESC scheme in Tan et al. [26] (Example-1).

schemes have faster convergence speeds to the extremum compared to the scheme in Tan et al. [26].



(a) $\hat{\theta}$ (Scheme-1)

(b) $\hat{\theta}$ (Scheme-2)

Figure 2.7: Simulation results for the proposed ESC schemes with different values of γ_1 and γ_2 (Example-1).

Figs. 2.7(a), 2.7(b) show results corresponding to the proposed schemes for different values of γ_1, γ_2, while all other parameters and conditions are kept the same (the ones utilized for the results in Figs. 2.4, 2.5). These results illustrate the following: (i) the proposed schemes are able to bypass the local extrema and reach a neighborhood of the global maximum for all the γ_1, γ_2 values chosen; (ii) it is possible to converge arbitrarily close to the global maximum by appropriate tuning of γ_1 and γ_2. The differences in performance, as shown in Fig. 2.7, can be attributed to the exponents in the adaptation laws, which utilize the gradient measures.

2.3.2 Example-2

For this example, we consider the same system as in Example-1 and choose a different objective function, given by

$$J(\theta) = -\theta^4 - \theta^3 + 20\theta^2 - 3\theta - 4, \tag{2.18}$$

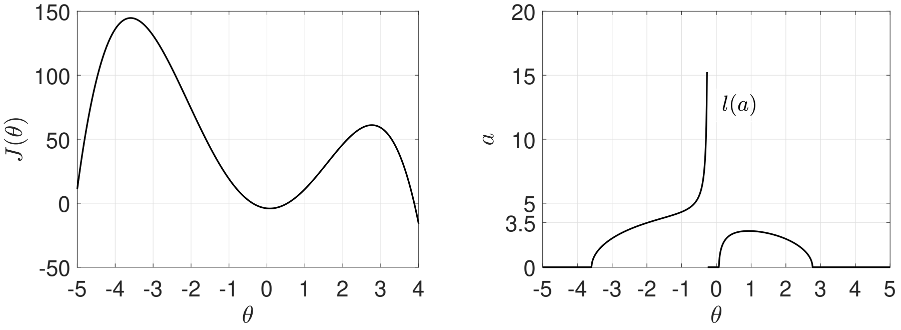
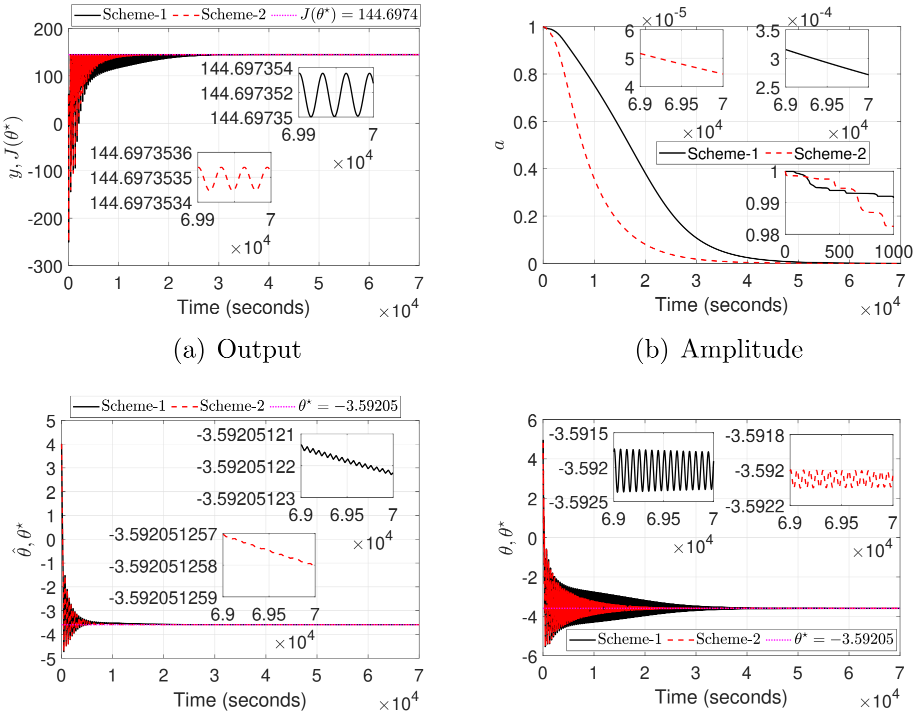


Figure 2.8: The objective function in (2.18) and the corresponding bifurcation diagram which is required for the scheme in Tan et al. [26].

which has a local maximum at $\theta = 2.76658$, a local minimum at $\theta = 0.07547$, and the global maximum at $\theta = -3.59205$, as shown in Fig. 2.8 (left). The global maximum value $J(\theta^{\star})$ is 144.6974. The simulation results for the proposed ESC schemes, with $t_0 = 0$, $g_1(a) = g_2(a) = a$, are shown in Fig. 2.9. We choose $\hat{\theta}(t_0) = 4$ for both the schemes in order to check if the proposed ESC schemes are able to achieve global maximum despite the presence of a local maximum at $\theta = 2.76658$. The parameter values utilized for the results are as follows: (i) for both the schemes, we utilize $a_0 = 1$, $\omega'_H = K' = 6$, $\omega'_L = \lambda'_1 = \lambda'_2 = 2$, $\omega = \epsilon = 0.1$, $\delta = 0.0075$; (ii) for scheme-1, we take $\gamma_1 = 0.1$, $\boldsymbol{Q} = 0.01\boldsymbol{I}$, $r = 0.01$; (iii) for scheme-2, we take $\gamma_2 = 1$. Also, for both the schemes, we choose $x_1(t_0) = x_2(t_0) = 0$. As shown in the results, both the schemes are successful in converging to the global maximum, bypassing the local extrema at $\theta = 2.76658$ and $\theta = 0.07547$. Again, the steady-state oscillations are attenuated to small amplitudes.

The results corresponding to the scheme in Tan et al. [26] are depicted in Fig. 2.10. For these results, we have used $\omega = \epsilon = 0.1$, $\delta = 0.0075$, $\hat{\theta}(t_0) = 4$, and $x_1(t_0) = x_2(t_0) = 0$ with $t_0 = 0$ seconds. Also, we choose $g(a) = a$ with $a_0 = 3.5$, which is sufficiently large to satisfy Assumption 4 in Tan et al. [26] (see Fig.



(a) Output

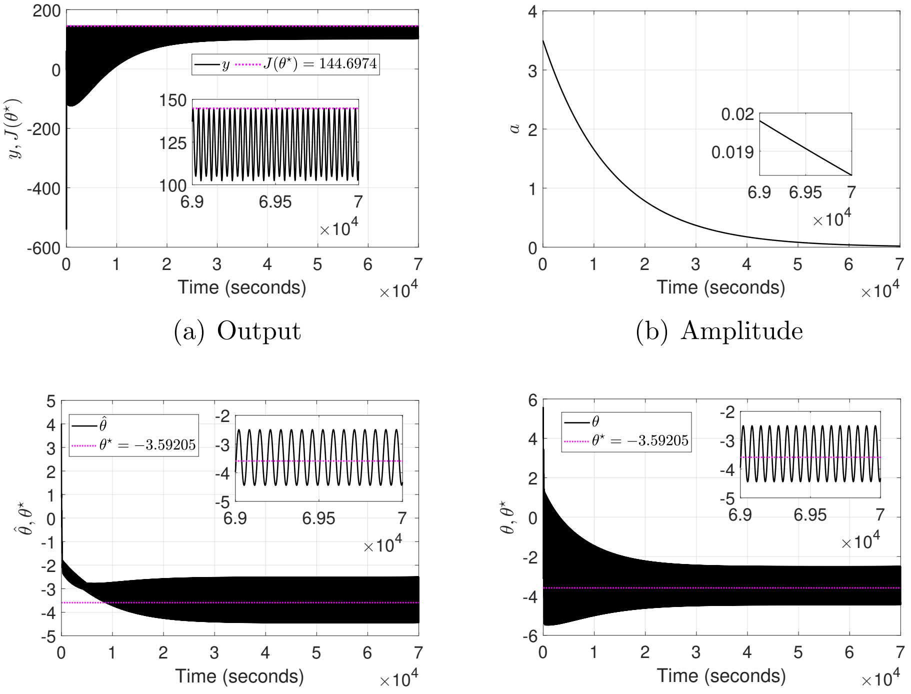
(b) Amplitude

(c) Estimated extremum $\hat{\theta}$

(d) Reference θ

Figure 2.9: Simulation results for the proposed ESC schemes (Example-2).

2.8 (right)). The ESC scheme is able to converge in a neighborhood of the global maximum, as shown in these results. However, we observe that θ keeps oscillating with an amplitude of approximately 1 (which is again roughly equal to $\delta J(\theta^\star)$ in this case), even when a is reduced to small values. As a result, the output oscillates about the maximum with a large amplitude. Note that $\hat{\theta}$s converge to and remain in a small neighborhood of the global maximum starting from 6000 seconds (approximately) for both the proposed schemes (Fig. 2.9(c)). In comparison, $\hat{\theta}$ for the scheme in Tan et al. [26] (see Fig. 2.10(c)) converges to and remains in a neighborhood of the global maximum starting from 10000 seconds (approximately). Thus, the proposed schemes have faster convergences speed to the extremum for this example as well.



(a) Output

(b) Amplitude

(c) Estimated extremum $\hat{\theta}$

(d) Reference θ

Figure 2.10: Simulation results for the ESC scheme in Tan et al. [26] (Example-2).

2.4 Chapter Summary

In this chapter, we have proposed two perturbation-based ESC schemes that are structurally similar to the classical ESC scheme. Moreover, we have proposed two novel adaptation laws for the amplitude of the excitation signal that would make the amplitude converge to zero once the extremum is reached. We have shown that it is possible for the proposed ESC schemes to achieve practical asymptotic convergence to the extremum. We have provided illustrative examples that show the effectiveness of the proposed schemes. There were two key observations from the examples: (a) the proposed schemes were able to bypass the local extremum (both local minimum and local maximum) and converge to the global maximum; (b) the steady-state oscillations in the outputs and reference inputs were attenuated to small values.

Chapter 3

A Flight Mechanics-Based Justification of the Unique Range of Strouhal Numbers for Avian Cruising Flight[*]

In this chapter, we provide details of the model proposed for avian-scale forward flapping flight. First, the proposed model is described. The following section includes derivation and verification of the analytical cycle-averaged force expressions for a rectangular planform. This section also elaborates the numerical integration scheme along with the cycle-averaged power and propulsive efficiency definitions and results. Next, the results corresponding to a representative avian wing planform are given. A discussion on the results obtained for both the rectangular and avian wing planforms is provided. Both sets of results are utilized to provide an explanation to the unique range of Strouhal numbers utilized in avian cruising flight. Finally, a summary of the results in the chapter is provided in the last section.

List of symbols used in the chapter

$(\cdot)(y)$	the quantity $(\cdot)$ corresponding to a section located at a distance y outboard of the wing root
AR	aspect ratio

[*] The materials of this chapter have been published in D. Bhattacharjee and K. Subbarao, "A Flight Mechanics-Based Justification of the Unique Range of Strouhal Numbers for Avian Cruising Flight," *Proc IMechE Part G: Journal of Aerospace Engineering*, p. 0954410020976597, 2020, DOI: 10.1177/0954410020976597 (reference [129]), and these are reproduced here under the permission granted by SAGE Publishing author re-use policies (Online, accessed in June, 2021: `https://us.sagepub.com/en-us/nam/journal-author-archiving-policies-and-re-use`).

b, $c(y)$	span or length of one wing and sectional chord length, respectively
e	Oswald's efficiency factor
$F_x(y)$, $F_y(y)$, $F_z(y)$	instantaneous sectional forces along the x, y, and z axis of the wing frames, respectively
$F_T(y)$, $F_L(y)$	instantaneous net thrust and lift forces at a section, respectively
$\bar{F}_T(y)$, $\bar{F}_L(y)$	cycle-averaged net thrust and lift forces at a section, respectively
$\bar{F}_T$, $\bar{F}_L$, $\bar{F}_D$	cycle-averaged net thrust, lift, and drag forces for a wing, respectively
$\boldsymbol{f_{R_B}}(y)$	sectional forces on the right wing, expressed in the body frame
$\boldsymbol{f_{L_B}}(y)$	sectional forces on the left wing, expressed in the body frame
$\boldsymbol{f_B}(y)$	net force generated at the sections on the right and left wings, expressed in the body frame
f	flapping frequency
$H_\nu(z)$	Struve function of order ν and argument z
$h(y)$, $\dot{h}(y)$	transverse displacement and speed at a section due to plunging, respectively
$J_\nu(z)$	Bessel function of the first kind of order ν and argument z
$P(y)$, $\bar{P}(y)$	instantaneous sectional power and cycle-averaged sectional power, respectively

$\bar{P}$	cycle-averaged power for the wing
St_y, St_0, St	scaled sectional Strouhal number, scaled Strouhal number for the wing, and Strouhal number, respectively
T	time period of flapping
V_∞	freestream speed
x_a	distance between the aerodynamic center and twist axis of a section, normalized with respect to $c(y)$
x_B, y_B, z_B	x, y, and z axes of the body frame, respectively
$\dot{z}(y)$	transverse speed at a section
$\alpha(y)$, $\dot{\alpha}(y)$, $\ddot{\alpha}(y)$	local angle of attack and its time derivatives
δ, $\dot{\delta}$, $\ddot{\delta}$	plunging angle, angular speed, and angular acceleration, respectively
κ	location of chordwise flow separation point on the upper surface of a section, normalized with respect to $c(y)$
ω_f	angular frequency of flapping
θ	twist angle
ζ	lagging or sweeping angle

Subscripts :

R	right wing frame
L	left wing frame
B	body frame

3.1 Preliminaries

In this section, the formulation required for the calculation of cycle-averaged forces, power, and propulsive efficiency is provided. While the formulation given in

Bhattacharjee and Subbarao [36] addressed the 2D airfoil, the extensions to the finite wing are addressed here.

3.1.1 Frames of Reference

In our formulation, three frames of reference are utilized. These are attached to the left and right wings, and the central fuselage (or body). Also, positive directions for the x, y, z axes of these frames are forward, towards the starboard side, and downwards, respectively (see figures 3 and 4 in Bhattacharjee and Subbarao [36]). A typical flapping vehicle undergoing rigid flapping motion has 6 degrees of freedom (DOFs) with the wings being holonomically constrained to the body (Orlowski and Girard [139]). It should be noted that the governing equations of motion are expressed in the body frame and so are the sectional lift and drag forces. The rotation matrices used to express the sectional lift and drag forces in the body frame are introduced in the next subsection.

3.1.2 Rotation Matrices

Rotation matrices for each wing, similar to the ones given in Orlowski and Girard [139], have been adopted with the 3-1-2 Euler angle sequence. The sweeping or lagging (3), plunging (1), and twisting (2) angles are the Euler angles. These rotation matrices transform the representation of a vector from the body frame to the wing frames. Plunging angles (δ_R or δ_L) are assumed positive down and lagging angles (ζ_R or ζ_L) are assumed to be positive forward. Similarly, we assume twisting up to be positive for the twist angles (θ_R or θ_L). As a result, the rotation ζ_R is negative for the right wing and the rotation δ_L is negative for the left wing. The detailed representation of the rotation matrices for the right and left wings are given in Bhattacharjee and Subbarao [36].

3.1.3 Aerodynamic Force Model

A strip theory-based formulation utilizing quasi-steady aerodynamics has been developed. Sectional or local lift forces are modeled using Goman and Khrabrov's model [140]. Hence, the sectional lift coefficient is given by

$$C_l\left(y\right) = C_{l_0}\sin\left[\left(1+\sqrt{\kappa}\right)^2\alpha\left(y\right)\right] = C_{l_0}\sin\left[\left(1+2\sqrt{\kappa}+\kappa\right)\alpha\left(y\right)\right] \tag{3.1}$$

where $C_{l_0} = \frac{\pi}{2}$ (Goman and Khrabrov [140]). The parameter κ denotes the location of flow separation on the upper surface of the section and varies between 0 and 1, where 0 signifies leading edge separation and 1 signifies trailing edge separation (Goman and Khrabrov [140]). The dynamic behavior of the separation point under unsteady flow conditions was studied in Goman and Khrabrov [140]. However, we assume the flow separation point to be static during the flapping cycle. The $C_l(y)$ formula in equation (3.1) is directly related to the well-known formula in thin airfoil theory. Assuming $\left(1+2\sqrt{\kappa}+\kappa\right)\alpha\left(y\right)$ to be small, the lift coefficient can be approximated as follows: $C_l(y) \approx C_{l_0}\left(1+2\sqrt{\kappa}+\kappa\right)\alpha\left(y\right) = \frac{\pi}{2}\left(1+2\sqrt{\kappa}+\kappa\right)\alpha\left(y\right)$. Thus, for $\kappa = 1$, one retrieves the well-known lift coefficient formula given by thin airfoil theory as $C_l(y) = 2\pi\alpha(y)$. As κ decreases from 1, the slope $\frac{dC_l}{d\alpha}$ decreases for a given value of α such that the change in the lift coefficient gets attenuated for a change in the angle of attack $\Delta\alpha$. Note that, for all the analysis presented in this chapter, κ values are restricted between 0.5 and 1, since a more elaborate model is required to account for the effects of flow separation close to the leading edge and development of an LEV.

Remark 3.1.1. *Tuncer and Platzer [141] carried out numerical simulations for a flapping airfoil and reported high propulsive efficiencies to be associated with cases where the flow stays mostly attached over the upper and lower surfaces of the airfoil for the entire flapping cycle. Further, Taylor et al. [28] argued that the St range observed in nature would likely correspond to high propulsive efficiencies as natural* *selection would favor that. By combining these two arguments, it is likely that the above-mentioned St range corresponds to mostly attached flows for the entire duration of flapping cycle. Following along similar lines, we assume the flow separation point to be static during the flapping cycle as one of the goals of this research is to explain this unique range of Strouhal numbers observed in nature. Also, for the completeness of the analysis, we consider κ values between 0.5 and 1, as mentioned earlier.*

Withers [142] showed that a parabolic drag polar is suitable for modeling the drag of avian wings by curve fitting experimentally obtained data. Paranjape et al. [143] utilized a parabolic drag polar for sectional drag modeling as well. Utilizing these examples from the existing literature, we model the sectional drag coefficient as

$$C_d(y) = C_{d_0} + KC_l^2(y) \tag{3.2}$$

where C_{d_0} is the profile drag coefficient and K is a proportionality constant that depends on the airfoil of the section, as shown in Hall et al. [144] Ideally, the value of K should be determined by performing wind-tunnel tests or by using CFD and will only be applicable for the airfoil studied. Paranjape et al. [143] proposed a solution to this problem by equating K with the induced drag coefficient of the flapping wing. This allows for the formulation to be generic and the same approach is adopted here. Hence, the proportionality constant K is given by $K = \frac{1}{\pi e AR}$. Withers [142] suggested that e for bird wings are typically lower (between 0.3 and 0.8) compared to aircraft wings (typically between 0.9 and 0.95). The profile drag coefficient can be decomposed into pressure drag coefficient (C_{d_p}) and skin friction drag coefficient (C_{d_f}). For $\kappa = 1$ (fully attached flow), the pressure drag is assumed negligible compared to the skin friction drag. In contrast, for $\kappa < 1$ (separated flow on the upper surface of the section), the pressure drag component can no longer be neglected. Motivated by the above discussion, the profile drag coefficient is expressed using an empirical formulation as

$$C_{d_0} = C_{d_f} + C_{d_p} = \frac{2}{\kappa^{\gamma}} \left(\frac{0.074}{(Re)^{\frac{1}{5}}} \right), \ \forall \kappa \in [0.5, 1] \tag{3.3}$$

where C_{d_f} is modeled assuming turbulent flow conditions (Ellington [145]) and as a function of the Reynolds number (Re) based on the sectional chord (Hoerner [146]), and $\gamma > 1$ is a positive constant that accounts for the increase in the drag coefficient under separated flow conditions due to pressure drag (C_{d_p}). Also, by substituting $\kappa = 1$ in equation (3.3), one retrieves the skin friction drag coefficient (Hoerner [146]). Note that γ can be determined using experimental or numerical data. The skin friction drag coefficient formula in equation (3.3) is valid for Reynolds numbers below 10^6 (Hoerner [146]) which is suitable for the current analysis as we are interested in the the Reynolds number range between 10^4 and 10^5. A verification of the above force modeling is presented next.

In order to demonstrate a verification of the above force modeling in the Reynolds number regime of interest, results of the proposed model are compared with the numerical and experimental data given in Viieru et al. [147] Figure 3.1 illustrates the comparison results and it can be concluded that the results of the proposed force modeling are comparable with the numerical and experimental data. Note that the discrepancy in the results for angles of attack (AoA) very close to zero is due to the lift coefficient being approximately zero, as shown in equation (3.1). Except for that, the results of the proposed model match adequately with the published data. This serves as a verification of the proposed force modeling.

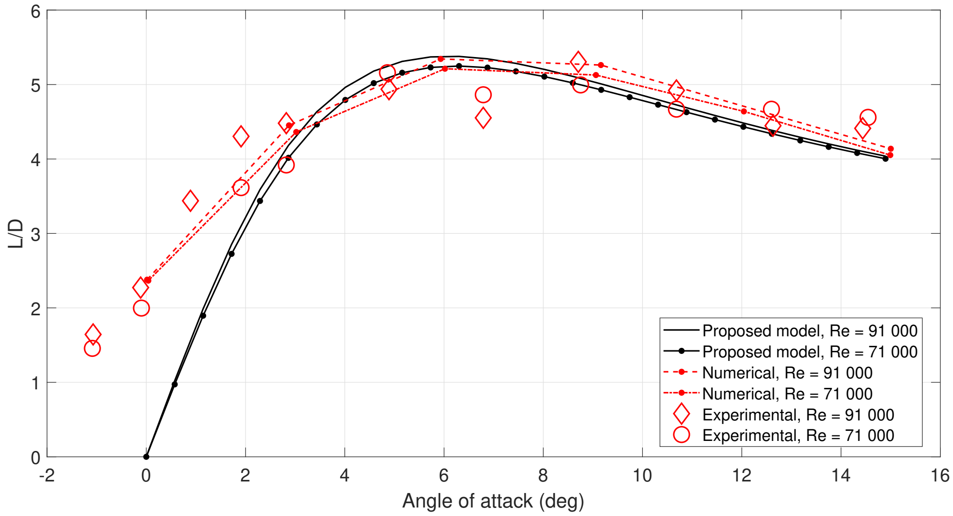


Figure 3.1: Lift-to-drag ratio plots for the proposed force model and the data given in Viieru et al. [147] For the proposed model, $e = 1$ (approximately elliptical planform), $\gamma = 2.5$, and $\kappa = 0.625$ were utilized.

3.1.4 Flapping Kinematics and Strouhal Number

For the present study, symmetrical, sinusoidal flapping kinematics are assumed. Thus, the angles induced due to the motions of flapping wings are equal in magnitude for the right and left wings. Since the sweeping motion is negligible for avian-scale forward flight (Paranjape et al. [148]), it has not been incorporated in the current analysis. The kinematics for plunging and twisting are given by (Bhattacharjee and Subbarao [36])

$$\begin{aligned} \delta_R = \delta_L = \delta = -\delta_0 \cos \omega_f t = \delta_0 \sin\left(\frac{3\pi}{2} + \omega_f t\right) \\ \theta_R = \theta_L = \theta = \bar{\theta} - \theta_0 \sin \omega_f t = \bar{\theta} + \theta_0 \sin\left(\pi + \omega_f t\right) \end{aligned} \tag{3.4}$$

where δ_0 is the amplitude of plunge angle, θ_0 is the amplitude of the twist profile, and $\bar{\theta}$ is the mean twist angle. It should be noted that the twist profile in equation (3.4) is constant across the span. Also, the phase difference between plunging and twisting motions in equation (3.4) has correlation with optimality of flapping (DeLaurier [42], Tuncer and Kaya [149], Isogai et al. [150]). Note that we only consider positive (or non-negative) values for θ_0 and δ_0 to maintain this favorable phase difference.

The transverse displacement and speed at a section are given by

$$\begin{aligned} h(y) &= y\sin\delta \approx y\delta = -y\delta_0\cos\omega_f t \\ \dot{h}(y) &= y\dot{\delta} = \omega_f y\delta_0 \sin\omega_f t \\ \dot{z}(y) &= \dot{h}(y)\cos\theta\cos\delta \approx \dot{h}(y) = (\omega_f y\delta_0)\sin\omega_f t \end{aligned} \tag{3.5}$$

where $\dot{h}(y)$ is aligned with the z_R or z_L axes and $\dot{z}(y)$ is aligned with the z_B axes (see figure 3 in Bhattacharjee and Subbarao [36]). The local AoA at a section are given by (Paranjape et al. [34], DeLaurier [42], Bhattacharjee and Subbarao [36])

$$\alpha(y) = \left(\theta + \arctan\left(\frac{\dot{z}(y)}{V_\infty}\right)\right) \approx \left(\theta + \arctan\left(\frac{\dot{h}(y)}{V_\infty}\right)\right) \approx \left(\theta + \frac{\dot{h}(y)}{V_\infty}\right). \tag{3.6}$$

The approximations introduced in equations (3.5) and (3.6) are commonly utilized by the candidate models in the existing literature (Paranjape et al. [34], DeLaurier [42]) and help make the problem analytically tractable. Next, we define a scaled sectional Strouhal number based on the plunge amplitude of the airfoil section as

$$St_y = \frac{\omega_f y\delta_0}{V_\infty}.$$

For a finite wing undergoing flapping motions, Strouhal number is defined in terms of the amplitude of the mid-span as (Heathcote et al. [35], Paranjape et al. [34])

$$St = \frac{2\,(0.5b\delta_0)\,f}{V_\infty} = \frac{bf\delta_0}{V_\infty}.$$

A scaled Strouhal number for the wing is defined and the relationships between several definitions of the Strouhal numbers are given by

$$St_0 = \frac{b\omega_f\delta_0}{2V_\infty}, \quad St_y = St_0\left(\frac{2y}{b}\right), \quad St = \frac{St_0}{\pi}. \tag{3.7}$$

We compute the local velocity at a section as the resultant of freestream velocity and velocity due to plunging (Paranjape et al. [34], Bhattacharjee and Subbarao [36]). Therefore, the magnitude of local velocity at a section is given by

$$V(y) = \sqrt{V_\infty^2 + \dot{h}^2(y)} = \sqrt{V_\infty^2 + \dot{\delta}^2 y^2} = V_\infty \sqrt{\left(1 + St_y^2 \sin^2 \omega_f t\right)}.$$

3.1.5 Forces Expressed in the Body Frame and Cycle Averaging

The instantaneous forces at a section of the wing (right or left) are expressed as

$$\begin{aligned} F_x(y) &= L(y)\sin\alpha(y) - D(y)\cos\alpha(y) \\ F_y(y) &= 0 \\ F_z(y) &= -\left(L(y)\cos\alpha(y) + D(y)\sin\alpha(y)\right) \end{aligned} \tag{3.8}$$

where the local lift $(L(y))$ and drag $(D(y))$ forces are as follows

$$\begin{aligned} L(y) &= \frac{1}{2}\rho V^2(y)c(y)C_l(y) + F_a(y) \\ &= \frac{1}{2}\rho V_\infty^2 c(y)\left(1 + St_y^2 \sin^2\omega_f t\right) C_{l_0} \sin\left(\left(1+\sqrt{\kappa}\right)^2 \alpha(y)\right) + F_a(y) \\ D(y) &= \frac{1}{2}\rho V^2(y)c(y)C_d(y) \\ &= \frac{1}{2}\rho V_\infty^2 c(y)\left(1 + St_y^2 \sin^2\omega_f t\right)\left(C_{d_0} + KC_l^2(y)\right) \end{aligned} \tag{3.9}$$

with $F_a(y)$ denoting the force due to the "added-mass effect". $F_a(y)$ models the contribution to the local lift force due to the acceleration of the fluid (or air) surrounding the section. Since this contribution is dependent on the instantaneous acceleration of the section, this term can be added to the quasi-steady lift as shown above. This term is given by (Paranjape et al. [143])

$$F_a(y) = \frac{\pi}{4}\rho c^2(y)\left(\ddot{\delta} y + V_\infty \dot{\alpha}(y) - (x_a - 0.25)c(y)\ddot{\alpha}(y)\right)$$

where $\frac{\pi}{4}\rho c^2(y)$ is the mass per unit length of air surrounding the section and the term inside the parenthesis is the normal acceleration of the section due to the kinematics.

The instantaneous forces generated by two sections, both located at distances y from the respective wing roots, are expressed in the body frame as

$$\begin{aligned}\boldsymbol{f_{R_B}}(y) &= \begin{bmatrix} F_x(y)\cos\theta + F_z(y)\sin\theta \\ \sin\delta\,(F_x(y)\sin\theta - F_z(y)\cos\theta) \\ \cos\delta\,(F_z(y)\cos\theta - F_x(y)\sin\theta) \end{bmatrix} \\ \boldsymbol{f_{L_B}}(y) &= \begin{bmatrix} F_x(y)\cos\theta + F_z(y)\sin\theta \\ -\sin\delta\,(F_x(y)\sin\theta - F_z(y)\cos\theta) \\ \cos\delta\,(F_z(y)\cos\theta - F_x(y)\sin\theta) \end{bmatrix}.\end{aligned} \tag{3.10}$$

Hence, the net instantaneous force generated by both the sections is given by $\boldsymbol{f_B}(y) = \boldsymbol{f_{R_B}}(y) + \boldsymbol{f_{L_B}}(y)$. Therefore,

$$\boldsymbol{f_B}(y) = \begin{bmatrix} 2F_x(y)\cos\theta + 2F_z(y)\sin\theta \\ 0 \\ 2\cos\delta\,(F_z(y)\cos\theta - F_x(y)\sin\theta) \end{bmatrix}.$$

As shown above, net force acting along the y_B axis is zero. Also, equal magnitude of forces are generated along the x_B and z_B axes by both the sections. These results are expected under symmetrical flapping conditions, as mentioned in Orlowski and Girard [139]. Substituting for the expressions of $F_x(y)$ and $F_z(y)$ given in equation (3.8), $\boldsymbol{f_B}(y)$ can be expressed as

$$\boldsymbol{f_B}(y) = \begin{bmatrix} F_{x_B}(y) \\ F_{y_B}(y) \\ F_{z_B}(y) \end{bmatrix} = \begin{bmatrix} 2L(y)\sin(\alpha(y)-\theta) - 2D(y)\cos(\alpha(y)-\theta) \\ 0 \\ -2\cos\delta\,[L(y)\cos(\alpha(y)-\theta) + D(y)\sin(\alpha(y)-\theta)] \end{bmatrix}. \tag{3.11}$$

Therefore, the forces due to *a section* (either of the right wing or the left wing) along the x_B and z_B axis are given by $\frac{1}{2}F_{x_B}(y)$ and $\frac{1}{2}F_{z_B}(y)$ respectively, where $F_{x_B}(y)$ and $F_{z_B}(y)$ are as shown in equation (3.11). Now, the freestream AoA are assumed negligible in the present study and only the AoA due to the flapping motions are considered. With that, it is easy to check that the freestream velocity vector, net thrust and lift forces act along the negative x_B axis, positive x_B axis, and negative z_B axis, respectively. Therefore, the instantaneous sectional net thrust and lift are expressed as

$$F_T(y) = \frac{1}{2}F_{x_B}(y), \quad F_L(y) = -\frac{1}{2}F_{z_B}(y). \tag{3.12}$$

Finally, we calculate the cycle-averaged net thrust and lift forces at a section of one of the wings (right or left) as

$$\begin{aligned}
\bar{F}_T(y) &= \frac{1}{T}\int_0^T \frac{1}{2}F_{x_B}(y)dt \\
&= \frac{1}{T}\int_0^T \Big[L(y)\sin\left(\alpha(y)-\theta\right) - D(y)\cos\left(\alpha(y)-\theta\right)\Big]dt \\
\bar{F}_L(y) &= \frac{1}{T}\int_0^T \left(-\frac{1}{2}F_{z_B}(y)\right)dt \\
&= \frac{1}{T}\int_0^T \cos\delta\Big[L(y)\cos\left(\alpha(y)-\theta\right) + D(y)\sin\left(\alpha(y)-\theta\right)\Big]dt.
\end{aligned} \tag{3.13}$$

Hence, the cycle-averaged net thrust and lift due to one of the flapping wings (right or left) are given by

$$\bar{F}_T = \int_0^b \bar{F}_T(y)dy, \quad \bar{F}_L = \int_0^b \bar{F}_L(y)dy. \tag{3.14}$$

Using the expressions given in equations (3.13) and (3.14), the analytical expressions for the cycle-averaged net thrust and lift for a rectangular planform are derived in the next section.

Remark 3.1.2. *The proposed formulation is quasi-steady, i.e., the force calculations do not depend on the time history of force generation and unsteady wake effects. Instead, the force calculation only depends on the instantaneous velocities and accelerations. It is well-known that a quasi-steady model produces reasonably accurate* *results without exhaustive modeling of the flow-field physics and provides a formulation with manageable sets of equations [44]. Further, we are only interested in the forward cruising flight of birds which is not dominated by unsteady phenomena as opposed to more complex flight situations like take-off or landing and hovering of insects. That is why, we are able ignore the effects of unsteady wake, vortex formation & shedding, dynamic stalling etc. in our formulation and obtain reasonably accurate results for the problem at hand using the simple analytical model proposed.*

3.2 Cycle-Averaged Quantities: Rectangular Planform

This section contains the details of deriving analytical expressions for the cycle-averaged forces of a rigid, untapered rectangular wing planform. The expressions derived are verified using a discrete element numerical integration scheme. Cycle-averaged power and propulsive efficiency are defined and computed numerically for the rectangular wing planform.

First, the expressions of cycle-average net thrust and lift forces corresponding to a section of one of the wings (right or left) are derived and subsequently, the calculations are performed for the forces generated by the wings. The following notations will be helpful in describing these expressions:

$$\begin{aligned} p = \left(1+\sqrt{\kappa}\right)^2, a = p(St_y - \theta_0), q_\infty = \frac{1}{2}\rho V_\infty^2, \\ c(y) = c, \alpha(y) \approx \left(\theta + \frac{\dot{\delta} y}{V_\infty}\right) = \bar{\theta} + (St_y - \theta_0)\sin\omega_f t. \end{aligned} \tag{3.15}$$

3.2.1 Cycle-Averaged Net Thrust

Performing the integration shown in equation (3.13) and utilizing the notations introduced in equation (3.15), we obtain (Bhattacharjee and Subbarao [36])

$$\bar{F}_T(y) = \frac{1}{2} q_\infty c C_{l_0} \cos(p\bar{\theta}) \left[St_y^2 \Big(-\frac{J_1(a - St_y)}{(a - St_y)} + \frac{J_1(a + St_y)}{(a + St_y)} \Big) + (1 + St_y^2) J_0(a - St_y) \right.$$

$$
\begin{aligned}
&- (1 + St_y^2) J_0(a + St_y)\Bigg] + (x_a - 0.25)(\frac{\pi}{4}\rho c^3)(St_y - \theta_0)\omega_f^2 J_1(St_y) \\
&- \frac{1}{2} q_\infty c \Bigg[(2C_{d_0} + KC_{l_0}^2)\big((1 + St_y^2) J_0(St_y) - St_y J_1(St_y)\big) \\
&+ \frac{1}{2} KC_{l_0}^2 \cos(2p\bar{\theta}) \Bigg(- (1 + St_y^2) J_0(2a - St_y) - (1 + St_y^2) J_0(2a + St_y) \\
&+ St_y^2 \Big(\frac{J_1(2a - St_y)}{(2a - St_y)} + \frac{J_1(2a + St_y)}{(2a + St_y)} \Big) \Bigg) \Bigg].
\end{aligned}
$$

Next, integrating this expression with respect to the span of the wing would yield the cycle-averaged net thrust generated by the wing. It should be noted that the arguments of the Bessel functions in the expression of $\bar{F}_T(y)$ are functions of the spanwise variable y. To avoid cumbersome expressions, we introduce the following notations.

$$
\begin{gathered}
a - St_y = A_1 + B_1 y, \;\; a + St_y = A_1 + B_2 y, \\
2a - St_y = A_2 + B_3 y, \;\; 2a + St_y = A_2 + B_4 y
\end{gathered}
$$

where

$$
\begin{aligned}
&A_1 = -p\theta_0, \;\; B_1 = (p - 1)\left(\frac{2St_0}{b}\right), \\
&B_2 = (p + 1)\left(\frac{2St_0}{b}\right), \;\; A_2 = -2p\theta_0, \\
&B_3 = (2p - 1)\left(\frac{2St_0}{b}\right), \;\; B_4 = (2p + 1)\left(\frac{2St_0}{b}\right).
\end{aligned}
$$

The integration with respect to the spanwise variable y is performed as shown in equation (3.14) and the cycle-averaged net thrust generated by one of the flapping wings (right or left) is expressed as

$$
\begin{aligned}
\bar{F}_T =& \frac{1}{2} q_\infty c C_{l_0} \cos(p\bar{\theta})\left(I_{L_I} + I_{L_{II}}\right) + F_{a_I} - \frac{1}{2} q_\infty c \Big((2C_{d_0} + KC_{l_0}^2) I_{D_I} \\
&+ \frac{1}{2} KC_{l_0}^2 \cos(2p\bar{\theta})\left(I_{D_{II}} + I_{D_{III}}\right) \Big)
\end{aligned}
\tag{3.16}
$$

where the expressions for the terms $I_{L_I}, I_{L_{II}}, F_{a_I}, I_{D_I}, I_{D_{II}}, I_{D_{III}}$ are given in Appendix C.

Note that the Struve functions appear in the expressions as a result of the integration of Bessel functions of the first kind. In the existing literature, Struve function and Bessel function of the first kind have appeared in the context of cycle-averaged forces in flapping flight (Theodorsen [37], Garrick [38], Doman et al. [151], Orlowski and Girard [152]). Finally, the cycle-averaged net thrust coefficient is given by

$$\bar{C}_T = \frac{2\bar{F}_T}{q_\infty(2b)c} = \frac{\bar{F}_T}{q_\infty bc}. \tag{3.17}$$

3.2.2 Cycle-Averaged Lift

Now, we derive the cycle-averaged lift expression using the equations (3.13) and (3.14). For the derivation, the term $\cos\delta$ cannot be allowed to vary during the cycle and has to be replaced with a constant (Bhattacharjee and Subbarao [36]). The task becomes analytically intractable otherwise. To this end, note that $\cos\delta$ is positive as the plunging angle has to satisfy the constraint $\delta \in [-90, 90]$ deg. Therefore, the approximation obtained by substituting $\cos\delta = \delta_c$ in equation (3.13) with $0 < \delta_c \leq 1$ is reasonable. For the analysis presented here, we take $\delta_c = 1$. Thus, we approximate the sectional cycle-averaged lift as (Bhattacharjee and Subbarao [36])

$$\bar{F}_L(y) = \frac{1}{T}\int_0^T \left(-\frac{1}{2}F_{z_B}\right) dt \approx \frac{1}{T}\int_0^T \Big(L(y)\cos\left(\alpha(y)-\theta\right) + D(y)\sin\left(\alpha(y)-\theta\right)\Big)dt.$$

After carrying out the integration, we derive the following expression (Bhattacharjee and Subbarao [36]):

$$\begin{aligned}\bar{F}_L(y) = \frac{1}{2}q_\infty c\Bigg[&C_{l_0}\sin(p\bar{\theta})\Bigg(- St_y^2\Big(\frac{J_1(a-St_y)}{(a-St_y)} + \frac{J_1(a+St_y)}{(a+St_y)}\Big)\\ &+ (1+St_y^2)J_0(a-St_y) + (1+St_y^2)J_0(a+St_y)\Bigg) + \frac{1}{2}KC_{l_0}^2\sin(2p\bar{\theta})\\ &\times\Bigg(St_y^2\Big(- \frac{J_1(2a-St_y)}{(2a-St_y)} + \frac{J_1(2a+St_y)}{(2a+St_y)}\Big)\\ &+ (1+St_y^2)J_0(2a-St_y) - (1+St_y^2)J_0(2a+St_y)\Bigg)\Bigg].\end{aligned} \tag{3.18}$$

Again, integrating the quantity $\bar{F}_L(y)$ with respect to the spanwise variable y results in the expression for the cycle-averaged lift generated by one of the flapping wings (right or left) and is given by

$$\bar{F}_L = \frac{1}{2}q_\infty c\Big(C_{l_0}\sin(p\bar{\theta})\left(I_{L_{III}} + I_{L_{IV}}\right) + \frac{1}{2}KC_{l_0}^2\sin(2p\bar{\theta})\left(I_{D_{IV}} + I_{D_V}\right)\Big) \tag{3.19}$$

where

$$\begin{aligned}I_{L_{III}} &= \int_0^b\Big[- St_y^2\left(\frac{J_1(a-St_y)}{(a-St_y)}\right) + (1+St_y^2)J_0(a-St_y)\Big]dy = I_{L_I}\\ I_{L_{IV}} &= \int_0^b\Big[- St_y^2\left(\frac{J_1(a+St_y)}{(a+St_y)}\right) + (1+St_y^2)J_0(a+St_y)\Big]dy = -I_{L_{II}}\\ I_{D_{IV}} &= \int_0^b\Big[- St_y^2\left(\frac{J_1(2a-St_y)}{(2a-St_y)}\right) + (1+St_y^2)J_0(2a-St_y)\Big]dy = -I_{D_{II}}\\ I_{D_V} &= \int_0^b\Big[St_y^2\left(\frac{J_1(2a+St_y)}{(2a+St_y)}\right) - (1+St_y^2)J_0(2a+St_y)\Big]dy = I_{D_{III}}.\end{aligned}$$

Similar to the cycle-averaged net thrust coefficient, the cycle-averaged lift coefficient is given by

$$\bar{C}_L = \frac{\bar{F}_L}{q_\infty bc} \tag{3.20}$$

where $\bar{F}_L$ is as shown in equation (3.19).

3.2.2.1 Remarks on the Analytical Expressions

Using these expressions of cycle-averaged net thrust and lift, as shown in equations (3.16) and (3.19) respectively, we make the following observations.

- A positive mean twist angle is required to produce positive cycle-averaged lift. This is not required for positive cycle-averaged net thrust generation.
- Cycle-averaged lift is independent of skin-friction drag, whereas the cycle-averaged net thrust is not. Also, the "added mass effect" contributes to the cycle-averaged net thrust and does not contribute to the cycle-averaged lift.
- The magnitude of force generation can be controlled by changing the speed of flight and the mean twist angle.
- The exact expression for $\bar{F}_L(y)$ is given by

  $$\bar{F}_L(y) = \frac{1}{T}\int_0^T \cos\delta\Big(L(y)\cos(\alpha(y)-\theta) + D(y)\sin(\alpha(y)-\theta)\Big)dt.$$

  Taking any convenient norm $||\cdot||$ leads to

  $$||\bar{F}_L(y)|| \leq \frac{1}{T}\int_0^T ||\cos\delta||\ ||\Big(L(y)\cos(\alpha(y)-\theta) + D(y)\sin(\alpha(y)-\theta)\Big)||dt.$$

  Since $||\cos\delta|| \leq 1$, we have

  $$||\bar{F}_L(y)|| \leq \frac{1}{T}\int_0^T ||\Big(L(y)\cos(\alpha(y)-\theta) + D(y)\sin(\alpha(y)-\theta)\Big)||dt.$$

  This clearly indicates that the magnitudes (positive or negative) of cycle-averaged sectional lift computed using equation (3.18) would be over-approximations of the actual values. Therefore, if one were to implement a numerical scheme to calculate the actual values of $\bar{F}_L(y)$ and $\bar{F}_L$, those would not match exactly with the corresponding estimates computed using equations (3.18) and (3.19). Still, we expect qualitative similarities between these two sets of results because $\cos\delta$ only acts as a positive scaling factor here (see the following subsection).

3.2.3 Verification of the Analytical Expressions Using Numerical Integration

In order to verify the analytical expressions derived in the previous subsections, a numerical integration scheme is required. For that purpose, a discrete element integration technique (Riemann sum) is developed. We remark that this integration technique is similar to the one given in DeLaurier [42] and had been utilized in Bhattacharjee et al. [153] However, we include an example to illustrate the procedure. To this end, the cycle-averaged lift generated by one wing (right or left) is approximated as

$$\bar{F}_L \approx \frac{1}{k}\sum_{j=1}^{k} F_{L_j}$$

where k is the number of time steps, j is the current time step and F_{L_j} is the lift force generated by the wing at jth time step. This is approximated as

$$F_{L_j} \approx \sum_{i=1}^{n} F_{L_{ij}} \Delta y$$

where n is the number of sections for one wing, i is the location of the section along the wingspan, $F_{L_{ij}}$ is the lift generated by the ith section at jth time step, and Δy is the spanwise width of one section. The quantity $F_{L_{ij}}$ is calculated using equations (3.11) and (3.12). Other cycle-averaged quantities are also calculated using a similar formulation in the sequel for the numerical results.

Since the analytical results are derived for a rectangular, untapered planform, a similar planform is chosen for the purposes of verifying those expressions. The values of the parameters implemented for the purposes of verification are given in table 3.2. We assume that the aerodynamic center and the twist axis are coincident at the quarter chord ($x_a = 0$). The cycle-averaged force coefficients corresponding to $\kappa = 1$ are depicted in figure 3.2. It should be noted that the $\arctan(x) \approx x$ approximation for the AoA definition (equation (3.6)) and the assumption $\cos\delta = 1$ for the lift calculation are retained for the numerical integration scheme. This

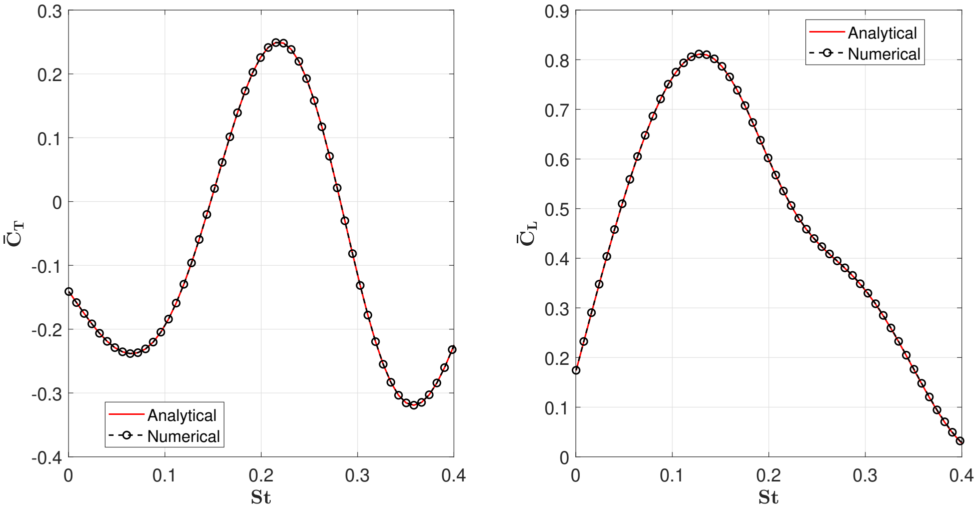


Figure 3.2: Verification of the analytical expressions derived for the cycle-averaged forces. These plots correspond to $\kappa = 1$ and $\cos\delta = 1$.

allows for a comparison between the results obtained by implementing the analytical expressions and the ones obtained numerically. The agreement between both sets of results, as shown in figure 3.2, confirm the accuracy of the analytical expressions derived.

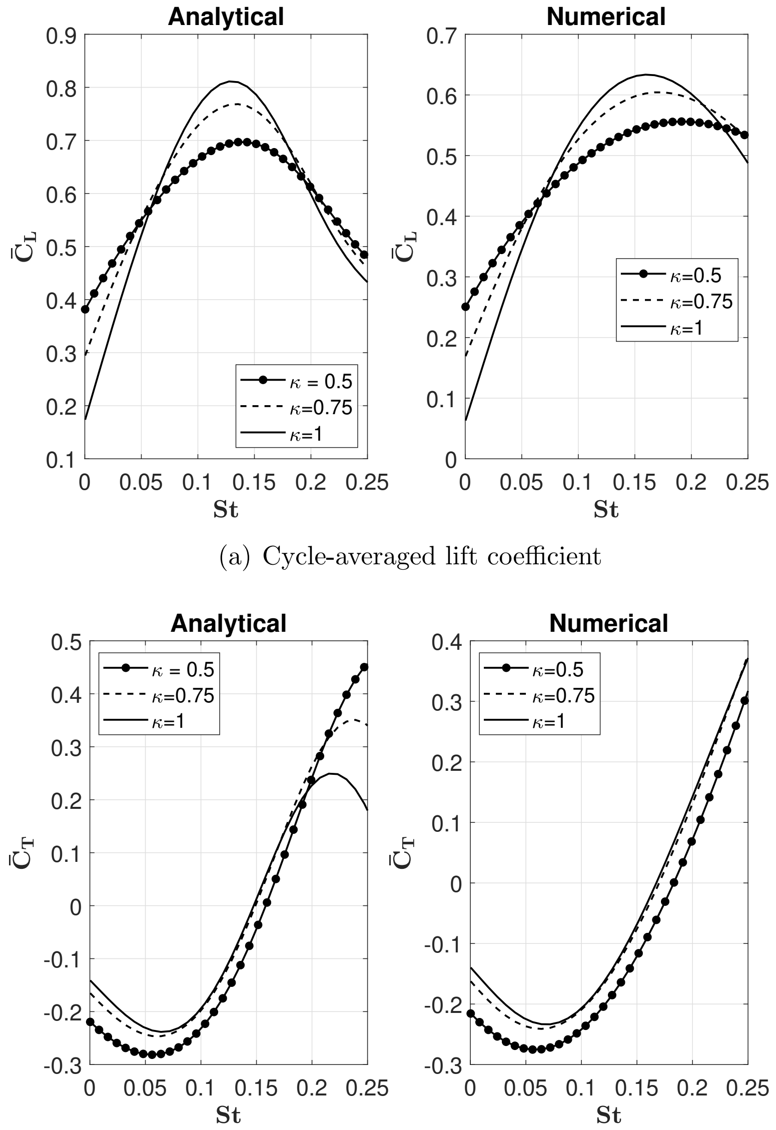
Next, numerically obtained results, without making any of the aforementioned simplifying approximations and assumptions, are compared with the ones obtained by implementing the analytical expressions. This comparison serves as a check for the effective range of validity of the analytical expressions. These results are shown in figures 3.3(a) and 3.3(b). As shown in these results, the two sets of results match qualitatively for Strouhal numbers up to 0.25 approximately (depending on the value of κ). In terms of the cycle-averaged lift coefficient, the analytical results predict a higher maximum value compared to the numerical results because of the $\cos\delta = 1$ assumption introduced in deriving the analytical expressions (figure 3.3(a)). However,

Table 3.2: Parameter values used for the rectangular planform

Item	Value	Units (SI)
ρ	1.225	kg/m^3
b	0.5	m
c	0.15	m
AR	6.67	None
e	0.5	None
f	5	Hz
V_∞	10	m/s
δ_0	60	deg
θ_0	30	deg
$\bar{\theta}$	10	deg
x_a	0	None
γ	2.5	None

the maximum amount of lift corresponds to $\kappa = 1$ in both sets of results, as shown in figure 3.3(a). The positive magnitudes of the cycle-averaged net thrust coefficient are also higher in the results of the analytical expressions (figure 3.3(b)). This is due to the $\arctan(x) \approx x$ approximation introduced in the AoA definition (equation (3.6)). In these plots, the Strouhal numbers corresponding to zero cycle-averaged net thrust and maximum lift are of particular interest. There is a close agreement demonstrated in both sets of results in terms of the St corresponding to zero cycle-averaged net thrust and maximum lift, with the numerical results showing a slightly higher value. For St higher than 0.25, the approximation in the AoA definition is not valid and as a result, there are discrepancies between the analytical and numerical results. Moreover, the effects of this simplification are most pronounced for κ values close to 1 as $p = (1+\sqrt{\kappa})^2$ increases with an increase in κ (see $C_l(y)$ in equation (3.1)).

From figure 3.3(a), we observe that the cycle-averaged lift coefficient decreases with an increase in κ for small values of St. This is counter-intuitive as the lift pro-



(a) Cycle-averaged lift coefficient

(b) Cycle-averaged net thrust coefficient

Figure 3.3: Comparison between the analytical results and numerical results generated without any simplifying assumptions.

duced is expected to go up as the flow separation is delayed towards the trailing edge. However, an explanation for this observation can be provided as follows. Consider the cycle-averaged sectional lift shown in equation (3.18) with $St = St_y = 0$. Substituting for $St_y = 0$ and carrying out the simplifications yield $\bar{F}_L(y) = q_\infty c C_{l_0} \sin(p\bar{\theta}) J_0(-p\theta_0)$. Therefore, the cycle-averaged lift and cycle-averaged lift coefficient are given by

$$\begin{aligned} \bar{F}_L &= q_\infty b c C_{l_0} \sin(p\bar{\theta}) J_0(-p\theta_0), \\ \bar{C}_L &= C_{l_0} \sin(p\bar{\theta}) J_0(-p\theta_0). \end{aligned} \tag{3.21}$$

Taking the partial derivative of $\bar{C}_L$ with respect to p yields

$$\frac{\partial \bar{C}_L}{\partial p} = C_{l_0} \Big[\bar{\theta} \cos(p\bar{\theta}) J_0(-p\theta_0) + \theta_0 \sin(p\bar{\theta}) J_1(-p\theta_0) \Big].$$

Evaluating the above expression for $\kappa = 0.5, 0.75, 1$, we have $\frac{\partial \bar{C}_L}{\partial p} = -0.11, -0.2, -0.26$, respectively. Thus, this mathematically explains why the cycle-averaged lift goes down with an increase in κ, as shown in figure 3.3(a). A fluid mechanics-based explanation for this observation might require concepts like vortex formation & shedding and vortex-wake interactions (see, for example, Cleaver et al. [154]), which are beyond the scope of the present model as pointed out in Remark 3.1.2. $\bar{C}_L$ exhibits a similar trend for $St \approx 0.25$ and a similar (and considerably more elaborate) analysis as given above can be performed to establish a mathematical explanation for that as well.

3.2.4 Cycle-Averaged Power and Propulsive Efficiency

The instantaneous power at a section is calculated using the rate of work done by the aerodynamic forces about the wing hinge or root (Betteridge and Archer [155], Phlips et al. [156]) and is expressed as (Bhattacharjee and Subbarao [36])

$$P(y) = (-F_z(y))\, \dot{h}(y) \tag{3.22}$$

where $F_z(y)$ and $\dot{h}(y)$ are aligned with the z_R or z_L axes (see figure 4 in Bhattacharjee and Subbarao [36]). The expressions for $\dot{h}(y)$ and $F_z(y)$ are given in equations (3.5) and (3.8), respectively. Therefore, the cycle-averaged sectional power and the cycle-averaged power for a wing (right or left) are given by (Bhattacharjee and Subbarao [36])

$$\bar{P}(y) = \frac{1}{T}\int_0^T P(y)dt = \frac{1}{T}\int_0^T \left(L(y)\cos\alpha(y) + D(y)\sin\alpha(y)\right)\dot{h}(y)dt$$
$$\bar{P} = \int_0^b \bar{P}(y)dy$$

where $\dot{h}(y) = \omega_f y\delta_0 \sin\omega_f t = V_\infty St_y \sin\omega_f t$. An analytical expression for $\bar{P}(y)$ can be found in Bhattacharjee and Subbarao [36]. The numerical integration scheme described earlier is implemented to calculate the cycle-averaged power $\bar{P}$ and that is utilized in defining the propulsive efficiency as

$$\bar{\eta} = \frac{\bar{F}_T V_\infty}{\bar{P}}. \tag{3.23}$$

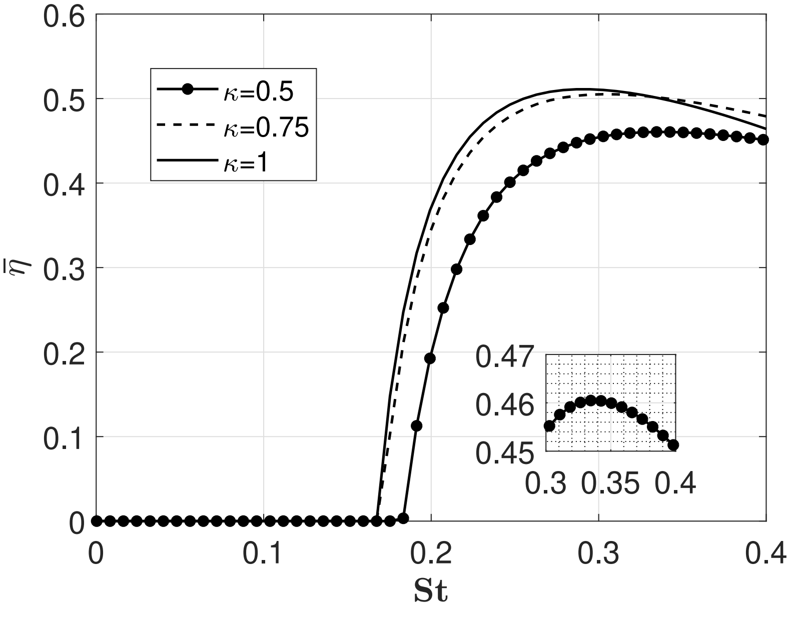


Figure 3.4: Propulsive efficiency as functions of St and κ for the rectangular wing.

It should be noted that no simplifying assumptions or approximations are made while calculating the propulsive efficiency using the numerical integration scheme.

In fact, all the subsequent results shown for the present model are derived numerically without any simplifying assumptions. Also, for all the propulsive efficiency plots shown in this appendix, the entries outside the range of 0 and 1 have been reduced to 0 to improve the visualization in the effective range. Propulsive efficiency plots as functions of St and κ for a flapping wing of rectangular planform (with values of the parameters given in table 3.2) are shown in figure 3.4. Compared to these results, the results shown in Phlips et al. [156] indicate an overprediction in the propulsive efficieny due to the inviscid flow modeling. On the other hand, experimental data in Heathcote et al. [35] show the maximum value of propulsive efficiency (based on cycle-averaged net thrust and power) for a rectangular wing to be approximately equal to 0.2. This discrepancy in the maximum efficiency value is expected as the experimental results would account for the losses in the flow-field. However, figure 3.4 illustrates that the propulsive efficiency maximizes between Strouhal numbers of 0.2 and 0.4, which is consistent with the results in the existing literature (Taylor et al. [28], Heathcote et al. [35]). Also, the maximum propulsive efficiency is obtained for $\kappa = 1$. To this end, although the plot for $\kappa = 0.5$ appears to flatten out for $St \in (0.3, 0.4)$ in figure 3.4, the zoomed-in plot shows that propulsive efficiency attains a maximum approximately at $St = 0.34$.

The corresponding parameter map for propulsive efficiency is shown in figure 3.5. Clearly, the area in the plot that corresponds to the maximum propulsive efficiency is approximately between $St = 0.2$ and $St = 0.4$. This result further supplements the results shown in figure 3.4.

3.3 Results: Avian Wing Planform

In this section, we discuss the results for a representative rigid avian wing planform and wing kinematics. First, a chord distribution representative of avian wings

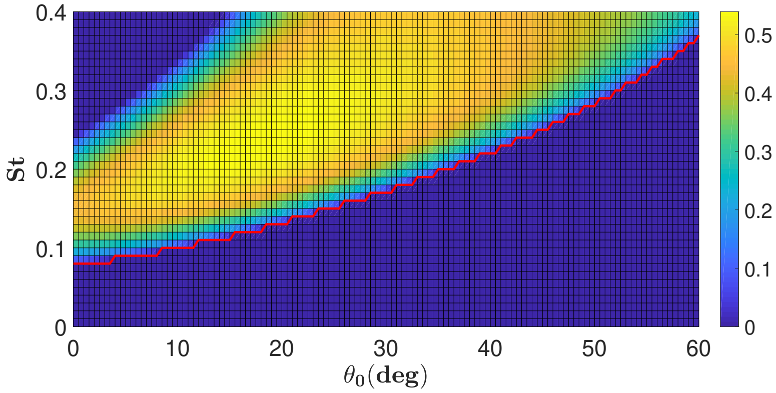


Figure 3.5: Parameter map of propulsive efficiency in terms of St & θ_0 for $\kappa = 1$ and $\bar{\theta} = 10\,\text{deg}$ (the rectangular wing) where the red line denotes $\bar{C}_T = 0$.

is incorporated in the study. According to Oehme and Kitzler [157], a planform with constant chord along the inner half-span of the wing and a parabolically decreasing chord along the outer half-span is a good enough representation of avian wing shapes. Therefore, the chord distribution is given by (Phlips et al. [156], Rayner [40])

$$c(y) = \begin{cases} c_0, \quad \forall y \in \left(0, \frac{b}{2}\right) \\ 4c_0\frac{y}{b}(1-\frac{y}{b}), \quad \forall y \in \left(\frac{b}{2}, b\right) \end{cases} \tag{3.24}$$

Also, other morphological parameters similar to that of a pigeon are adopted from Tobalske and Dial [158]. All the morphological and kinematic parameter values are given in table 3.3. The values of e, ρ, and γ are kept the same as shown in table 3.2. A linear twist profile is introduced for the avian wing planform and the twist angle is given by

$$\theta(y) = \bar{\theta} - y\theta_0 \sin \omega_f t. \tag{3.25}$$

The linear twist distribution is expected to improve the performance of the avian wing planform as the amplitude of twist gradually increases from the wing root towards the wingtips (Bhattacharjee et al. [153]). This is crucial as substantially higher local AoA are induced towards the wingtips (compared to the wing roots) due to the

Table 3.3: Morphological and kinematic parameter values for the avian wing planform

Item	Value	Units (SI)
b	0.279	m
c_0	0.112	m
AR	5.5	None
f	6.5	Hz
V_∞	12	m/s
δ_0	60	deg
θ_0	150	deg /m
$\bar{\theta}$	10	deg
x_a	0	None

plunging motion. Thus, a sufficient amplitude of twist profile, along with a favorable phase angle between the plunging and twisting motions, is required such that the sections located at the outboard parts of the wings produce forces that are favourable during the flapping cycle. The amplitude of time-dependent part of the twist is set at 150 deg /m which is apparently high. But, it should be noted that the maximum twist, i.e., the twist at the wing-tip, would be approximately 42 deg and the average twist angle would be approximately 21 deg. The flight speed chosen is 12 m/s and the corresponding flapping frequency is chosen as 6.5 Hz based on the results in Tobalske and Dial [158]. Parslew and Crowther [159] simulated the crusing flight of pigeons and published results on the amplitude of twisting and plunging angles as functions of flight speed. The magnitude of average amplitude of twist profile chosen for this study (21 deg) is very close to the results in Parslew and Crowther [159], corresponding to the flight speed of 12 m/s. The amplitude of plunging motion is also chosen based on the results in Parslew and Crowther [159] and is given in table 3.3.

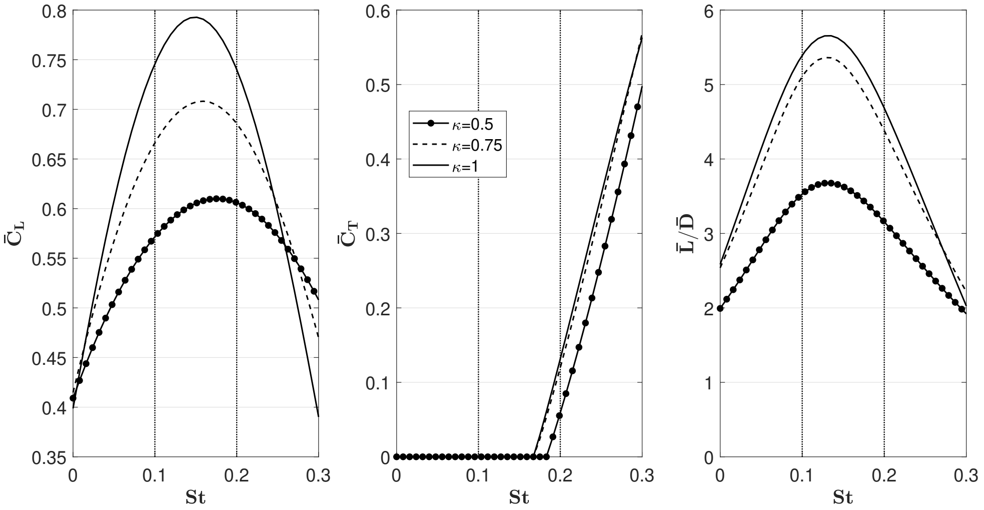


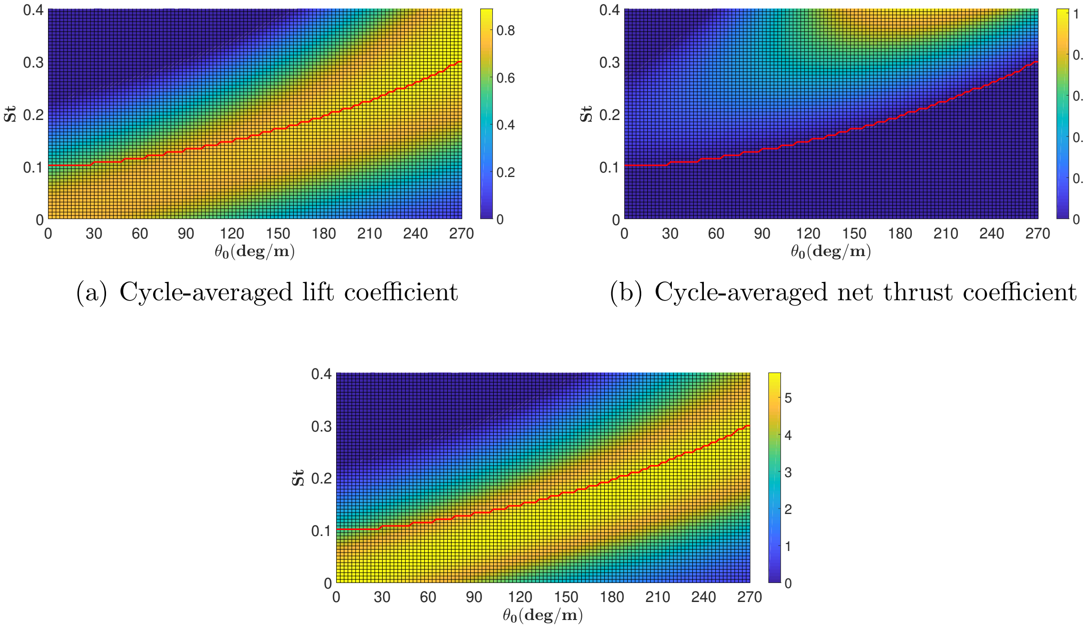
Figure 3.6: Results for the avian wing planform: Cycle-averaged force coefficients and lift-to-drag ratio as functions of St and κ. Note that negative entries are shown as zero.

Lift-to-drag ratio is selected as a performance metric where drag is the component of net thrust that acts opposite to the direction of motion (equation (3.11)). The cycle-averaged drag for one wing (right or left) is given by

$$\bar{F}_D = \int_0^b \left[\frac{1}{T} \int_0^T D(y) \cos\left(\alpha(y) - \theta(y)\right) dt \right] dy \tag{3.26}$$

The cycle-averaged lift and drag are used to define an equivalent lift-to-drag ratio (i.e., $\bar{\mathrm{L}}/\bar{\mathrm{D}} = \bar{F}_L/\bar{F}_D$) for the current analysis. For the values of the parameters given in table 3.3, cycle-averaged force coefficients and lift-to-drag ratio are plotted as functions of St and κ, and are shown in figure 3.6. These results suggest that the lift generation becomes increasingly poor for $St \in (0, 0.25)$ (approximately) as the parameter κ is decreased from 1, i.e., the flow separation point moves closer to the leading edge. It has been argued by Ellington [145] that avian-scale fliers typically operate in the turbulent Reynolds number regime. In turn, the turbulent flow conditions help delay flow separation and preclude the possibility of an LEV formation. Taylor et al. [28] presented a similar argument based on the results in Anderson et al. [30] These results in Anderson et al. [30] indicate formation of a very weak LEV for Strouhal numbers upto 0.2 and only for very high values of maximum AoA. To this end, the discussion in Remark 3.1.1 is also relevant. *Hence, it seems plausible that birds typically operate in such a way that allows them to delay the separation towards the trailing edge (Bhattacharjee and Subbarao [36]). Now, with this presumption and by analyzing the results shown in figure 3.6, we can conclude that there exists a narrow Strouhal number range where maximum lift is generated as well as the net thrust is approximately zero. Moreover, the lift-to-drag ratio maximizes for approximately the same range of Strouhal numbers (figure 3.6), demonstrating a local optimization in the force generation for those Strouhal numbers. These are very suitable for the requirements of cruising.*

The abovementioned remarks hold true for Strouhal numbers approximately between 0.1 and 0.3, depending upon the amplitude of twisting profile for a given mean twist angle, as shown in the parameter maps in figure 3.7. These parameter maps should be read along the red line. It seems likely that this optimality in force generation plays a role in the choice of the unique Strouhal number range for avian cruising flight. Thus, based on the results in this appendix, we argue that avian-scale cruising flight is restricted to the unique Strouhal number range approximately between 0.1 and 0.3 so that the avian creatures are able to benefit from the optimization in lift and thrust generation, provided the flow separations on the upper surfaces of the wings are delayed towards the trailing edge. Note, it is possible to draw another $\bar{C}_T = 0$ contour in the parameter maps in figure 3.7, apart from the one already shown. However, the cycle-averaged lift is negative at the corresponding values of St and θ_0. Thus, these are not considered in our analysis.

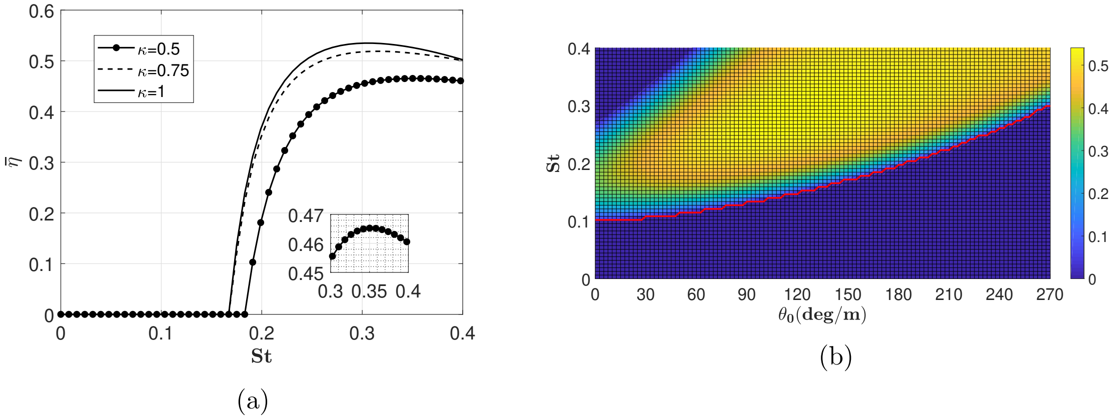


(a) Cycle-averaged lift coefficient

(b) Cycle-averaged net thrust coefficient

(c) Lift-to-drag ratio

Figure 3.7: Results for the avian wing planform: Parameter maps of cycle-averaged force coefficients and lift-to-drag ratio in terms of St and θ_0 for $\bar{\theta} = 10\,\text{deg}$ and $\kappa = 1$. Note that negative entries are shown as zero and the red line denotes $\bar{C}_T = 0$.



(a)

(b)

Figure 3.8: Results for the avian wing planform: (a) Plots of propulsive efficiency as functions of St & κ for $\theta_0 = 150$ deg/m, (b) Parameter map of propulsive efficiency in terms of St & θ_0 for $\kappa = 1$ and $\bar{\theta} = 10\,\text{deg}$ where the red line denotes $\bar{C}_T = 0$.

3.4 Discussion

In this section results for the rectangular planform will be compared with the equivalent results for the avian wing planform. The propulsive efficiency plots for the avian wing planform, equivalent to the ones depicted in figures 3.4 and 3.5, are shown in figure 3.8. It should be noted that the morphological and kinematic parameter values utilized to generate these results are the ones shown in table 3.3. It is obvious that the avian wing planform offers very minimal improvement over the rectangular planform in terms of the propulsive efficiency. However, it is again shown that the propulsive efficiency maximizes for Strouhal numbers approximately between 0.2 and 0.4 (figure 3.8).

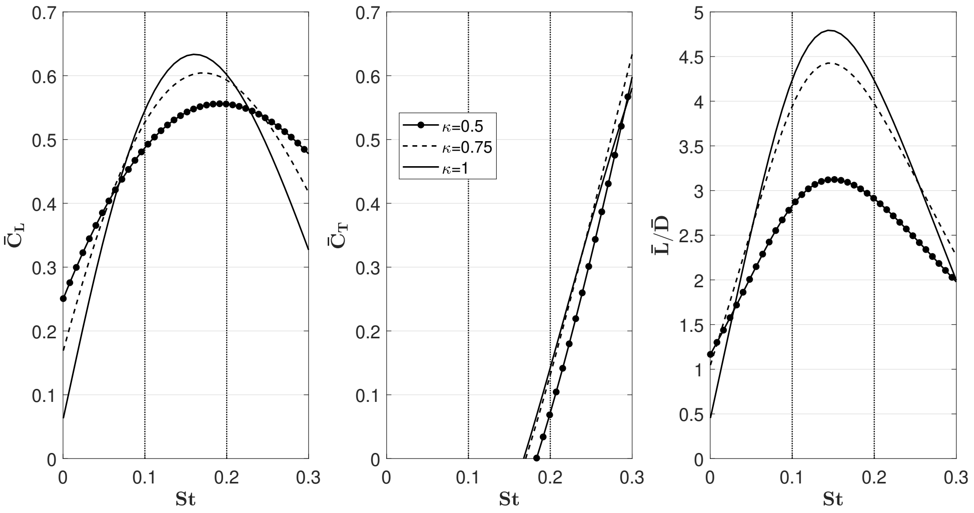
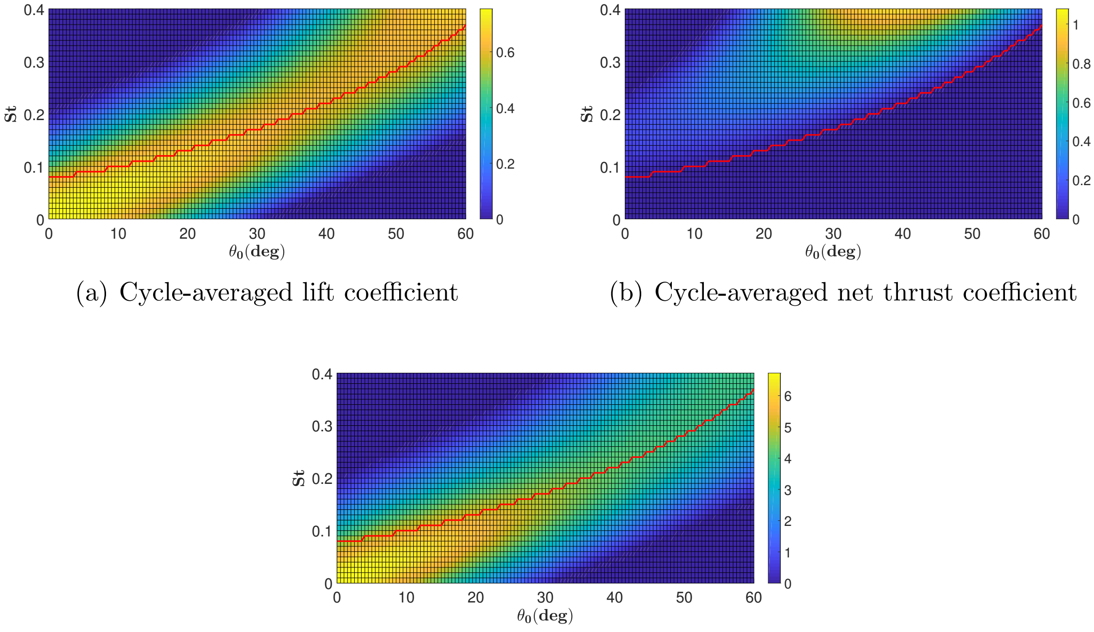


Figure 3.9: Results for the rectangular planform: Cycle-averaged force coefficients and lift-to-drag ratio as functions of St and κ. Note that the negative entries are shown as zero.

Next, the cycle-averaged force coefficients and lift-to-drag ratio for the rectangular planform are shown in figure 3.9. Again, it should be noted that values of the parameters for these results are shown in table 3.2. It is very interesting to observe that the Strouhal numbers for zero cycle-averaged net thrust are approximately the same for both the planforms. Also, the zero crossings in the net thrust are closely accompanied by the maximizations in cycle-averaged lift and lift-to-drag ratio for the rectangular planform as well (for κ values close to 1, as shown in figure 3.9). One of the reasons behind these qualitative similarities between the results of the two planforms is the amplitudes of the twist profiles: magnitude of the average amplitude of the linear twist profile for the avian planform is fairly close to the magnitude of the constant amplitude of twist for the rectangular planform (tables 3.2 and 3.3). Also, the magnitudes of the mean twist angle are the same for both the planforms (tables 3.2 and 3.3).



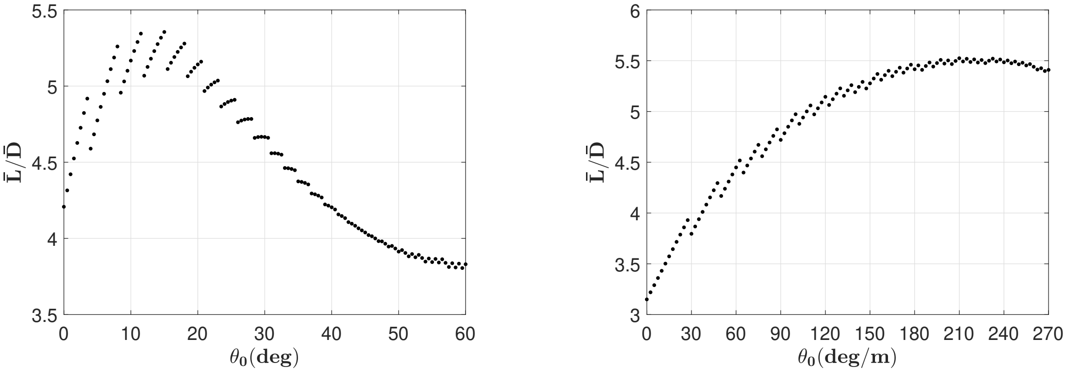
(a) Cycle-averaged lift coefficient

(b) Cycle-averaged net thrust coefficient

(c) Lift-to-drag ratio

Figure 3.10: Results for the rectangular wing planform: Parameter maps of cycle-averaged force coefficients and lift-to-drag ratio in terms of St and θ_0 for $\bar{\theta} = 10\,\mathrm{deg}$ and $\kappa = 1$. Note that negative entries are shown as zero and the red line denotes $\bar{C}_T = 0$.

The aforementioned qualitative agreements in force generation and lift-to-drag ratio results corresponding to two significantly different planforms have potential correlation with the pattern observed in the nature, i.e., despite having so many different morphological properties of the wings of the avian fliers, a large number of them choose to cruise in a narrow range of Strouhal numbers between 0.1 and 0.3 (Taylor et al. [28]). Also, it is shown in the propulsive efficiency results for both the planforms that the efficiency improves rapidly as the St increases from its corresponding zero net thrust value (figures 3.5 and 3.8(b)). That means positive net thrust can be generated optimally, if required intermittently during cruising. This observation further supplements the explanation presented here.

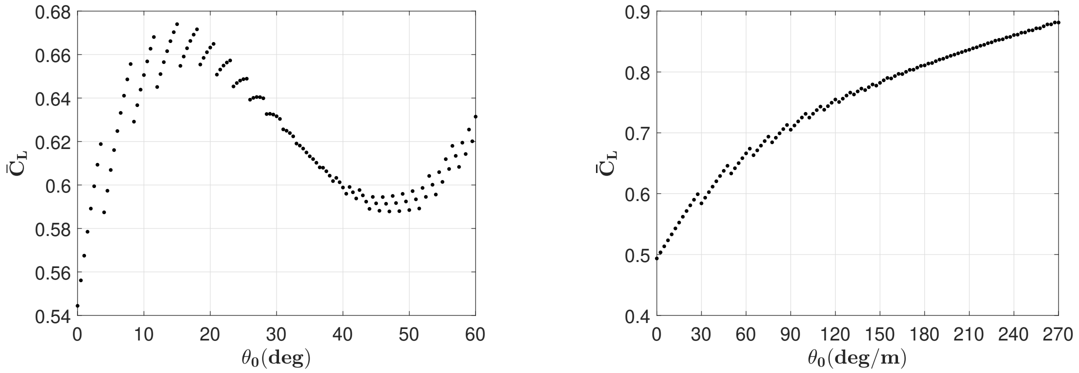


(a) Rectangular wing planform

(b) Avian wing planform

Figure 3.11: The lift-to-drag ratio results corresponding to $\bar{C}_T = 0$.

Finally, the parameter maps for the rectangular planform are shown in figure 3.10 which are qualitatively similar to the results shown in figure 3.7 for the avian wing planform. Again, these parameter maps in figure 3.10 should be read along the red line. Now, these red lines in figures 3.7 and 3.10 can be utilized to extract the lift-to-drag ratio values corresponding to $\bar{C}_T = 0$ in the grids and plotted against θ_0 for



(a) Rectangular wing planform

(b) Avian wing planform

Figure 3.12: The cycle-averaged lift coefficient results corresponding to $\bar{C}_T = 0$.

both the planforms. These results are shown in figure 3.11. Comparing the maxima in the two plots, we can deduce that the avian planform offers approximately 3% increase in the lift-to-drag ratio, which is not very significant. Similarly, extracting the cycle-averaged lift coefficients corresponding to $\bar{C}_T = 0$, we obtain the results shown in figure 3.12. With the θ_0 ranges considered, the maximum $\bar{C}_L$ for the avian wing planform is approximately 31% higher compared to the maximum obtained for the rectangular wing planform. This means that the avian wing planform, with its optimized semi-elliptic chord distribution and linear twist profile, offers approximately 31% improvement in terms of lift over the rectangular planform, having a constant chord distribution and a constant twist profile. Thus, the avian wing planform would be able to support more weight in cruising flight. An analysis similar to this can be performed for designing an efficient bio-inspired flapping vehicle.

3.5 Chapter Summary

Analytical expressions for cycle-averaged aerodynamic forces were derived for flapping wings having untapered, rectangular planforms with proper modeling of the local forces for the wings. *As shown in the results, these analytical expressions are valid for Strouhal numbers between 0 and 0.25, depending on κ.* Cycle-averaged power and propulsive efficiency have been defined and propulsive efficiency estimates are shown to be in agreement with the published results. A wing having the planform representative of avian creatures was incorporated in the analysis and corresponding results for cycle-averaged forces, propulsive efficiency, and lift-to-drag ratio were derived. Utilizing both sets of results for the rectangular and avian wing planforms, we have shown that there exists a narrow Strouhal number range where the cycle-averaged net thrust is approximately equal to zero, both cycle-averaged lift and lift-to-drag ratio maximize, given the chordwise flow separations on the upper surfaces of the wings are delayed towards the trailing edge. This narrow Strouhal number range was found to be varying between 0.1 and 0.3 which corresponds to the unique cruising range for a large number of avian creatures. Based on the results shown, it was argued that birds fly in the unique Strouhal number range to benefit from the local optimization in lift and thrust generation. Further, we showed that the avian wing planform can support significantly higher weights for cruising compared to the rectangular wing planform.

Chapter 4

Hypothesis: Birds Employ Online Optimization for Cruising Flight

The proposed hypothesis regarding likely in-flight mechanism used by birds to converge to cruising Strouhal numbers is provided in this chapter. Constructive arguments behind the hypothesis, as well as the hypothesis itself, are described in Section 4.1. The simulation setup, using one of the proposed extremum seeking schemes, is detailed afterwards in Section 4.2. Simulation results verifying the hypothesis are shown in Section 4.3, and the findings of this chapter are summarized in Section 4.4.

4.1 Background and Description of the Hypothesis

Taylor et al. [28] demonstrated that several flying (and swimming) creatures operate at Strouhal numbers between 0.2 and 0.4 while cruising (see figure 2 in [28]). It is quite remarkable that several species with different wing geometries, wing-body morphologies, and flight speeds have converged to this narrow range, indicating a universal pattern that might be true for any creature using flapping as the means for cruising [28]. We are interested in the cruising flight of birds, and direct cruising flights of birds are largely restricted between Strouhal numbers of approximately 0.1 and 0.3 [28], a range that permits high propulsive efficiency which natural selection might favor [28]. However, it seems logical that cruise performance should also be a contributing factor for the above choice of Strouhal numbers. It therefore follows that this choice has a correlation with optimal cruise performance as well. Taking this route, one can approach optimality of flapping flight in nature, which hitherto has been primarily investigated from fluid dynamics viewpoints, from a flight mechanics standpoint. Such a study has been carried out recently for bird-scale flapping in [129], and it has been shown that the Strouhal number range between 0.1 and 0.3 corresponds to maximum cycle-averaged lift, small cycle-averaged net-thrust, and high cycle-averaged lift-to-drag ratio [129], all of which are suitable for cruising flight, provided the flow separation on the upper surfaces of the wings remain close to the trailing edge [129].

The focus of this chapter is on the in-flight mechanism that birds use to optimize their flapping kinematics such that the operating Strouhal numbers converge to the unique range. We hypothesize that the mechanism is some variant of an online (or real-time) information feedback-based (or data-driven) optimization process that aims at maximizing the cruise performance objective with flapping kinematics as the variable. This hypothesis is based on the underlying conjecture that the use of optimal flapping kinematics for cruising is a learned behavior, and it stands to reason that birds are able to sense their environment and eventually learn from it to converge to a unique range of flapping motions (Strouhal numbers) every time. Following along this argument, one can assume that the creatures are able to perceive their performance in real-time and associate this perception with a value of the objective based on the learning, similar to the methodology implemented for modern artificially intelligent robots. The value of the objective is then passed along to the optimization process, and the recursion continues until sufficient performance is obtained, i.e., the operating flapping kinematics are close enough to the optima.

4.2 Simulation Setup

We intend to carry out a simulation study investigating the above hypothesis, and we adopt the model of bird-scale flapping flight (along with the relevant symbols) provided in Chapter 3 for that purpose. Let us first recall a few things from Chapter

3 for the convenience of our discussion here. As in Chapter 3, here we consider symmetric flapping motions of the wings, comprising of plunging and twisting. These motions are mathematically described as

$$\begin{aligned} \delta &= -\delta_0 \cos \omega_f t \\ \theta &= \theta(y) = \bar{\theta} - y\theta_0 \sin \omega_f t \end{aligned} \tag{4.1}$$

where $\omega_f = 2\pi f$ with f as the flapping frequency, y is the spanwise distance from the wing root, δ_0 is the plunging amplitude, θ_0 is the twisting amplitude, and $\bar{\theta}$ is the mean twist angle. The linear twist profile as well as the phase difference between plunging and twisting motions are correlated with optimal performance [19,129,149]. Now, let us recall the definition of Strouhal number for a flapping wing from Chapter 3, given by

$$St = \frac{b\delta_0 f}{V_\infty} \tag{4.2}$$

where b is the span of one wing and V_∞ is the freestream speed. We consider cycle-averaged lift-to-drag ratio as the objective for cruise performance. Cycle-averaged forces (lift and drag) for the wings are numerically computed using the results in Chapter 3. To this end, note that the simplifying assumptions introduced in Chapter 3 for analytical results therein are not relevant to the present study, and we do not include those simplifying assumptions here. With that, let $\bar{F}_{L_w}$ and $\bar{F}_{D_w}$ denote the cycle-averaged lift and drag for each wing, respectively. We assume that the body contributes a constant drag throughout the flapping cycle. Therefore, the net cycle-averaged drag for the wings and body is given by

$$\bar{F}_D = 2\bar{F}_{D_w} + F_{D_b} = 2\bar{F}_{D_w} + \frac{1}{2}\rho V_\infty^2 C_{d_b} S_w \tag{4.3}$$

where ρ stands for the atmospheric density of air (1.225 kg/m^3), C_{d_b} is the drag coefficient of the body, and S_w is the wing planform area (both wings). In this regard, we assume that the drag forces acting on the tail are negligible. Then, the cycle-averaged lift-to-drag ratio is calculated as follows:

$$\left(\bar{L}/\bar{D}\right) = \frac{2\bar{F}_{L_w}}{\bar{F}_D} = \frac{2\bar{F}_{L_w}}{\left(2\bar{F}_{D_w} + \frac{1}{2}\rho V_\infty^2 C_{d_b} S_w\right)} \tag{4.4}$$

where it is assumed that the lift forces generated by the tail and body are negligible compared to the wings. Alternatively, the cycle-averaged lift-to-drag ratio can be expressed in a more generic and compact form as

$$\left(\bar{L}/\bar{D}\right) = \pi(\boldsymbol{p}_{fs}, \boldsymbol{p}_f, \boldsymbol{p}_w, \boldsymbol{p}_a) \tag{4.5}$$

where $\boldsymbol{p}_{fs}$, $\boldsymbol{p}_f$, $\boldsymbol{p}_w$, and $\boldsymbol{p}_a$ denote vectors of freestream parameters (including V_∞), flapping parameters (including δ_0 and f), wing geometry parameters (including b), and aerodynamic parameters, respectively. However, note that an analytical description of the function $\pi(\cdot)$ does not exist in general. This means that despite the obvious correlation between cycle-averaged lift-to-drag ratio and Strouhal number (see, for example, the illustrations in Chapter 3), the mapping between the two quantities is not known analytically. It is worth pointing out that although we numerically compute cycle-averaged lift-to-drag ratio as described above, the simulation study is suitable for any method used to generate this quantity.

Next, we make a simplifying assumption that the framework emloyed by birds for the optimization can access current values of the cycle-averaged lift-to-drag ratio. Then, under this assumption, we regard a perturbation-based extremum seeking control [5, 6, 24, 128] scheme as that optimization framework. As mentioned previously, this technique is especially useful in scenarios where the mathematical form of the objective function is partially or completely unknown. This is the case for cycle-averaged lift-to-drag ratio in flapping flight, as explained above. Specifically, we choose ESC scheme-2 from Chapter 2 for the present study, and take flapping frequency (f) as the variable of optimization, keeping all other Strouhal number-related parameters (wingspan, amplitude of flapping, and freestream speed) constant (see figure 4.1).

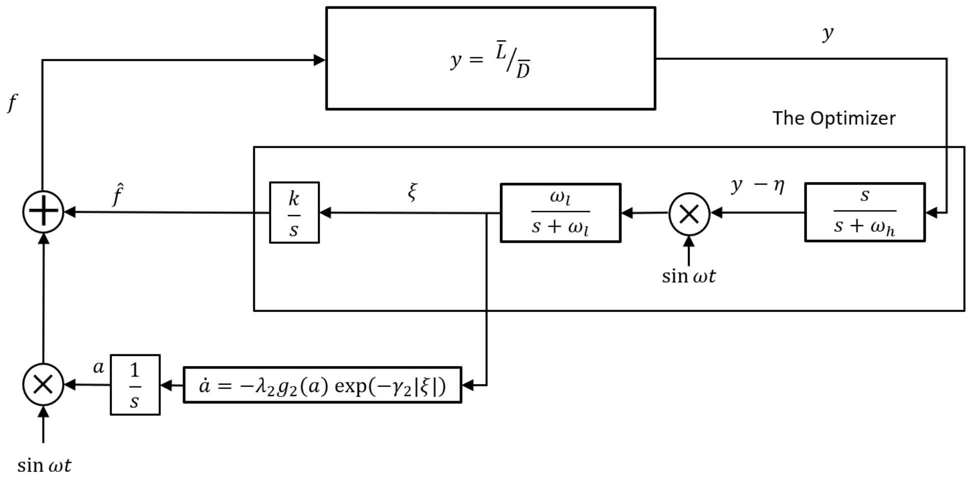


Figure 4.1: Schematic of the ESC scheme-2 for the present simulation setup where $g_2(a) = a$.

We consider the flight of pigeons (*Columbidae*) and gulls (*Laridae*) in the simulation study. Both of these bird species were included in the analysis by Taylor et al. [28] and were shown to cruise between Strouhal numbers of approximately 0.1 and 0.3 [28]. Overall geometry of a wing is captured through the chord distribution of its respective planform along the wingspan. Chord distribution for the pigeon wing planform is adopted from [157] and is as shown in equation (3.24). For convenience, it is included here as well:

$$c(y) = \begin{cases} c_0, & \forall y \in \left(0, \dfrac{b}{2}\right) \\ 4c_0 \dfrac{y}{b}(1 - \dfrac{y}{b}), & \forall y \in \left(\dfrac{b}{2}, b\right) \end{cases} \tag{4.6}$$

where c_0 is the root chord. It has been remarked that this is a good enough representation of any avian wing planform [40, 156, 157]. On the other hand, an empirical formula, which provides improvements over the chord distribution in equation (4.6), is selected for the chord distribution of a gull wing planform [160]. Morphometric characteristics of the pigeon and gull wing planforms are chosen based on the experimental data for common pigeons (*Columba livia* [158,161]) and common gulls (*Larus canus* [162]), respectively (see Table 4.1). It is assumed, without loss of generality, that the freestream speed is the same as the flight speed, and we take flight speeds equal to 12 m/s and 11.6 m/s for the pigeon and gull wing planforms, respectively, again based on experimental data [158,162]. The flapping kinematics-related parameters are listed in Table 4.2. Although θ_0 is set at 165 deg /m for the pigeon wing planform, the maximum twist (occurs at the wing-tips) is approximately equal to 46 deg and the average twist angle is approximately 23 deg.

Table 4.1: Morphological parameters for the wing planforms where entries adopted from experimental data in the literature are referenced. All the other entries are calculated based on the chord distribution selected for the respective wing planforms.

Item	Pigeon wing planform	Gull wing planform
b	0.297 m [158]	0.55 m [162]
c_0	0.13 m [161]	0.2134 m
S_w	0.0605 m^2	0.194 m^2
AR	5.5 [158]	6.22

Table 4.2: Flapping kinematics-related parameters for the chosen wing planforms.

Item	Pigeon wing planform	Gull wing planform
δ_0	60 deg	50 deg
θ_0	165 deg/m	67.5 deg/m
$\bar{\theta}$	9.5 deg	7.5 deg
C_{d_b}	0.015	0.015

The numerically computed cycle-averaged lift-to-drag ratio results are shown in figure 4.2. The gull wing planform yields higher values, likely due to its higher aspect ratio (more streamlined and less induced drag) compared to the pigeon wing planform (see Table 4.1).

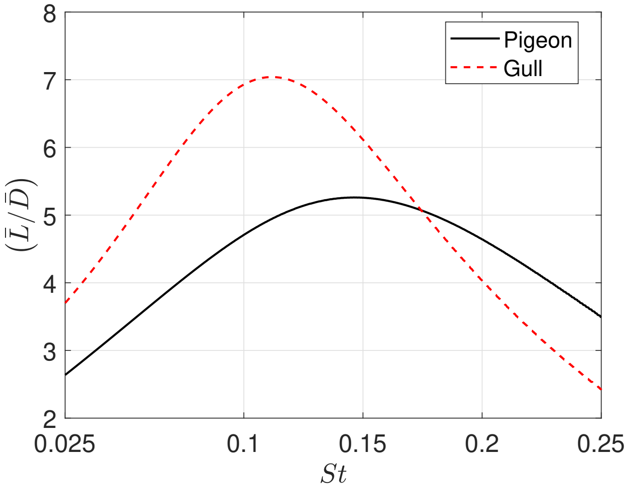
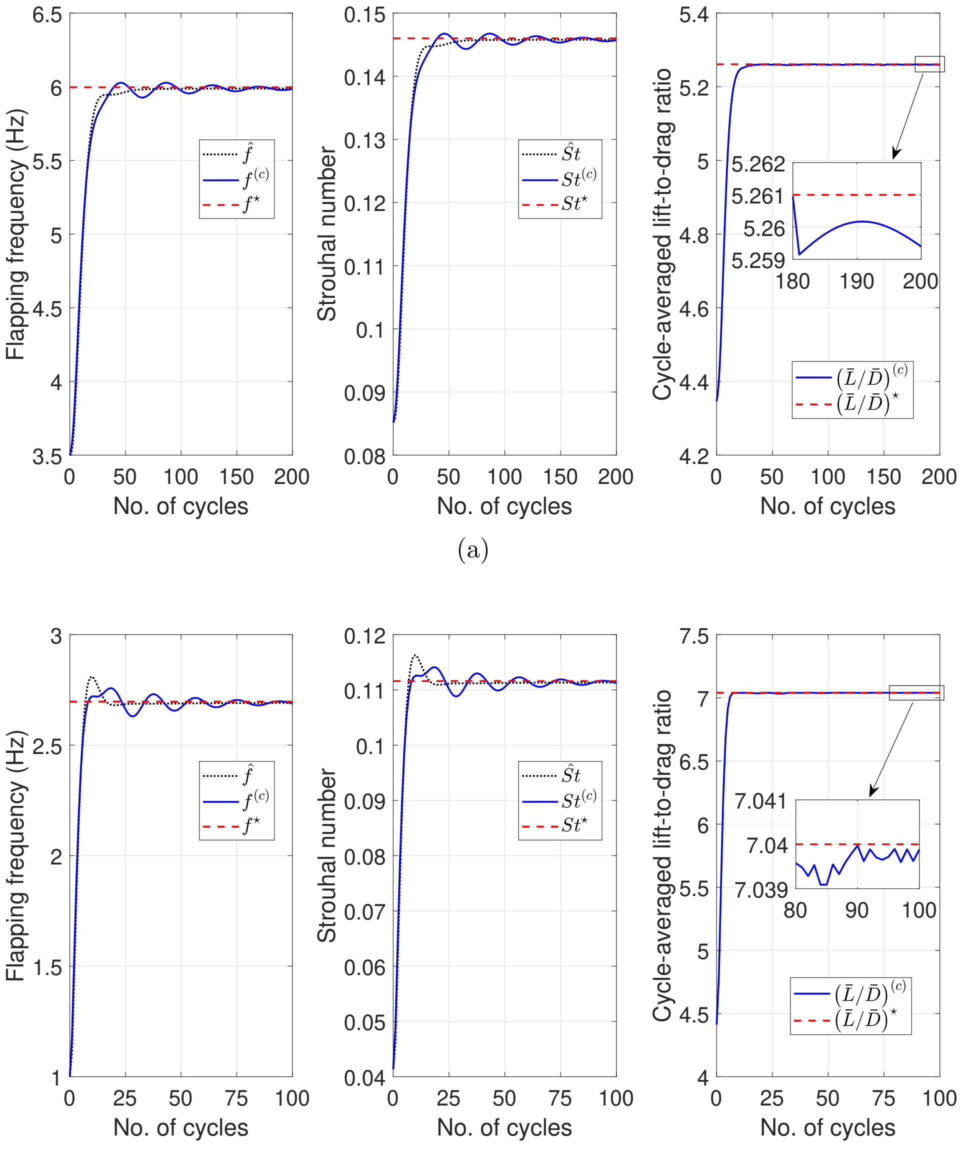


Figure 4.2: Cycle-averaged lift-to-drag ratio plots for the pigeon and gull wing planforms chosen.

4.3 Simulation Results

The simulation results are shown in figure 4.3, and the corresponding ESC scheme parameters and initial conditions are listed in Table 4.3 where a_0 stands for the initial value of the variable a in figure 4.1 (the perturbation signal amplitude) and $\hat{f}_0$ denotes the initial estimate of the optimal flapping frequency. It is noteworthy that we only needed to adjust the integration gain k across the two planforms, with all other ESC parameters remaining the same. These results in figure 4.3 clearly demonstrate that the operating (and estimated optimal) flapping frequencies converge to a small neighborhood of the respective true optimal values for both the pigeon and gull wing planforms. As a result, the cycle-averaged lift-to-drag ratio gets maximized and the operating (and estimated optimal) Strouhal numbers converge to (a small



(a)

(b)

Figure 4.3: The flapping frequency, Strouhal number and cycle-averaged lift-to-drag ratio plots for the **(a)** pigeon and **(b)** gull wing planforms. Note that the hatted quantities $\hat{(\cdot)}$ denote current estimates of the optimal values as determined by the ESC scheme. Also, superscripts (c) and $\star$ stand for the current operating values and the true optima, respectively. Following are the true optimum values: **(a)** $St^{\star} = 0.146$, $f^{\star} = 5.996$ Hz, $\left(\bar{L}/\bar{D}\right)^{\star} = 5.261$ and **(b)** $St^{\star} = 0.1116$, $f^{\star} = 2.697$ Hz, $\left(\bar{L}/\bar{D}\right)^{\star} = 7.04$.

neighborhood of) their respective true optima. The oscillating nature of the operating values is due to the perturbation introduced by the ESC scheme to extract the gradient information of the objective (cycle-averaged lift-to-drag ratio). The oscillations are clearly attenuated as the optimum is reached, and this is a feature of our ESC schemes (see Chapter 2). The pigeon and gull wing planforms first converge close to the respective true optima approximately after 21 and 6 flapping cycles, respectively (figure 4.3). However, in nature, birds might be capable of expediting this further. These results nonetheless corroborate our hypothesis and show that it is possible to implement a data-driven technique like ESC to converge to the optimal operating point in-flight. Going forward, this approach might help design efficient biomimetic flapping vehicles.

Table 4.3: Parameters and initial conditions of the ESC scheme utilized for the simulations.

Item	Pigeon wing planform	Gull wing planform
ω	0.9 rad/s	0.9 rad/s
k	0.761	0.513
ω_l	0.54 rad/s	0.54 rad/s
ω_h	0.69 rad/s	0.69 rad/s
λ_2	0.08	0.08
γ_2	5	5
$\hat{f}_0$	3.5 Hz	1 Hz
a_0	0.1	0.1

Furthermore, the estimated optimal flapping frequencies are approximately equal to 6 Hz and 2.7 Hz for the pigeon and gull wing planforms, respectively. Note that these values are within 10% of the flapping frequencies reported in the experimental studies for pigeons (*Columba livia* [158], *Columba palumbus* [162]) and gulls

(*Larus canus* [162]) flying at speeds similar to those used in the simulations. This serves as a verification of our simulation results here.

4.4 Chapter Summary

We considered the direct cruising flight of birds restricted to the unique range of Strouhal numbers between 0.1 and 0.3. Taking a flight mechanics approach, it is postulated that birds use some form of an information feedback-based real-time optimization framework to get to these Strouhal numbers in-flight and maximize their cruise performance in the process. A study simulating the flights of two different species of birds is carried out with an extremum seeking control scheme as the optimization framework and flapping frequency as the optimization variable. The simulation results show successful convergence to the respective true optima, and the corresponding flapping frequencies obtained are consistent with the experimental data for these birds.

Chapter 5

Set-Membership Filter for Discrete-Time Nonlinear Systems Using State Dependent Coefficient Parameterization*

A recursive set-membership filter (SMF) utilizing the state dependent coefficient (SDC) parameterization (termed SDC-SMF) is derived in this chapter for discrete-time nonlinear systems subject to unknown but bounded process and measurement noises. Note that an abbreviated version of the materials in this chapter has been published in reference [130]. In this chapter, the symbol $||\cdot||$ denotes the spectral norm for matrices and the Euclidean norm for vectors.

The chapter is organized as follows. Section 5.1 describes the preliminaries and problem formulation for the SDC-SMF. Section 5.2 discusses the main results for the proposed SDC-SMF and formulates the semi-definite programs (SDPs) to be solved at each time step to find the ellipsoidal sets containing the true state of the system. The theoretical observer properties for the SDC-SMF are assessed in Section 5.3. Finally, Section 5.4 includes a simulation example and Section 5.5 presents a summary of the contents in the chapter.

*Copyright © 2021 IEEE. Reprinted, with permission, from D. Bhattacharjee and K. Subbarao, "Set-Membership Filter for Discrete-Time Nonlinear Systems Using State Dependent Coefficient Parameterization," *IEEE Transactions on Automatic Control*, Early access, May 2021. DOI: 10.1109/TAC.2021.3082504 (reference [130])

5.1 Preliminaries and Problem Formulation

Consider discrete-time, nonlinear dynamical systems of the form

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{f}_d(\boldsymbol{x}_k) + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{h}_d(\boldsymbol{x}_k) + \boldsymbol{v}_k \end{aligned} \tag{5.1}$$

where $k \in \mathbb{Z}_\star$, $\boldsymbol{x}_k \in \mathbb{R}^n$ is the state of the system, $\boldsymbol{w}_k \in \mathbb{R}^n$ is the process noise or input disturbance, $\boldsymbol{y}_k \in \mathbb{R}^p$ is the measured output, and $\boldsymbol{v}_k \in \mathbb{R}^p$ is the measurement noise. We make the following standing assumption on the nonlinear functions $\boldsymbol{f}_d : \mathbb{R}^n \to \mathbb{R}^n$ and $\boldsymbol{h}_d : \mathbb{R}^n \to \mathbb{R}^p$.

Assumption 5.1.1. *$\boldsymbol{f}_d(\boldsymbol{0}_n) = \boldsymbol{0}_n$, $\boldsymbol{h}_d(\boldsymbol{0}_n) = \boldsymbol{0}_p$, and $\boldsymbol{f}_d(\cdot) \in C^t$, $\boldsymbol{h}_d(\cdot) \in C^t$ where $t \geq 2$.*

Under Assumption 5.1.1, the nonlinear functions can be put into corresponding pseudo-linear forms using the SDC parameterization as

$$\begin{aligned} \boldsymbol{f}_d(\boldsymbol{x}_k) &= \boldsymbol{A}(\boldsymbol{x}_k)\boldsymbol{x}_k \\ \boldsymbol{h}_d(\boldsymbol{x}_k) &= \boldsymbol{H}(\boldsymbol{x}_k)\boldsymbol{x}_k \end{aligned} \tag{5.2}$$

where $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and $\boldsymbol{H} : \mathbb{R}^n \to \mathbb{R}^{p\times n}$ are nonlinear matrix-valued functions. To this end, we recall the following useful result.

Proposition 5.1.1 ([72, 163])**.** *Under Assumption 5.1.1, SDC parameterizations of $\boldsymbol{f}_d(\boldsymbol{x}_k)$, $\boldsymbol{h}_d(\boldsymbol{x}_k)$ as in (5.2) always exist for some C^{t-1} matrix-valued functions $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and $\boldsymbol{H} : \mathbb{R}^n \to \mathbb{R}^{p\times n}$. This property is satisfied by the following parameterizations*

$$\begin{aligned} \boldsymbol{A}(\boldsymbol{x}_k) &= \int_0^1 \left.\frac{\partial \boldsymbol{f}_d(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\right|_{\boldsymbol{x}_k=\lambda\boldsymbol{x}_k} d\lambda \\ \boldsymbol{H}(\boldsymbol{x}_k) &= \int_0^1 \left.\frac{\partial \boldsymbol{h}_d(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\right|_{\boldsymbol{x}_k=\lambda\boldsymbol{x}_k} d\lambda \end{aligned} \tag{5.3}$$

where λ is a dummy variable of integration. The parameterizations in (5.3) *are guaranteed to exist under Assumption 5.1.1. Furthermore, any SDC parameterization of* $\boldsymbol{f}_d(\boldsymbol{x}_k)$, $\boldsymbol{h}_d(\boldsymbol{x}_k)$ *as in* (5.2) *satisfies* $\boldsymbol{A}(\boldsymbol{0}_n) = \frac{\partial \boldsymbol{f}_d(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\big|_{\boldsymbol{x}_k=\boldsymbol{0}_n}$, $\boldsymbol{H}(\boldsymbol{0}_n) = \frac{\partial \boldsymbol{h}_d(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}\big|_{\boldsymbol{x}_k=\boldsymbol{0}_n}$.

Note that multiple SDC parameterizations of the form (5.2) are possible for $n > 1$ using mathematical factorization [72]. However, we choose the SDC parameterizations given in (5.3) under Assumption 5.1.1 and describe the nonlinear system (5.1) in an equivalent pseudo-linear form as

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{H}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{v}_k. \end{aligned} \tag{5.4}$$

For a detailed discussion on the SDC parameterization, refer to [71,72] and references therein. We make the following assumption on the state dynamics of system (5.4).

Assumption 5.1.2. *[57, Section V] There exist compact sets* $\mathbb{D}_0, \mathbb{D} \subset \mathbb{R}^n$ *and* $\epsilon_1 > 0$ *such that* $\boldsymbol{x}_0 \in \mathbb{D}_0$ *implies*

$$\boldsymbol{x}_k + \epsilon_1 \mathcal{B}(\boldsymbol{x}_k) \subset \mathbb{D}, \quad \forall k \in \mathbb{Z}_\star$$

where $\mathcal{B}(\boldsymbol{x}_k)$ *is the closed unit ball in* $\mathbb{R}^n$ *centered at* $\boldsymbol{x}_k$.

The above assumption implies that the state $\boldsymbol{x}_k$ evolves within a compact set $\mathbb{D}$ which is not necessarily small [57]. Now, we state the following assumptions for system (5.4) where $\mathbb{D}_0$ is as described in Assumption 5.1.2.

Assumption 5.1.3. $\boldsymbol{x}_0$ *is unknown but belongs to a known ellipsoid, i.e.,* $\boldsymbol{x}_0 \in \mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0) \subseteq \mathbb{D}_0$ *where* $\hat{\boldsymbol{x}}_0$ *is a given initial estimate and* $\boldsymbol{P}_0$ *is known.*

Assumption 5.1.4. $\boldsymbol{w}_k$ *and* $\boldsymbol{v}_k$ *are unknown but bounded and belong to known ellipsoids, i.e.,* $\boldsymbol{w}_k \in \mathcal{E}(\boldsymbol{0}_n, \boldsymbol{Q}_k)$ *and* $\boldsymbol{v}_k \in \mathcal{E}(\boldsymbol{0}_p, \boldsymbol{R}_k)$, $\forall k \in \mathbb{Z}_\star$ *where* $\boldsymbol{Q}_k$, $\boldsymbol{R}_k$ *are known and satisfy* $||\boldsymbol{Q}_k|| \leq q$ *and* $||\boldsymbol{R}_k|| \leq r$, $\forall k \in \mathbb{Z}_\star$ *with some* $q, r > 0$.

Assumption 5.1.4 means that the process and measurement noises in system (5.4) are uniformly upper bounded. Now, we introduce the final assumption on system (5.4).

Assumption 5.1.5. *Along any trajectory of system* (5.4) *under Assumption 5.1.2, define*

$$\begin{aligned} \boldsymbol{\phi}_{k+s,k} &= \boldsymbol{A}(\boldsymbol{x}_{k+s-1})\ \boldsymbol{A}(\boldsymbol{x}_{k+s-2})\cdots\boldsymbol{A}(\boldsymbol{x}_k) \\ \mathcal{O}_{k,k+s} &= \begin{bmatrix} \boldsymbol{H}(\boldsymbol{x}_k) \\ \boldsymbol{H}(\boldsymbol{x}_{k+1})\boldsymbol{\phi}_{k+1,k} \\ \vdots \\ \boldsymbol{H}(\boldsymbol{x}_{k+s})\boldsymbol{\phi}_{k+s,k} \end{bmatrix} \end{aligned} \tag{5.5}$$

for any $s \in \mathbb{Z}_\star \backslash \{0\}$*. There exists an* $N_o \in \mathbb{Z}_\star \backslash \{0\}$ *such that*

$$\mathit{rank}(\mathcal{O}_{k,k+N_o-1}) = n, \quad \forall k \in \mathbb{Z}_\star. \tag{5.6}$$

This is an observability assumption where $N_o = n$ might be possible. The above assumption leads to the following result.

Proposition 5.1.2. *Under Assumption 5.1.5, there exist* $\mu_1, \mu_2 > 0$ *such that*

$$\mu_1 \boldsymbol{I}_n \le \mathcal{O}^T_{k,k+N_o-1}\mathcal{O}_{k,k+N_o-1} \le \mu_2 \boldsymbol{I}_n. \tag{5.7}$$

Proof. Follows directly from Proposition 5.1 in [57] (or see the proof of Proposition 4.1 in [164, Section 4]). □

Remark 5.1.3. *With the knowledge of set* $\mathbb{D}$*, Assumption 5.1.5 requires one to check if the rank condition in* (5.6) *is satisfied for all* $\boldsymbol{x}_k \in \mathbb{D}$ *with some* $N_o \in \mathbb{Z}_\star \backslash \{0\}$*. This can be done by carrying out a theoretical analysis (cf., Section 5.4) or by implementing a numerical routine.*

Remark 5.1.4. *With our compactness and observability assumptions, theoretical properties of the proposed SDC-SMF for system* (5.4) *(in a sense similar to that* *of Definition 3.1 in [57]) can be assessed by appropriately modifying the analysis and results in [57] (cf., Section 5.3).*

5.1.1 SDC-SMF Objectives

The objective is to develop an SDC-SMF for system (5.4) having a correction-prediction form, similar to the Kalman Filter variants [45]. This helps to obtain an accurate estimate of the state and a reliable evaluation of the estimation error. The filtering objectives are as follows.

5.1.1.1 Correction Step

At each time step $k \in \mathbb{Z}_\star$, upon receiving the measurement $\boldsymbol{y}_k$ with $\boldsymbol{v}_k \in \mathcal{E}(\mathbf{0}_p, \boldsymbol{R}_k)$ and given $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, the objective is to find a *correction ellipsoid* such that $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$. The corrected state estimate is given by

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{L}_k(\boldsymbol{y}_k - \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\hat{\boldsymbol{x}}_{k|k-1}) \tag{5.8}$$

where $\boldsymbol{L}_k$ is the filter gain.

5.1.1.2 Prediction Step

At each time step $k \in \mathbb{Z}_\star$, given $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\boldsymbol{w}_k \in \mathcal{E}(\mathbf{0}_n, \boldsymbol{Q}_k)$, the objective is to find a *prediction ellipsoid* such that $\boldsymbol{x}_{k+1} \in \mathcal{E}(\hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{P}_{k+1|k})$ where the predicted state estimate is given by

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k}. \tag{5.9}$$

Initialization is provided by $\hat{\boldsymbol{x}}_{0|-1} = \hat{\boldsymbol{x}}_0$ and $\boldsymbol{P}_{0|-1} = \boldsymbol{P}_0$ [45] which form the *initial prediction ellipsoid* due to Assumption 5.1.3.

5.1.2 Matrix Taylor Expansions of the SDC Matrices

Assume that the state of system (5.4) at time step k belongs to the prediction ellipsoid of time step $k-1$, i.e., $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$ where $\hat{\boldsymbol{x}}_{k|k-1}$ and $\boldsymbol{P}_{k|k-1}$ are known. Then, there exists a $\boldsymbol{z}_{k|k-1} \in \mathbb{R}^n$ with $||\boldsymbol{z}_{k|k-1}|| \leq 1$ such that

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \tag{5.10}$$

where $\boldsymbol{E}_{k|k-1}$ is the Cholesky factorization of $\boldsymbol{P}_{k|k-1}$, i.e., $\boldsymbol{P}_{k|k-1} = \boldsymbol{E}_{k|k-1}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}$ [52, 54]. Utilizing the matrix Taylor expansion in [165], $\boldsymbol{H}(\boldsymbol{x}_k) = \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1}+\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1})$ can be expanded about the state estimate $\hat{\boldsymbol{x}}_{k|k-1}$ as

$$\boldsymbol{H}(\boldsymbol{x}_k) = \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{K}_1(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{\Delta}_1(\boldsymbol{\xi}_{k|k-1}) + \boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k) \tag{5.11}$$

where $\boldsymbol{K}_1(\hat{\boldsymbol{x}}_{k|k-1}) = \mathcal{D}_{\boldsymbol{x}^{\mathrm{T}}}\boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})$ is the derivative matrix evaluated at $\hat{\boldsymbol{x}}_{k|k-1}$, $\boldsymbol{\Delta}_1(\boldsymbol{\xi}_{k|k-1}) = \left(\boldsymbol{\xi}_{k|k-1} \otimes \boldsymbol{I}_n\right)$ with $\boldsymbol{\xi}_{k|k-1} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1}$, and $\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)$ is the remainder (see Section 6 in [165]). Similarly, the matrix $\boldsymbol{A}(\boldsymbol{x}_k) = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k})$ is expanded as

$$\boldsymbol{A}(\boldsymbol{x}_k) = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k}) + \boldsymbol{K}_2(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{\Delta}_2(\boldsymbol{\xi}_{k|k}) + \boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k) \tag{5.12}$$

where $\boldsymbol{K}_2(\hat{\boldsymbol{x}}_{k|k}) = \mathcal{D}_{\boldsymbol{x}^{\mathrm{T}}}\boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})$, $\boldsymbol{\Delta}_2(\boldsymbol{\xi}_{k|k}) = \left(\boldsymbol{\xi}_{k|k} \otimes \boldsymbol{I}_n\right)$ with $\boldsymbol{\xi}_{k|k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k}$ where $\boldsymbol{P}_{k|k} = \boldsymbol{E}_{k|k}\boldsymbol{E}_{k|k}^{\mathrm{T}}$ and $||\boldsymbol{z}_{k|k}|| \leq 1$. As $\boldsymbol{A}(\boldsymbol{x}_k)$, $\boldsymbol{H}(\boldsymbol{x}_k)$ are calculated using (5.3) under Assumption 5.1.1, $\boldsymbol{K}_1(\cdot)$, $\boldsymbol{K}_2(\cdot)$ are at least continuous matrix-valued functions.

5.1.3 Upper Bounds on the Norms of Remainders in Matrix Taylor Expansions

At each time step, upper bounds on the norms of the remainders in (5.11)-(5.12) are calculated and utilized in the SDC-SMF design. Thus, we require the following quantities:

$$\begin{aligned} \bar{r}_{A_k} &= \sup_{\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})} ||\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)|| \\ \bar{r}_{H_k} &= \sup_{\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})} ||\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)||. \end{aligned} \tag{5.13}$$

Next, we state an important result regarding $\bar{r}_{A_k}$ and $\bar{r}_{H_k}$.

Proposition 5.1.5. *$\bar{r}_{A_k}$ and $\bar{r}_{H_k}$ belong to the boundaries of the ellipsoids $\mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, respectively.*

Proof. Let us denote $\mathcal{E}_{c_k} = \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\mathcal{E}_{p_k} = \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$. Due to the continuity of the matrix-valued functions and compactness of the ellipsoids, we have

$$\begin{aligned} h_k &= \sup_{\boldsymbol{x} \in \mathcal{E}_{p_k}} ||\boldsymbol{H}(\boldsymbol{x})||, \quad k_{1_k} = \sup_{\boldsymbol{x} \in \mathcal{E}_{p_k}} ||\boldsymbol{K}_1(\boldsymbol{x})||, \\ a_k &= \sup_{\boldsymbol{x} \in \mathcal{E}_{c_k}} ||\boldsymbol{A}(\boldsymbol{x})||, \quad k_{2_k} = \sup_{\boldsymbol{x} \in \mathcal{E}_{c_k}} ||\boldsymbol{K}_2(\boldsymbol{x})|| \end{aligned}$$

where $0 < a_k, h_k, k_{1_k}, k_{2_k} < \infty$. Now, consider the remainder $\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)$ in (5.12), expressed as

$$\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k) = \boldsymbol{A}(\boldsymbol{x}_k) - \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k}) - \boldsymbol{K}_2 \boldsymbol{\Delta}_2 \tag{5.14}$$

where the arguments of $\boldsymbol{K}_2(\cdot)$ and $\boldsymbol{\Delta}_2(\cdot)$ have been dropped. Taking the norm leads to

$$||\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)|| \leq 2a_k + k_{2_k} ||\boldsymbol{\Delta}_2||.$$

Utilizing $\boldsymbol{\Delta}_2 = \big(\boldsymbol{\xi}_{k|k} \otimes \boldsymbol{I}_n\big)$, the following holds:

$$||\boldsymbol{\Delta}_2|| = ||\boldsymbol{\xi}_{k|k}|| \; ||\boldsymbol{I}_n|| \leq ||\boldsymbol{E}_{k|k}|| \; ||\boldsymbol{z}_{k|k}||. \tag{5.15}$$

Denoting $||\boldsymbol{E}_{k|k}|| = \gamma_{k|k}$, (5.15) becomes $||\boldsymbol{\Delta}_2|| \leq \gamma_{k|k}||\boldsymbol{z}_{k|k}||$. Then, the norm of the remainder satisfies

$$||\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)|| \leq 2a_k + k_{2_k}\gamma_{k|k}||\boldsymbol{z}_{k|k}||. \tag{5.16}$$

Clearly, the upper bound corresponds to $||\boldsymbol{z}_{k|k}|| = 1$, i.e., $\bar{r}_{A_k}$ belongs to the boundary of $\mathcal{E}_{c_k}$. Carrying out a similar analysis for $\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)$ yields

$$||\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)|| \leq 2h_k + k_{1_k}\gamma_{k|k-1}||\boldsymbol{z}_{k|k-1}||$$

with $||\boldsymbol{E}_{k|k-1}|| = \gamma_{k|k-1}$ which shows that $\bar{r}_{H_k}$ belongs to the boundary of $\mathcal{E}_{p_k}$. This completes the proof. □

Therefore, $\bar{r}_{A_k}$ can be obtained by solving the optimization problem

$$\begin{aligned} &\sup_{\boldsymbol{z}_{k|k}} ||\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k})|| \\ &\text{subject to } ||\boldsymbol{z}_{k|k}|| = 1 \end{aligned} \tag{5.17}$$

where the feasible set is non-convex. The non-convex problem can be convexified and solved using the primal-dual methods numerically (see, e.g., [166]). Alternatively, a much simpler approach, so-called *non-adaptive random search algorithm* [167, 168], can be utilized to obtain an approximate solution to (5.17). Adopting this approach, the norm of the remainder is evaluated N times by randomly sampling N number of points on the unit circle $||\boldsymbol{z}_{k|k}|| = 1$. Then, the upper bound on the remainder norm is given by the *empirical maximum* [167] as

$$r_{A_k} = \max_{i=1,2,...,N} ||\boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k_i})|| \tag{5.18}$$

where $||\boldsymbol{z}_{k|k_i}|| = 1, \ i = 1, 2, ..., N$. Similarly, the upper bound on the norm of $\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)$ is determined as

$$r_{H_k} = \max_{i=1,2,...,N} ||\boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1_i})|| \tag{5.19}$$

where $||\boldsymbol{z}_{k|k-1_i}|| = 1,\ i = 1, 2, ..., N$. Moreover, as $N \to \infty$, we have $r_{A_k} \to \bar{r}_{A_k}$ and $r_{H_k} \to \bar{r}_{H_k}$ (see Theorem 7.4 in [167]). The arguments of $\boldsymbol{K}_i(\cdot)$ and $\boldsymbol{\Delta}_i(\cdot)$ $(i = 1, 2)$ have been dropped in the subsequent analysis.

Remark 5.1.6. *Using the matrix Taylor expansions in* (5.11)-(5.12)*, the governing equations utilized for the SDC-SMF design for system* (5.4) *can be expressed as*

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{x}_k + \tilde{\boldsymbol{w}}_k \\ \boldsymbol{y}_k &= \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{x}_k + \tilde{\boldsymbol{v}}_k \end{aligned} \tag{5.20}$$

where

$$\begin{aligned} \tilde{\boldsymbol{w}}_k &= \boldsymbol{w}_k + \boldsymbol{K}_2\boldsymbol{\Delta}_2\boldsymbol{x}_k + \boldsymbol{R}_{\boldsymbol{A}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)\boldsymbol{x}_k \\ \tilde{\boldsymbol{v}}_k &= \boldsymbol{v}_k + \boldsymbol{K}_1\boldsymbol{\Delta}_1\boldsymbol{x}_k + \boldsymbol{R}_{\boldsymbol{H}_2}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{x}_k)\boldsymbol{x}_k. \end{aligned}$$

The governing equations in (5.20) *are different from the governing equations utilized in EKF-like approach-based SMF frameworks (see, e.g., Section 3 in [69]). The bounds on the terms in* $\tilde{\boldsymbol{w}}_k$*,* $\tilde{\boldsymbol{v}}_k$ *and the ellipsoidal set description of the true state* $\boldsymbol{x}_k$ *are utilized in the next section to derive the SDC-SMF.*

5.2 Main Results

This section formulates the SDPs to be solved at each time step for the correction and prediction steps. The arguments of $\boldsymbol{R}_{\boldsymbol{A}_2}(\cdot)$ and $\boldsymbol{R}_{\boldsymbol{H}_2}(\cdot)$ are omitted in the subsequent analysis for notational simplicity. With that, let us state Theorem 5.2.1 that summarizes the filtering problem at the correction step.

Theorem 5.2.1. *Consider system* (5.4) *under Assumptions 5.1.3 and 5.1.4. At each time step* $k \in \mathbb{Z}_\star$*, upon receiving the measurement* $\boldsymbol{y}_k$ *with* $\boldsymbol{v}_k \in \mathcal{E}(\boldsymbol{0}_p, \boldsymbol{R}_k)$ *and given* $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$*, the state* $\boldsymbol{x}_k$ *is contained in the optimal correction ellipsoid*

$\mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$, if there exist $\boldsymbol{P}_{k|k} > 0$, $\boldsymbol{L}_k$, $\tau_i \geq 0$, $i = 1, 2, 3, 4, 5, 6$ as solutions to the following SDP:

$$\begin{aligned}
&\min_{\boldsymbol{P}_{k|k}, \boldsymbol{L}_k, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6} \quad trace(\boldsymbol{P}_{k|k}) \\
&\text{subject to} \\
&\boldsymbol{P}_{k|k} > 0 \\
&\tau_i \geq 0, \ i = 1, 2, 3, 4, 5, 6 \\
&\begin{bmatrix} -\boldsymbol{P}_{k|k} & \boldsymbol{\Pi}_{k|k-1} \\ & \\ \boldsymbol{\Pi}^{\mathrm{T}}_{k|k-1} & -\boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6) \end{bmatrix} \leq 0
\end{aligned} \tag{5.21}$$

where $\boldsymbol{\Pi}_{k|k-1}$ and $\boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)$ are given by

$$\begin{aligned}
&\boldsymbol{\Pi}_{k|k-1} \\
&= \Big[\boldsymbol{0}_n \quad (\boldsymbol{E}_{k|k-1} - \boldsymbol{L}_k \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{E}_{k|k-1}) \quad -\boldsymbol{L}_k \quad -\boldsymbol{L}_k\boldsymbol{K}_1 \quad -\boldsymbol{L}_k \quad -\boldsymbol{L}_k\boldsymbol{K}_1 \\
&\qquad -\boldsymbol{L}_k\Big] \\
&\boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6) \\
&= \mathrm{diag}\ (1 - \tau_1 - \tau_2 - \tau_5\gamma^2_{k|k-1}\hat{\boldsymbol{x}}^{\mathrm{T}}_{k|k-1}\hat{\boldsymbol{x}}_{k|k-1} - \tau_6 r^2_{H_k}\hat{\boldsymbol{x}}^{\mathrm{T}}_{k|k-1}\hat{\boldsymbol{x}}_{k|k-1}, \tau_1\boldsymbol{I}_n \\
&\quad - \tau_3\gamma^2_{k|k-1}\boldsymbol{E}^{\mathrm{T}}_{k|k-1}\boldsymbol{E}_{k|k-1} - \tau_4 r^2_{H_k}\boldsymbol{E}^{\mathrm{T}}_{k|k-1}\boldsymbol{E}_{k|k-1}, \tau_2\boldsymbol{R}^{-1}_k, \tau_3\boldsymbol{I}_{n^2}, \tau_4\boldsymbol{I}_p, \tau_5\boldsymbol{I}_{n^2}, \tau_6\boldsymbol{I}_p).
\end{aligned} \tag{5.22}$$

Furthermore, center of the correction ellipsoid is given by the corrected state estimate in (5.8).

Proof. See Appendix D. □

The next Theorem summarizes the filtering problem at the prediction step.

Theorem 5.2.2. *Consider system* (5.4) *under Assumption 5.1.4 with the current state $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\boldsymbol{w}_k \in \mathcal{E}(\boldsymbol{0}_n, \boldsymbol{Q}_k)$. Then, the successor state $\boldsymbol{x}_{k+1}$ belongs* *to the optimal prediction ellipsoid* $\mathcal{E}(\hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{P}_{k+1|k})$*, if there exist* $\boldsymbol{P}_{k+1|k} > 0$*,* $\tau_i \geq 0,\ i = 7, 8, 9, 10, 11, 12$ *as solutions to the following SDP:*

$$\begin{aligned}
&\min_{\boldsymbol{P}_{k+1|k}, \tau_7, \tau_8, \tau_9, \tau_{10}, \tau_{11}, \tau_{12}} \quad \textit{trace}(\boldsymbol{P}_{k+1|k}) \\
&\textit{subject to} \\
&\boldsymbol{P}_{k+1|k} > 0 \\
&\tau_i \geq 0, i = 7, 8, 9, 10, 11, 12 \\
&\begin{bmatrix} -\boldsymbol{P}_{k+1|k} & \boldsymbol{\Pi}_{k|k} \\ & \\ \boldsymbol{\Pi}_{k|k}^{\mathrm{T}} & -\boldsymbol{\Psi}(\tau_7, \tau_8, \tau_9, \tau_{10}, \tau_{11}, \tau_{12}) \end{bmatrix} \leq 0
\end{aligned} \tag{5.23}$$

where $\boldsymbol{\Pi}_{k|k}$ *and* $\boldsymbol{\Psi}(\tau_7, \tau_8, \tau_9, \tau_{10}, \tau_{11}, \tau_{12})$ *are given by*

$$\begin{aligned}
&\boldsymbol{\Pi}_{k|k} \\
&= \begin{bmatrix} \boldsymbol{0}_n & \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{E}_{k|k} & \boldsymbol{I}_n & \boldsymbol{K}_2 & \boldsymbol{I}_n & \boldsymbol{K}_2 & \boldsymbol{I}_n \end{bmatrix} \\
&\boldsymbol{\Psi}(\tau_7, \tau_8, \tau_9, \tau_{10}, \tau_{11}, \tau_{12}) \\
&= \mathrm{diag}\,(1 - \tau_7 - \tau_8 - \tau_9 \gamma_{k|k}^2 \hat{\boldsymbol{x}}_{k|k}^{\mathrm{T}} \hat{\boldsymbol{x}}_{k|k} - \tau_{10} r_{A_k}^2 \hat{\boldsymbol{x}}_{k|k}^{\mathrm{T}} \hat{\boldsymbol{x}}_{k|k}, \tau_7 \boldsymbol{I}_n - \tau_{11} \gamma_{k|k}^2 \boldsymbol{E}_{k|k}^{\mathrm{T}} \boldsymbol{E}_{k|k} \\
&\quad - \tau_{12} r_{A_k}^2 \boldsymbol{E}_{k|k}^{\mathrm{T}} \boldsymbol{E}_{k|k}, \tau_8 \boldsymbol{Q}_k^{-1}, \tau_9 \boldsymbol{I}_{n^2}, \tau_{10} \boldsymbol{I}_n, \tau_{11} \boldsymbol{I}_{n^2}, \tau_{12} \boldsymbol{I}_n).
\end{aligned}$$

Furthermore, center of the prediction ellipsoid is given by the predicted state estimate in (5.9).

Proof. Utilizing (5.4) and (5.9), we have

$$\begin{aligned}
&\boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1|k} \\
&= \boldsymbol{A}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{w}_k - \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k} \\
&= (\boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k}) + \boldsymbol{K}_2 \boldsymbol{\Delta}_2 + \boldsymbol{R}_{\boldsymbol{A}_2})(\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k}) + \boldsymbol{w}_k - \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k} \\
&= \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k} + \boldsymbol{K}_2 \boldsymbol{\Delta}_2 \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{R}_{\boldsymbol{A}_2} \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{K}_2 \boldsymbol{\Delta}_2 \boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k} + \boldsymbol{R}_{\boldsymbol{A}_2} \boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k} + \boldsymbol{w}_k
\end{aligned} \tag{5.24}$$

Denote the unknowns in (5.24) as

$$\begin{aligned}\boldsymbol{\Delta}_7 &= \boldsymbol{\Delta}_2 \boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k} \\ \boldsymbol{\Delta}_8 &= \boldsymbol{R}_{\boldsymbol{A}_2} \boldsymbol{E}_{k|k} \boldsymbol{z}_{k|k} \\ \boldsymbol{\Delta}_9 &= \boldsymbol{\Delta}_2 \hat{\boldsymbol{x}}_{k|k} \\ \boldsymbol{\Delta}_{10} &= \boldsymbol{R}_{\boldsymbol{A}_2} \hat{\boldsymbol{x}}_{k|k}.\end{aligned} \tag{5.25}$$

The rest of the proof can be completed by carrying out steps similar to the ones carried out for the proof of Theorem 5.2.1. □

These SDPs in (5.21) and (5.23) can be solved efficiently using interior point methods [169]. In terms of practical efficiency, interior point methods roughly require 5-50 iterations to solve each SDP with each iteration requiring solution to a least-squares problem of the same size as the original problem [169]. The recursive SDC-SMF algorithm for system (5.4) is summarized in Algorithm 1.

Algorithm 1 SDC-SMF Algorithm

```
1: (Initialization) Choose a time-horizon $T_f$. Given the initial values $(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$, set $k = 0$, $\hat{\boldsymbol{x}}_{k|k-1} = \hat{\boldsymbol{x}}_0$, $\boldsymbol{E}_{k|k-1} = \boldsymbol{E}_0$ where $\boldsymbol{P}_0 = \boldsymbol{E}_0 \boldsymbol{E}_0^{\mathrm{T}}$, and $\gamma_{k|k-1} = ||\boldsymbol{E}_0||$.
2: Calculate $r_{H_k}$ by solving (5.19). Find $\boldsymbol{P}_{k|k}$ and $\boldsymbol{L}_k$ by solving the SDP in (5.21).
3: Calculate $\hat{\boldsymbol{x}}_{k|k}$ using (5.8). Also, calculate $\boldsymbol{E}_{k|k}$ using $\boldsymbol{P}_{k|k} = \boldsymbol{E}_{k|k} \boldsymbol{E}_{k|k}^{\mathrm{T}}$ and set $\gamma_{k|k} = ||\boldsymbol{E}_{k|k}||$.
4: Calculate $r_{A_k}$ by solving (5.18). With that, given $\hat{\boldsymbol{x}}_{k|k}$, $\boldsymbol{E}_{k|k}$, $\gamma_{k|k}$, solve the SDP in (5.23) to obtain $\boldsymbol{P}_{k+1|k}$.
5: Calculate $\hat{\boldsymbol{x}}_{k+1|k}$ using (5.9). Set $\boldsymbol{E}_{k+1|k}$ using $\boldsymbol{P}_{k+1|k} = \boldsymbol{E}_{k+1|k} \boldsymbol{E}_{k+1|k}^{\mathrm{T}}$ and $\gamma_{k+1|k} = ||\boldsymbol{E}_{k+1|k}||$.
6: If $k = T_f$ stop. Otherwise, set $k = k + 1$ and go to Step 2.
```

Remark 5.2.3. *Note that the upper bounds calculated using* (5.18) *and* (5.19) *are conservative since the points are sampled from the boundary of the ellipsoids, whereas the true state of the system might belong to the interior of these sets. Assumption 5.1.4 means* $||\boldsymbol{w}_k|| \leq \sqrt{q}$ *and* $||\boldsymbol{v}_k|| \leq \sqrt{r}$ *for all* $k \in \mathbb{Z}_\star$*. Therefore, higher values of* q *and* r *would indicate that the available bounds on the noises are large, which would also introduce some degree of conservativeness to the SDC-SMF.*

Remark 5.2.4. *After the correction step is executed, the SDC-SMF guarantees the following:* $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$*. Thus, by denoting* $\boldsymbol{e}_k = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}$ *as the estimation error at the correction step, we deduce that* $\boldsymbol{e}_k$ *belongs to the ellipsoidal set* $\mathcal{E}(\mathbf{0}_n, \boldsymbol{P}_{k|k})$*, i.e.,* $\boldsymbol{e}_k \in \mathcal{E}(\mathbf{0}_n, \boldsymbol{P}_{k|k})$*. Note the conceptual similarity of this with the error covariance in Kalman filtering. Utilizing* $\boldsymbol{e}_k \in \mathcal{E}(\mathbf{0}_n, \boldsymbol{P}_{k|k})$*, estimation error bounds can be calculated which would serve as the confidence bounds in the context of the SDC-SMF. A similar notion applies for the prediction step as well. Note that ellipsoids smaller in 'size' would result in tighter error bounds.*

Remark 5.2.5. *Both the SDPs in* (5.21) *and* (5.23) *involve constraints such that the shape matrices are positive definite. However, in order to solve the SDPs using optimization toolboxes/packages such as YALMIP [170] and CVX [171], strict matrix inequalities* $\boldsymbol{P}_{k|k} > 0$ *and* $\boldsymbol{P}_{k+1|k} > 0$ *have to be replaced with non-strict matrix inequalities. Taking a practical approach for that, one can select* $\boldsymbol{P}_{k|k} \geq a_1 \boldsymbol{I}_n$ *and* $\boldsymbol{P}_{k+1|k} \geq a_2 \boldsymbol{I}_n$ *with* $a_1 > 0$*,* $a_2 > 0$ *as the tuning parameters that can be chosen for a given system.*

Until this point, we have discussed the SDC-SMF for system (5.4). Now, let us discuss the application of SDC-SMF to systems with known control inputs, i.e., systems of the form

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{f}_d(\boldsymbol{x}_k) + \sum_{i=1}^{m} \boldsymbol{g}_{d_i}(\boldsymbol{x}_k) u_{k_i} + \boldsymbol{w}_k = \boldsymbol{A}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{B}(\boldsymbol{x}_k)\boldsymbol{u}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{h}_d(\boldsymbol{x}_k) + \boldsymbol{v}_k = \boldsymbol{H}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{5.26}$$

where $\boldsymbol{u}_k \in \mathbb{R}^m$ is a vector of known control inputs and $\boldsymbol{f}_d(\cdot)$, $\boldsymbol{h}_d(\cdot)$ again satisfy Assumption 5.1.1. To be consistent with our earlier formulation, we choose the SDC parameterizations given in (5.3) and state the following assumption regarding the state dynamics of system (5.26).

Assumption 5.2.1. *There exist compact sets* $\mathbb{D}_{u_0}, \mathbb{D}_u \subset \mathbb{R}^n$, $\mathbb{U} \subset \mathbb{R}^m$, *and* $\epsilon_u > 0$ *such that* $\boldsymbol{x}_0 \in \mathbb{D}_{u_0}$ *and* $\boldsymbol{u}_k \in \mathbb{U}$ *together imply*

$$\boldsymbol{x}_k + \epsilon_u \mathcal{B}(\boldsymbol{x}_k) \subset \mathbb{D}_u, \quad \forall k \in \mathbb{Z}_\star.$$

The implication of the above assumption is similar to that of Assumption 5.1.2, i.e., the system (5.26) evolves within a compact set $\mathbb{D}_u$ which is not necessarily small. Then, the filtering problem at the correction step is as in Theorem 5.2.1 with system (5.4) replaced by system (5.26) and $\mathbb{D}_0$ in Assumption 5.1.3 replaced by $\mathbb{D}_{u_0}$. However, the SDP for the prediction step would have to be modified due to the control inputs acting through the state dependent control matrix. To this end, similar to the matrix Taylor expansion of $\boldsymbol{A}(\boldsymbol{x}_k)$ in (5.12), let us expand $\boldsymbol{B}(\boldsymbol{x}_k)$ as

$$\boldsymbol{B}(\boldsymbol{x}_k) = \boldsymbol{B}(\hat{\boldsymbol{x}}_{k|k}) + \boldsymbol{K}_3(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{\Delta}_3(\boldsymbol{\xi}_{k|k}) + \boldsymbol{R}_{\boldsymbol{B}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k) \tag{5.27}$$

where $\boldsymbol{K}_3(\hat{\boldsymbol{x}}_{k|k}) = \mathbb{D}_{\boldsymbol{x}^{\mathrm{T}}}\boldsymbol{B}(\hat{\boldsymbol{x}}_{k|k})$, $\boldsymbol{\Delta}_3(\boldsymbol{\xi}_{k|k}) = \big(\boldsymbol{\xi}_{k|k} \otimes \boldsymbol{I}_m\big)$ with $\boldsymbol{\xi}_{k|k}$ as in (5.12). Again, similar to (5.18)- (5.19), let us calculate the upper bound on the norm of remainder $\boldsymbol{R}_{\boldsymbol{B}_2}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{x}_k)$ as

$$r_{B_k} = \max_{i=1,2,\ldots,N} ||\boldsymbol{R}_{\boldsymbol{B}_2}(\hat{\boldsymbol{x}}_{k|k}, \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k_i})|| \tag{5.28}$$

where $||\boldsymbol{z}_{k|k_i}|| = 1,\ i = 1, 2, ..., N$. Finally, the next result summarizes the filtering problem at the prediction step for systems with state dynamics as in (5.26) where we have dropped the argument of $\boldsymbol{K}_3(\cdot)$.

Corollary 5.2.5.1. *Consider system* (5.26) *under Assumption 5.1.4 with the current state* $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ *and* $\boldsymbol{w}_k \in \mathcal{E}(\boldsymbol{0}_n, \boldsymbol{Q}_k)$*. Then, the successor state* $\boldsymbol{x}_{k+1}$ *belongs to the optimal prediction ellipsoid* $\mathcal{E}(\hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{P}_{k+1|k})$*, if there exist* $\boldsymbol{P}_{k+1|k} > 0$*,* $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8 \geq 0$ *as solutions to the following SDP:*

$$\min_{\boldsymbol{P}_{k+1|k}, \tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8} \quad \mathrm{trace}(\boldsymbol{P}_{k+1|k})$$

subject to

$$\boldsymbol{P}_{k+1|k} > 0$$

$$\tau_i \geq 0,\ i = 1, 2, 3, 4, 5, 6, 7, 8$$

$$\begin{bmatrix} -\boldsymbol{P}_{k+1|k} & \boldsymbol{\Pi}_{k|k} \\ \\ \boldsymbol{\Pi}^{\mathrm{T}}_{k|k} & -\boldsymbol{\Psi}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8) \end{bmatrix} \leq 0$$

where $\boldsymbol{\Pi}_{k|k}$ *and* $\boldsymbol{\Psi}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8)$ *are given by*

$$\begin{aligned} &\boldsymbol{\Pi}_{k|k} \\ &= \begin{bmatrix} \boldsymbol{0}_n & \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{E}_{k|k} & \boldsymbol{I}_n & \boldsymbol{K}_2 & \boldsymbol{I}_n & \boldsymbol{K}_2 & \boldsymbol{I}_n & \boldsymbol{K}_3 & \boldsymbol{I}_n \end{bmatrix} \\ &\boldsymbol{\Psi}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6, \tau_7, \tau_8) \\ &= \mathrm{diag}\,(1 - \tau_1 - \tau_2 - \tau_3\gamma^2_{k|k}\hat{\boldsymbol{x}}^{\mathrm{T}}_{k|k}\hat{\boldsymbol{x}}_{k|k} - \tau_4 r^2_{A_k}\hat{\boldsymbol{x}}^{\mathrm{T}}_{k|k}\hat{\boldsymbol{x}}_{k|k} - \tau_7\gamma^2_{k|k}\boldsymbol{u}^{\mathrm{T}}_k\boldsymbol{u}_k - \tau_8 r^2_{B_k}\boldsymbol{u}^{\mathrm{T}}_k\boldsymbol{u}_k, \\ &\quad \tau_1\boldsymbol{I}_n - \tau_5\gamma^2_{k|k}\boldsymbol{E}^{\mathrm{T}}_{k|k}\boldsymbol{E}_{k|k} - \tau_6 r^2_{A_k}\boldsymbol{E}^{\mathrm{T}}_{k|k}\boldsymbol{E}_{k|k}, \tau_2\boldsymbol{Q}^{-1}_k, \tau_3\boldsymbol{I}_{n^2}, \tau_4\boldsymbol{I}_n, \tau_5\boldsymbol{I}_{n^2}, \tau_6\boldsymbol{I}_n, \tau_7\boldsymbol{I}_{mn}, \tau_8\boldsymbol{I}_n). \end{aligned}$$

Furthermore, the center of the prediction ellipsoid is given by the predicted state estimate

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{B}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{u}_k. \tag{5.29}$$

Proof. Follows from that of Theorem 5.2.2 and is omitted. □

5.3 Theoretical Properties of SDC-SMF

In this section, we provide an elaborate sketch of the proof of the theoretical properties satisfied by the proposed SDC-SMF (which are similar to the observer properties described in Definition 3.1 in [57]). To this end, we utilize the approach outlined in Sections IV, V in [57] and adopt the symbols used to represent some variables in [57] so that it is easy to draw parallels between the results given here and the results in [57]. Further, for a sequence $\boldsymbol{x} = \{\boldsymbol{x}_k\}_{k\in\mathbb{Z}_\star}$ with $\boldsymbol{x}_k \in \mathbb{R}^n$ and $k \in \mathbb{Z}_\star$, we denote $||\boldsymbol{x}||_{l^\infty} = \sup_{k\in\mathbb{Z}_\star} ||\boldsymbol{x}_k||$. Due to the differences in the notations, simple modifications have to be introduced for Definition 3.1 in [57] and it is understood that those changes have already been carried out.

Remark 5.3.1. *Initialization step of the Algorithm 4.1 in [57] is similar to the initial correction step for the SDC-SMF at $k = 0$. Then, the the Algorithm 4.1 in [57] employs a one-step estimation wherein the correction and prediction are combined into one single step. For the SDC-SMF, we have two distinct steps for correction and prediction. However, for the analysis shown here, we would only consider the correction step with the corrected state estimate explicitly. The prediction step is only considered implicitly in the subsequent analysis. With this approach, we show that the corrected state estimate satisfies properties similar to the ones given in Definition 3.1 in [57]. Then, the same applies for the predicted state estimate under the conditions/assumptions described in the sequel.*

Consider the simplified version of system (5.4) given by

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}(\boldsymbol{x}_k)\boldsymbol{x}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{5.30}$$

where the state dependent matrix $\boldsymbol{H}(\boldsymbol{x}_k)$ is replaced by the constant matrix $\boldsymbol{H}$. This obviously introduces some loss of generality (which is remarked by the authors of [57] as well), but is crucial for establishing the theoretical properties, as shown in the sequel. Now, let (i) Assumption 5.1.2 hold for the state dynamics of system (5.30); (ii) Assumptions 5.1.3 and 5.1.4 hold for system (5.30); (iii) Assumption 5.1.5 hold with system (5.4) replaced by system (5.30) and $\boldsymbol{H}(\boldsymbol{x}_{(\cdot)})$ replaced by $\boldsymbol{H}$. With that, let us implement the proposed SDC-SMF for system (5.30). Note that Assumptions 3.1 and 5.2.1 in [57] are replaced by our Assumption 5.1.4. Under our Assumption 5.1.4, we have $||\boldsymbol{w}||_{l^\infty} \leq \sqrt{q}$ and $||\boldsymbol{v}||_{l^\infty} \leq \sqrt{r}$.

Before discussing the theoretical properties of the SDC-SMF for system (5.30), we give the next two assertions (Claims 5.3.1 and 5.3.2) under our above assumptions. First, we adopt the following claim from [57, Section V] which is asserted to hold due to the time-invariance and compactness assumptions.

Claim 5.3.1. *Let* $\alpha > 0$ *and* $\epsilon_2 > 0$ *be such that* $\forall k \in \mathbb{Z}_\star$

- $||\boldsymbol{A}(\boldsymbol{x})|| \leq \alpha,\ \forall \boldsymbol{x} \in \mathbb{D}$
- $||\boldsymbol{x} - \hat{\boldsymbol{x}}||_{l^\infty} \leq \epsilon_2$ *with* $\hat{\boldsymbol{x}} = \{\hat{\boldsymbol{x}}_{k|k}\}_{k\in\mathbb{Z}_\star}$ *implies*

$$\mu_1^o \boldsymbol{I}_n \leq \hat{\mathcal{O}}_{k,k+N_o-1}^T \hat{\mathcal{O}}_{k,k+N_o-1} \leq \mu_2^o \boldsymbol{I}_n$$

where $\mu_1^o(\mu_1, \mu_2) > 0$, $\mu_2^o(\mu_1, \mu_2) > 0$, *and*

$$\hat{\mathcal{O}}_{k,k+s} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{H}\hat{\boldsymbol{\phi}}_{k+1,k} \\ \vdots \\ \boldsymbol{H}\hat{\boldsymbol{\phi}}_{k+s,k} \end{bmatrix}$$

with

$$\hat{\boldsymbol{\phi}}_{k+s,k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k+s-1|k+s-1})\ \boldsymbol{A}(\hat{\boldsymbol{x}}_{k+s-2|k+s-2}) \cdots \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})$$

defined along the corrected state estimate trajectory for any $s \in \mathbb{Z}_\star \backslash \{0\}$.

Using the matrix Taylor expansion of $\boldsymbol{A}(\boldsymbol{x}_k)$, we have the state dynamics of the form (cf., (5.20))

$$\boldsymbol{x}_{k+1} = \ \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{x}_k + \boldsymbol{K}_2\boldsymbol{\Delta}_2\boldsymbol{x}_k + \boldsymbol{R}_{\boldsymbol{A}_2}\boldsymbol{x}_k + \boldsymbol{w}_k$$

where $\boldsymbol{\Delta}_2 = (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}) \otimes \boldsymbol{I}_n$ and we define

$$\begin{aligned} \boldsymbol{d}_{f_k} &= \boldsymbol{K}_2\boldsymbol{\Delta}_2\boldsymbol{x}_k + \boldsymbol{R}_{\boldsymbol{A}_2}\boldsymbol{x}_k \\ \mathbb{E}_k &= \{\boldsymbol{\nu} : \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{\nu} \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})\} \\ \rho_k &= \sup_{\boldsymbol{\nu}\in\mathbb{E}_k} ||\boldsymbol{\nu}||. \end{aligned}$$

Next, the following claim is related to the norm of the term $\boldsymbol{d}_{f_k}$ where ϵ_1 and $\mathbb{D}$ are as in Assumption 5.1.2.

Claim 5.3.2. *Define* $\bar{\epsilon} = \max\{\epsilon_1, \epsilon_2\}$*. Also, define a compact subset* $\bar{\mathbb{D}} \subset \mathbb{R}^n$ *such that* $\bar{\mathbb{D}} = \mathbb{D}$ *if* $\bar{\epsilon} = \epsilon_1$*, otherwise* $\bar{\mathbb{D}} \supseteq \mathbb{D}$ *with* $d_H(\mathbb{D}, \bar{\mathbb{D}}) \leq \bar{\epsilon}$ *where* $d_H(\cdot,\cdot)$ *is the Hausdorff distance. Let there exist* $\bar{a} > 0$ *such that* $||\boldsymbol{A}(\boldsymbol{x}_1) - \boldsymbol{A}(\boldsymbol{x}_2)|| \leq \bar{a}||\boldsymbol{x}_1 - \boldsymbol{x}_2||$ *for all* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \bar{\mathbb{D}}$*. Then,*

$$||\boldsymbol{d}_{f_k}|| \leq \delta\rho_k$$

for some $\delta > 0$*.*

Proof. The remainder of the matrix Taylor expansion can be expressed as in (5.14) with $\boldsymbol{K}_2 \equiv \boldsymbol{K}_2(\hat{\boldsymbol{x}}_{k|k})$. With the assertion in Claim 5.3.1 and the definition of the set $\bar{\mathbb{D}}$, we have $\boldsymbol{x}_k, \hat{\boldsymbol{x}}_{k|k} \in \bar{\mathbb{D}}$. Thus, under the assumption that $||\boldsymbol{A}(\boldsymbol{x}_1) - \boldsymbol{A}(\boldsymbol{x}_2)|| \leq \bar{a}||\boldsymbol{x}_1 - \boldsymbol{x}_2||$, we have

$$||\boldsymbol{R}_{\boldsymbol{A}_2}|| \leq \bar{a}||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| + ||\boldsymbol{K}_2||\ ||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}||$$

since $\boldsymbol{\Delta}_2 = (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}) \otimes \boldsymbol{I}_n$. Also, $||\boldsymbol{K}_2|| \leq k_2$ for some $k_2 > 0$ holds due to the continuity of $\boldsymbol{K}_2$ and compactness of $\bar{\mathbb{D}}$. Collecting all these, we deduce

$$||\boldsymbol{R}_{\boldsymbol{A}_2}|| \leq (\bar{a} + k_2)||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| \leq \alpha_r\rho_k$$

with some $\alpha_r > 0$. Also, $||\boldsymbol{x}_k|| \leq \alpha_x$ with some $\alpha_x > 0$ holds due to the compactness of $\mathbb{D}$. Therefore,

$$||\boldsymbol{d}_{f_k}|| \leq ||\boldsymbol{K}_2||\ ||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}||\ ||\boldsymbol{x}_k|| + ||\boldsymbol{R}_{\boldsymbol{A}_2}||\ ||\boldsymbol{x}_k|| \leq k_2\alpha_x\rho_k + \alpha_r\alpha_x\rho_k.$$

Combining all the above results, we conclude that there is a constant $\delta > 0$ such that

$$||\boldsymbol{d}_{f_k}|| \leq \delta\rho_k$$

holds $\forall k \in \mathbb{Z}_\star$. □

Remark 5.3.2. *Note that we have shown that the norm of the remainder term remains uniformly bounded under the Lipschitz continuity assumption on the matrix valued function $\boldsymbol{A}(\cdot)$. We stress that this assumption would hold due the continuous differentiability of the function and compactness of the sets. Furthermore, note that the above bound on the remainder term is developed using the methodology in Section 5.1 to calculate r_{A_k} at each time step. Thus, r_{A_k} would implicitly obey the above bound as well.*

We are now ready to establish the theoretical properties of the SDC-SMF for system (5.30). To this end, we first show that the SDC-SMF is nondivergent in the presence of the process and measurement noises and is unbiased and asymptotically convergent in the absence of the noises.

5.3.1 Nondivergence for $\boldsymbol{w}_k \neq \boldsymbol{0}_n$ and $\boldsymbol{v}_k \neq \boldsymbol{0}_p$

First, let us redefine the 'false' system in [57, Section IV.B]. Consider the following system

$$\begin{aligned} \boldsymbol{x}_{f_{k+1}} &= \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{x}_{f_k} + \boldsymbol{d}_{f_k} + \boldsymbol{u}_{f_k} + \boldsymbol{w}_k \\ \boldsymbol{x}_{f_0} &= \boldsymbol{0}_n \\ \boldsymbol{y}_{f_k} &= \boldsymbol{H}\boldsymbol{x}_{f_k} + \boldsymbol{v}_k \end{aligned} \qquad (5.31)$$

where $\boldsymbol{x}_{f_k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}$, $\boldsymbol{u}_{f_k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k} - \hat{\boldsymbol{x}}_{k+1|k+1}$. Note that this system is noncausal as in [57] and we have implicitly utilized the predicted state estimate in $\boldsymbol{u}_{f_k}$ as $\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k}$. Next, we state an important result that is subsequently utilized to show that the SDC-SMF is nondivergent for the case under consideration.

Proposition 5.3.3. *Given any $j \in \mathbb{Z}_\star$, let*

$$g_1 = \max\{1, \theta^j(\alpha+\delta)^j\}$$

$$g_2 = \sum_{k=1}^{j} \theta^k(\alpha+\delta)^{k-1}$$

$$g_3 = \sum_{k=1}^{j} \bar{l}\ \theta^{k-1}(\alpha+\delta)^{k-1}$$

where $\theta = (1+\bar{l}\ ||\boldsymbol{H}||)$ with $||\boldsymbol{L}_k|| \le \bar{l}$, $\forall k \in [1,j]$ for some $\bar{l} > 0$. Then,

$$\max_{0\le k\le j} \rho_k \le g_1\rho_0 + g_2||\boldsymbol{w}||_{l^\infty} + +g_3||\boldsymbol{v}||_{l^\infty}.$$

Proof. For any $k \in \mathbb{Z}_\star \backslash \{0\}$, we have

$$\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{H}\ (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) - \boldsymbol{L}_k\boldsymbol{v}_k$$

where

$$\boldsymbol{x}_k = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\boldsymbol{x}_{k-1} + \boldsymbol{d}_{f_{k-1}} + \boldsymbol{w}_{k-1}$$

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\hat{\boldsymbol{x}}_{k-1|k-1}.$$

Therefore, we can write

$$\begin{aligned}
&\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} \\
&= \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\boldsymbol{x}_{k-1} + \boldsymbol{d}_{f_{k-1}} + \boldsymbol{w}_{k-1} - \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\hat{\boldsymbol{x}}_{k-1|k-1} \\
&\quad - \boldsymbol{L}_k\boldsymbol{H}\Big(\boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\boldsymbol{x}_{k-1} + \boldsymbol{d}_{f_{k-1}} + \boldsymbol{w}_{k-1} - \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\hat{\boldsymbol{x}}_{k-1|k-1}\Big) - \boldsymbol{L}_k\boldsymbol{v}_k \\
&= \ \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{d}_{f_{k-1}} + \boldsymbol{w}_{k-1} \\
&\quad - \boldsymbol{L}_k\boldsymbol{H}\ \Big(\boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\Big(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\Big) + \boldsymbol{d}_{f_{k-1}} + \boldsymbol{w}_{k-1}\Big) - \boldsymbol{L}_k\boldsymbol{v}_k.
\end{aligned}$$

Let $\bar{l} > 0$ be such that $||\boldsymbol{L}_k|| \leq \bar{l},\ \forall k \in [1, j]$. Hence, we derive

$$\begin{aligned}
\rho_k =& \max_{\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})} ||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| \\
&\leq \alpha\rho_{k-1} + \delta\rho_{k-1} + ||\boldsymbol{w}||_{l^\infty} + ||\boldsymbol{L}_k||\ ||\boldsymbol{H}||(\alpha\rho_{k-1} + \delta\rho_{k-1} + ||\boldsymbol{w}||_{l^\infty}) + ||\boldsymbol{L}_k||\ ||\boldsymbol{v}||_{l^\infty} \\
&\leq (1 + \bar{l}\ ||\boldsymbol{H}||)\left((\alpha+\delta)\rho_{k-1} + ||\boldsymbol{w}||_{l^\infty}\right) + \bar{l}\ ||\boldsymbol{v}||_{l^\infty}.
\end{aligned}$$

Carrying out these calculations recursively yields

$$\rho_j \leq \theta^j(\alpha+\delta)^j\rho_0 + \sum_{k=1}^{j} \theta^k(\alpha+\delta)^{k-1}||\boldsymbol{w}||_{l^\infty} + \sum_{k=1}^{j} \bar{l}\ \theta^{k-1}(\alpha+\delta)^{k-1}||\boldsymbol{v}||_{l^\infty}$$

where $\theta = (1 + \bar{l}\ ||\boldsymbol{H}||)$. Then, collecting all the required bounds leads to the desired result. □

Remark 5.3.4. *Note that we have used a uniform bound $||\boldsymbol{L}_k|| \leq \bar{l}$ for the filter gain. This is guaranteed to hold as the filter gain is a solution to a convex optimization problem (namely, SDP) at each time step.*

Let $\epsilon^\star = \min\{\epsilon_1, \epsilon_2\}$. We need to show that for $k \in [0, N_o - 1]$

$$||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| \leq \epsilon^\star.$$

Then,

$$\mu_1^o \boldsymbol{I}_n \leq \hat{\mathcal{O}}_{0,N_o-1}^{\mathrm{T}} \hat{\mathcal{O}}_{0,N_o-1} \leq \mu_2^o \boldsymbol{I}_n. \tag{5.32}$$

Making straightforward modifications to the result in Proposition 5.3.3, we can assure

$$\max_{0 \leq k \leq N_o - 1} \rho_k \leq \epsilon^\star$$

whenever $\boldsymbol{x}_0 \in \mathbb{D}_0$, $\rho_0 \leq \bar{\rho}_1$, $||\boldsymbol{w}||_{l^\infty} \leq \bar{d}$, and $||\boldsymbol{v}||_{l^\infty} \leq \bar{n}$ where

$$\begin{aligned}
\bar{\rho}_1 &= \min\left\{\frac{\epsilon^\star}{3}, \frac{(\epsilon^\star/3)}{\theta^{N_o-1}(\alpha+\delta)^{N_o-1}}\right\} \\
\bar{d} &= \frac{(\epsilon^\star/3)}{\sum_{k=1}^{N_o-1} \theta^k(\alpha+\delta)^{k-1}} \\
\bar{n} &= \frac{(\epsilon^\star/3)}{\sum_{k=1}^{N_o-1} \bar{l}\ \theta^{k-1}(\alpha+\delta)^{k-1}}.
\end{aligned}$$

This, in turn, implies (5.32).

Next, let us implement the gramian-based observer for the 'false' system (5.31), as in [57]. Doing so, we have

$$\begin{aligned}
&||\boldsymbol{x} - \hat{\boldsymbol{x}}_{N_o|N_o} - \hat{\boldsymbol{x}}_{g_{N_o}}|| \\
&\leq \beta_1(||\boldsymbol{w}||_{l^\infty} + \max_{0\leq k\leq N_o-1} ||\boldsymbol{d}_{f_k}||) + \beta_2||\boldsymbol{v}||_{l^\infty} \\
&\leq \beta_1(||\boldsymbol{w}||_{l^\infty} + \delta \max_{0\leq k\leq N_o-1} \rho_k) + \beta_2||\boldsymbol{v}||_{l^\infty}
\end{aligned}$$

for any $\boldsymbol{x} \in \mathcal{E}(\hat{\boldsymbol{x}}_{N_o|N_o}, \boldsymbol{P}_{N_o|N_o})$ and with $\beta_1, \beta_2 > 0$. Note that the above bound holds for $\boldsymbol{x} = \hat{\boldsymbol{x}}_{N_o|N_o}$. With this, we derive

$$\begin{aligned}
\rho_{N_o} &= \max_{\boldsymbol{x}\in\mathcal{E}(\hat{\boldsymbol{x}}_{N_o|N_o}, \boldsymbol{P}_{N_o|N_o})} ||\boldsymbol{x} - \hat{\boldsymbol{x}}_{N_o|N_o}|| \\
&\leq ||\boldsymbol{x} - \hat{\boldsymbol{x}}_{N_o|N_o} - \hat{\boldsymbol{x}}_{g_{N_o}}|| + ||\hat{\boldsymbol{x}}_{g_{N_o}}|| \\
&\leq 2\beta_1(||\boldsymbol{w}||_{l^\infty} + \delta \max_{0\leq k\leq N_o-1} \rho_k) + 2\beta_2||\boldsymbol{v}||_{l^\infty}.
\end{aligned}$$

Utilizing the result in Proposition 5.3.3 with $j = N_o - 1$, we have

$$\rho_{N_o} \leq 2\beta_1\left(||\boldsymbol{w}||_{l^\infty} + \delta(g_1\rho_0 + g_2||\boldsymbol{w}||_{l^\infty} + g_3||\boldsymbol{v}||_{l^\infty})\right) + 2\beta_2||\boldsymbol{v}||_{l^\infty}$$

which upon rearranging becomes

$$\rho_{N_o} \leq c_1\rho_0 + c_2||\boldsymbol{w}||_{l^\infty} + c_3||\boldsymbol{v}||_{l^\infty}$$

where $c_1 = 2\beta_1\delta g_1$, $c_2 = 2\beta_1(1 + \delta g_2)$, and $c_3 = 2(\beta_2 + \beta_1\delta g_3)$. Thus,

$$\begin{aligned}
\rho_0 &\leq \frac{\bar{\rho}_1}{(c_1+c_2+c_3)(c_1+c_2+c_3+1)} \\
||\boldsymbol{w}||_{l^\infty} &\leq \min\left\{\bar{d}, \frac{\bar{\rho}_1}{(c_1+c_2+c_3)(c_1+c_2+c_3+1)}\right\} \\
||\boldsymbol{v}||_{l^\infty} &\leq \min\left\{\bar{n}, \frac{\bar{\rho}_1}{(c_1+c_2+c_3)(c_1+c_2+c_3+1)}\right\}
\end{aligned}$$

together imply

$$\rho_{N_o} \leq \frac{\bar{\rho}_1}{(c_1+c_2+c_3+1)}$$

which is similar to the result (12) in [57]. Therefore, the rest of the proof of uniform boundedness of ρ_k and nondivergence of the corrected state estimate follows from arguments similar to the ones outlined in [57].

5.3.2 Unbiasedness and Asymptotic Convergence for $\boldsymbol{w}_k = \boldsymbol{0}_n$ and $\boldsymbol{v}_k = \boldsymbol{0}_p$

In this case, the SDC-SMF is clearly unbiased for $\mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0) = \boldsymbol{x}_0$. Next, let us redefine the 'false' system of Proposition 4.1 in [57]. Consider the following system

$$\begin{aligned} \boldsymbol{x}_{f_{k+1}} &= \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\boldsymbol{x}_{f_k} + \boldsymbol{d}_{f_k} + \boldsymbol{u}_{f_k} \\ \boldsymbol{y}_{f_k} &= \boldsymbol{H}\boldsymbol{x}_{f_k} \end{aligned} \tag{5.33}$$

where $\boldsymbol{x}_{f_k} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}$, $\boldsymbol{u}_{f_k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k|k})\hat{\boldsymbol{x}}_{k|k} - \hat{\boldsymbol{x}}_{k+1|k+1}$. This system is obviously similar to the earlier 'false' system (5.31). Now, we state a result similar to the one in Proposition 5.3.3.

Proposition 5.3.5. *Given any $j \in \mathbb{Z}_\star$, let*

$$g = \max\left\{1, \theta^j(\alpha+\delta)^j\right\}$$

where $\theta = (1 + \bar{l}\ ||\boldsymbol{H}||)$ with $||\boldsymbol{L}_k|| \le \bar{l}, \forall k \in [1, j]$. Then,

$$\max_{0\le k\le j} \rho_k \le g\rho_0.$$

Proof. For any $k \in \mathbb{Z}_\star\backslash\{0\}$, we have

$$\begin{aligned} \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = \boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{d}_{f_{k-1}} \\ - \boldsymbol{L}_k\boldsymbol{H}\left(\boldsymbol{A}(\hat{\boldsymbol{x}}_{k-1|k-1})\left(\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}\right) + \boldsymbol{d}_{f_{k-1}}\right). \end{aligned}$$

As earlier, let $\bar{l} > 0$ be such that $||\boldsymbol{L}_k|| \le \bar{l}$, $\forall k \in [1, j]$. With that, the above expression implies

$$\begin{aligned} \rho_k &= \max_{\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})} ||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| \\ &\le \alpha\rho_{k-1} + \delta\rho_{k-1} + ||\boldsymbol{L}_k||\ ||\boldsymbol{H}||(\alpha\rho_{k-1} + \delta\rho_{k-1}) \\ &\le (1 + \bar{l}\ ||\boldsymbol{H}||)(\alpha+\delta)\rho_{k-1}. \end{aligned}$$

Proceeding recursively for $k = 1, 2, \ldots, j$ leads to the desired result. □

Same as earlier, we need to show that for $k \in [0, N_o - 1]$

$$||\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}|| \leq \epsilon^\star.$$

To this end, using the result in Proposition 5.3.5, we have

$$\max_{0 \leq k \leq N_o - 1} \rho_k \leq \epsilon^\star$$

whenever $\boldsymbol{x}_0 \in \mathbb{D}_0$, $\rho_0 \leq \bar{\rho}_2$ where

$$\bar{\rho}_2 = \min \left\{ \epsilon^\star, \frac{\epsilon^\star}{\theta^{N_o - 1}(\alpha + \delta)^{N_o - 1}} \right\}$$

which, in turn, implies (5.32).

Next, we implement the gramian-based observer, as in [57], for the 'false' system (5.33) and derive

$$||\boldsymbol{x} - \hat{\boldsymbol{x}}_{N_o|N_o} - \hat{\boldsymbol{x}}_{g_{N_o}}|| \leq \beta \max_{0 \leq k \leq N_o - 1} ||\boldsymbol{d}_{f_k}|| \leq \beta\delta \max_{0 \leq k \leq N_o - 1} \rho_k$$

for any $\boldsymbol{x} \in \mathcal{E}(\hat{\boldsymbol{x}}_{N_o|N_o}, \boldsymbol{P}_{N_o|N_o})$ and with $\beta > 0$. Therefore, utilizing the result in Proposition 5.3.5 with $j = N_o - 1$, we have

$$\rho_{N_o} = \max_{\boldsymbol{x} \in \mathcal{E}(\hat{\boldsymbol{x}}_{N_o|N_o}, \boldsymbol{P}_{N_o|N_o})} ||\boldsymbol{x} - \hat{\boldsymbol{x}}_{N_o|N_o}||$$

$$\leq 2\beta\delta \max_{0 \leq k \leq N_o - 1} \rho_k \leq 2\beta\delta g \rho_0.$$

Then, for

$$\beta \leq \frac{\lambda}{2\delta g}, \ \lambda \in (0, 1), \ \rho_0 \leq \bar{\rho}_2,$$

we have

$$\rho_{N_o} \leq \lambda \rho_0.$$

The above inequality also implies that $\rho_{N_o} < \bar{\rho}_2$. Thus, the uniform boundedness in Claim 5.3.1 holds and the above process can be repeated to derive the following:

$$\rho_{kN_o} \leq \lambda \rho_{(k-1)N_o} \leq \cdots \leq \lambda^k \rho_0$$

which is similar to the result given in [57]. This clearly establishes the asymptotic convergence property, i.e., $\lim_{k\to\infty} \rho_k = 0$.

Finally, we note that the above analyses also imply boundedness of the correction ellipsoid shape matrices. To this end, we note that $\boldsymbol{\nu} \equiv \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k}$ and $\rho_k \equiv \gamma_{k|k}$ where $\boldsymbol{E}_{k|k}$, $\boldsymbol{z}_{k|k}$, and $\gamma_{k|k}$ are as in Section 5.1. Now, consider the case of nondivergence. Since ρ_k is uniformly bounded, so is $\gamma_{k|k}$. This implies that the correction ellipsoid shape matrices remain uniformly bounded. Next, consider the asymptotic convergence case. For this, $\lim_{k\to\infty} \rho_k = 0$ means $\lim_{k\to\infty} \boldsymbol{z}_{k|k} = \mathbf{0}_n$. Then, due to the nature of set-membership filtering technique (i.e., at every time step, the correction ellipsoid is synthesized by solving a convex optimization problem that guarantees to contain the true state with the corrected state estimate at the corresponding center), we again have $\gamma_{k|k}$ bounded. A similar set of arguments can be made for the prediction ellipsoid shape matrices as well. This completes our discussion on the theoretical properties of the SDC-SMF for system (5.30).

5.4 Simulation Example

A simulation example is provided in this section to illustrate the effectiveness of the proposed approach. All the simulations are carried out on a laptop computer with 8.00 GB RAM and 1.60-1.80 GHz Intel(R) Core(TM) i5-8250U processor running MATLAB R2019b. The SDPs in (5.21) and (5.23) are solved utilizing 'YALMIP' [170] with the 'SDPT3' solver in the MATLAB framework.

Let us consider the Van der Pol equation in [57] and express the discrete-time system as

$$\begin{aligned}\boldsymbol{x}_{k+1} &= \begin{bmatrix} x_{1_k} + \Delta t x_{2_k} \\ x_{2_k} + \Delta t(-9x_{1_k} + \mu(1 - x_{1_k}^2)x_{2_k}) \end{bmatrix} + \begin{bmatrix} 0 \\ w_k \end{bmatrix}, \\ &= \boldsymbol{f}_d(\boldsymbol{x}_k) + \boldsymbol{w}_k, \\ y_k &= x_{1_k} + v_k = \boldsymbol{h}_d(\boldsymbol{x}_k) + v_k \end{aligned}$$

where $\boldsymbol{x}_{(\cdot)} = [x_{1_{(\cdot)}} \quad x_{2_{(\cdot)}}]^{\mathrm{T}}$ and Δt is the discretization time step. Clearly, the functions in the above system satisfy Assumption 5.1.1. Then, utilizing (5.3), we have

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{x}_k) &= \begin{bmatrix} 1 & \Delta t \\ -9\Delta t - \frac{2}{3}\mu\Delta t x_{1_k} x_{2_k} & 1 + \mu\Delta t(1 - \frac{1}{3}x_{1_k}^2) \end{bmatrix} \\ \boldsymbol{H}(\boldsymbol{x}_k) &= \begin{bmatrix} 1 & 0 \end{bmatrix}. \end{aligned}$$

With these SDC matrices, we have

$$\mathcal{O}_{k,k+1} = \begin{bmatrix} 1 & 0 \\ 1 & \Delta t \end{bmatrix} \tag{5.34}$$

which is full-rank for all $\Delta t \neq 0$. Thus, the rank condition in (5.6) is satisfied with $N_o = 2$. We take $\mu = 2$ for which the Van der Pol equation (nominal part) admits a unique and stable limit cycle, thus satisfying Assumption 5.1.2. Also, we set $\Delta t = 0.05$ seconds and use $N = 1000$ for calculating r_{A_k}. With the above SDC parameterizations, the matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are given by

$$\begin{aligned}\boldsymbol{K}_1 &= \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \\ \boldsymbol{K}_2 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\frac{2}{3}\mu\Delta t\hat{x}_{2_{k|k}} & -\frac{2}{3}\mu\Delta t\hat{x}_{1_{k|k}} & -\frac{2}{3}\mu\Delta t\hat{x}_{1_{k|k}} & 0 \end{bmatrix}. \end{aligned}$$

In this example, the initial condition is given by $\boldsymbol{P}_0 = \boldsymbol{I}_2$, $\boldsymbol{x}_0 = [1.5 \quad 1.25]^{\mathrm{T}}$, and

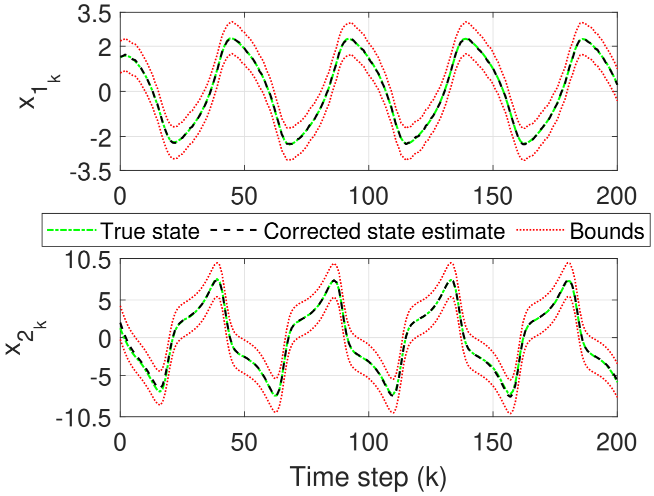


Figure 5.1: Simulation results corresponding to the SDC-SMF.

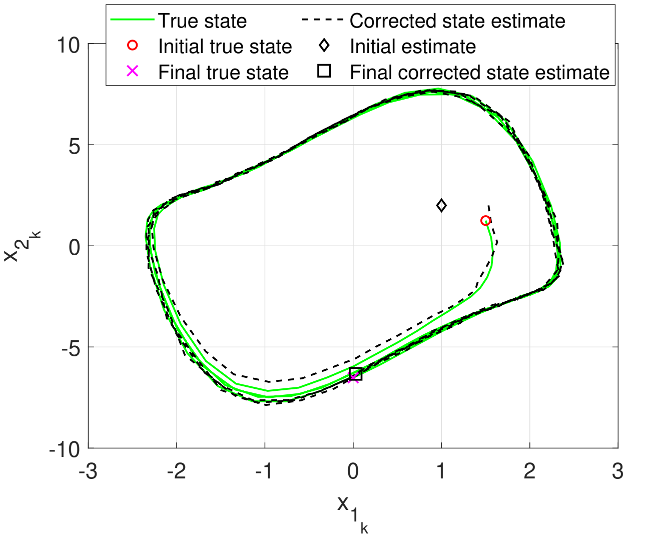


Figure 5.2: True state and corrected state estimate trajectories in the phase plane.

$\hat{\boldsymbol{x}}_0 = [1 \quad 2]^{\mathrm{T}}$. For Assumption 5.1.3, we can consider $\mathbb{D}_0 = \mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$. In terms of Assumption 5.1.4, let us choose $q = r = 0.01$. Then, Assumption 5.1.4 is satisfied with (i) w_k and v_k randomly varying (uniform distribution) between -0.05 and 0.05; (ii) $\boldsymbol{Q}_k = 0.01\boldsymbol{I}_2$, $\boldsymbol{R}_k = 0.01$. The true state components along with the corresponding corrected state estimates and bounds are shown in Fig. 5.1 as functions of time steps.

Clearly, x_{1_k}, x_{2_k} remain within the bounds for the entire time-horizon considered which mean that the true state is successfully contained in the correction ellipsoids. Fig. 5.2 depicts the true state trajectory and the corrected state estimate trajectory in the phase plane. Note that, at $k = 0$, the correction step brings the corrected state estimate close to the initial true state. Also, it is obvious that the corrected state estimate trajectory converges to and remains in a neighborhood of the true state trajectory after a few recursions of the filter.

Table 5.1: Mean trace and estimation error comparisons over 200 time steps

Item	SDC-SMF	Wang et al. [67]
Mean trace	5.5007	6.2616
MAE	0.1142	0.1761
MSE	0.0277	0.0643

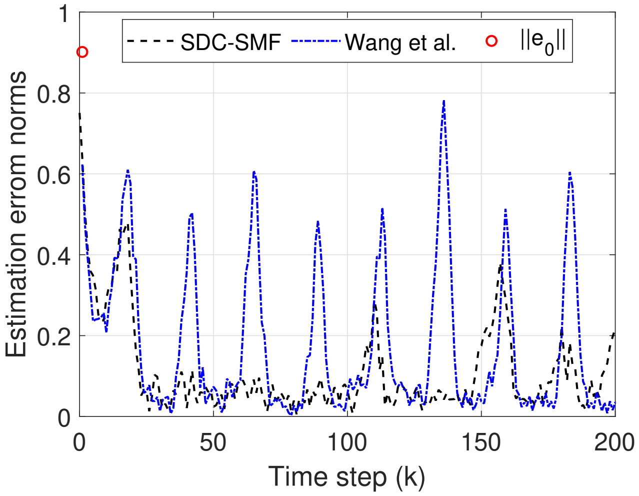


Figure 5.3: Estimation error norms for the SDC-SMF and the SMF in [67] (Wang et al.).

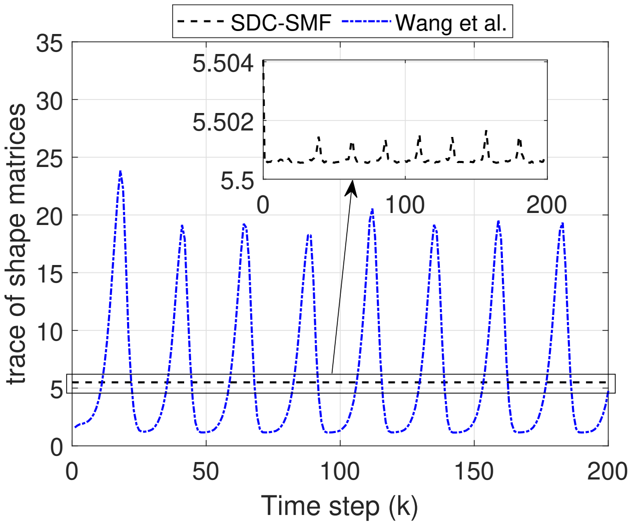


Figure 5.4: Trace of correction ellipsoid shape matrices for the SDC-SMF and state estimation ellipsoid shape matrices for the SMF in [67] (Wang et al.).

Next, for comparison, we implement the SMF in [67] for the above example with the remainder bounding ellipsoids synthesized using 50 constraints. Let us consider the estimation errors at the correction steps for the SDC-SMF and at the measurement update steps for the SMF in [67]. The comparison in these estimation error norms is shown in Fig. 5.3 where $||\boldsymbol{e}_0|| = ||\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0||$ is the initial error norm and the comparison in trace of the corresponding ellipsoid shape matrices is shown in Fig. 5.4. The results in Figs. 5.3, 5.4 demonstrate that the SDC-SMF outperforms the SMF in [67]. This is further illustrated in the results given in Table 5.1 where MAE and MSE stand for mean absolute error and mean squared error, respectively. The SDC-SMF performs much better in terms of these two metrics, as shown in Table 5.1. Also, the mean trace value for the SDC-SMF correction ellipsoid shape matrices is smaller compared to that of the state estimation ellipsoid shape matrices for the SMF in [67]. In summary, the SDC-SMF results in lower estimation errors with lower error bounds for this example.

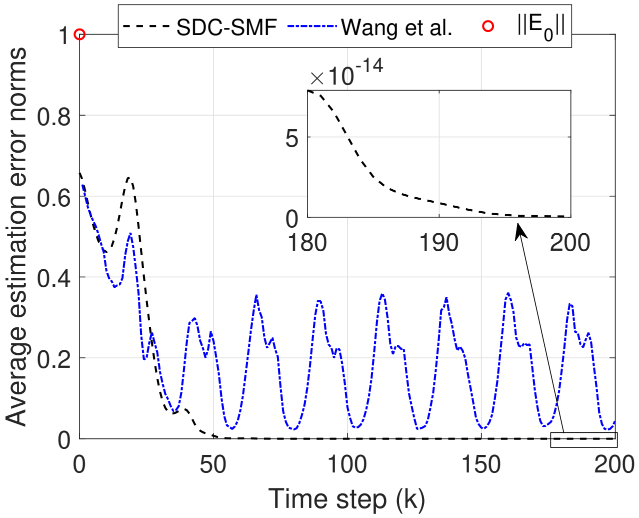


Figure 5.5: Average estimation error norms with $w_k = v_k = 0$ and $n_r = 10$.

Finally, to demonstrate that the SDC-SMF is asymptotically convergent if there are no process and measurement noises (cf., Section 5.3), we implement the SDC-SMF for the above example with the initial state randomly chosen from the boundary of the initial ellipsoid $\mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$ and with $w_k = v_k = 0$. We repeat this process n_r times. The same is done for the SMF in [67] as well. The average estimation error norms of these runs with the random initializations are shown in Fig. 5.5 where $\boldsymbol{E}_0$ is the Cholesky factorization of $\boldsymbol{P}_0$. Note that the upper bound of the initial error norm for the random initializations is $||\boldsymbol{E}_0||$, which is shown in Fig. 5.5. The results in Fig. 5.5 show that the SDC-SMF is asymptotically convergent with the estimation error tending to zero. However, the SMF in [67] does not exhibit this property, as shown in Fig. 5.5.

5.5 Chapter Summary

A recursive set-membership filtering algorithm for discrete-time nonlinear dynamical systems subject to unknown but bounded process and measurement noise has been derived utilizing the state dependent coefficient (SDC) parameterization. At each time step, the filtering problem has been transformed into two semi-definite programs (SDPs) using the S-procedure and Schur complement. Optimal (minimum trace) ellipsoids have been constructed that contain the true state of the system at the correction and prediction steps. Finally, a simulation example is provided which demonstrates that the proposed filter performs better compared to an existing set-membership filter for discrete-time nonlinear systems.

Chapter 6

Set-Membership Filtering-Based Leader-Follower Synchronization of Discrete-time Linear Multi-Agent Systems[*]

In this chapter, we discuss a leader-follower synchronization protocol design for high-order discrete-time linear multi-agent systems using set-membership filtering. In this regard, the set-membership filter (SMF) is a linear version of the SDC-SMF in Chapter 5. Note that an abbreviated version of the materials in this chapter has been reported in reference [131]. The symbol $|\cdot|$ denotes standard Euclidean norm for vectors and induced matrix norm for matrices. For any function $\boldsymbol{\theta} : \mathbb{Z}_\star \to \mathbb{R}^n$, we have $||\boldsymbol{\theta}|| = \sup\{|\boldsymbol{\theta}_k| : k \in \mathbb{Z}_\star\}$. This is the standard l_∞ norm for a bounded $\boldsymbol{\theta}$. The rest of this chapter is organized as follows. Section 6.1 describes the preliminaries required for the SMF design. The formulation of the SMF is given in Section 6.2. The control input synthesis and related results for synchronization are given in Section 6.3. Finally, Section 6.4 includes the simulation examples and Section 6.5 presents the concluding remarks.

[*] Most of the materials of this chapter have been published in D. Bhattacharjee and K. Subbarao, "Set-Membership Filtering-Based Leader-Follower Synchronization of Discrete-Time Linear Multi-Agent Systems," *ASME Journal of Dynamic Systems, Measurement, and Control*, Vol. 143, Issue 6, p. 064502, 2021, DOI: 10.1115/1.4049553, Copyright © 2021 ASME (reference [131]), and these are included here under the permission granted by ASME.

6.1 Preliminaries

Consider the discrete-time dynamical systems of the form

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}_k \boldsymbol{x}_k + \boldsymbol{B}_k \boldsymbol{u}_k + \boldsymbol{G}_k \boldsymbol{w}_k, \\ \boldsymbol{y}_k &= \boldsymbol{C}_k \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{v}_k, \quad k \in \mathbb{Z}_\star \end{aligned} \tag{6.1}$$

where $\boldsymbol{x}_k \in \mathbb{R}^{\bar{n}}$ is the state, $\boldsymbol{u}_k \in \mathbb{R}^{\bar{m}}$ is the control input, $\boldsymbol{w}_k \in \mathbb{R}^{\bar{w}}$ is the input disturbance, $\boldsymbol{y}_k \in \mathbb{R}^{\bar{p}}$ is the measured output, $\boldsymbol{v}_k \in \mathbb{R}^{\bar{v}}$ is the output disturbance. Also, $\boldsymbol{A}_k$, $\boldsymbol{B}_k$, $\boldsymbol{G}_k$, $\boldsymbol{C}_k$ and $\boldsymbol{D}_k$ are system matrices of appropriate dimensions. Following are the assumptions for systems of the form given in Eq. (6.1).

Assumption 6.1.1. *The initial state $\boldsymbol{x}_0$ is unknown. However, it satisfies $\boldsymbol{x}_0 \in \mathcal{E}(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$ where $\hat{\boldsymbol{x}}_0$ is a given initial estimate and $\boldsymbol{P}_0$ is known.*

Assumption 6.1.2. *$\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are unknown-but-bounded for all $k \in \mathbb{Z}_\star$. Also, $\boldsymbol{w}_k \in \mathcal{E}(\boldsymbol{0}_{\bar{w}}, \boldsymbol{Q}_k)$ and $\boldsymbol{v}_k \in \mathcal{E}(\boldsymbol{0}_{\bar{v}}, \boldsymbol{R}_k)$ for all $k \in \mathbb{Z}_\star$ where $\boldsymbol{Q}_k$, $\boldsymbol{R}_k$ are known.*

We intend to develop an SMF for systems of the form in Eq. (6.1), having a correction-prediction structure similar to the Kalman filter variants (see, for example, [45]). We construct such an SMF by simplifying the SDC-SMF design in Chapter 5, and the relevant details are included here.

Following are the filtering objectives where the corrected and predicted state estimates at time-step k are denoted by $\hat{\boldsymbol{x}}_{k|k}$ and $\hat{\boldsymbol{x}}_{k+1|k}$, respectively.

6.1.1 Correction Step

At each time-step $k \in \mathbb{Z}_\star$, upon receiving the measured output $\boldsymbol{y}_k$ with $\boldsymbol{v}_k \in \mathcal{E}(\boldsymbol{0}_{\bar{v}}, \boldsymbol{R}_k)$ and given $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, the objective is to find the optimal correction ellipsoid, characterized by $\hat{\boldsymbol{x}}_{k|k}$ and $\boldsymbol{P}_{k|k}$, such that $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$. The corrected state estimate is given by

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{L}_k(\boldsymbol{y}_k - \boldsymbol{C}_k \hat{\boldsymbol{x}}_{k|k-1}) \tag{6.2}$$

where $\boldsymbol{L}_k$ is the filter gain. Since $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, there exists a $\boldsymbol{z}_{k|k-1} \in \mathbb{R}^{\bar{n}}$ with $|\boldsymbol{z}_{k|k-1}| \leq 1$ such that

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \tag{6.3}$$

where $\boldsymbol{E}_{k|k-1}$ is the Cholesky factorization of $\boldsymbol{P}_{k|k-1}$, i.e., $\boldsymbol{P}_{k|k-1} = \boldsymbol{E}_{k|k-1}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}$ [52, 54].

6.1.2 Prediction Step

At each time-step $k \in \mathbb{Z}_\star$, given $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\boldsymbol{w}_k \in \mathcal{E}(\boldsymbol{0}_{\bar{w}}, \boldsymbol{Q}_k)$, the objective is to find the optimal prediction ellipsoid, characterized by $\hat{\boldsymbol{x}}_{k+1|k}$ and $\boldsymbol{P}_{k+1|k}$, such that $\boldsymbol{x}_{k+1} \in \mathcal{E}(\hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{P}_{k+1|k})$ where the predicted state estimate is given by

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{A}_k\hat{\boldsymbol{x}}_{k|k} + \boldsymbol{B}_k\boldsymbol{u}_k \tag{6.4}$$

Again, since $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$, we have

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_{k|k} + \boldsymbol{E}_{k|k}\boldsymbol{z}_{k|k} \tag{6.5}$$

where $\boldsymbol{P}_{k|k} = \boldsymbol{E}_{k|k}\boldsymbol{E}_{k|k}^{\mathrm{T}}$ and $|\boldsymbol{z}_{k|k}| \leq 1$. Initialization is provided by $\hat{\boldsymbol{x}}_{0|-1} = \hat{\boldsymbol{x}}_0$ and $\boldsymbol{P}_{0|-1} = \boldsymbol{P}_0$ [45].

Remark 6.1.1. *As mentioned in the filtering objectives, we are interested in finding the optimal ellipsoids, i.e., the minimum-'size' ellipsoids, at each time-step. There are two criteria for the 'size' of an ellipsoid in terms of its shape matrix: trace criterion and log-determinant criterion [54]. In this chapter, we have considered the trace criterion (see Theorems 6.2.1 and 6.2.2) which represents the sum of squared lengths of semi-axes of an ellipsoid [54]. As a result, the corresponding optimization problems are convex (see the SDPs in Eqs.* (6.6) *and* (6.15)*). Alternatively, for minimum-volume ellipsoids, one can consider the log-determinant criterion. However,* *this would render the optimization problems non-convex and additional modifications might be required to restore convexity (see, for example, [54]).*

6.2 Set-Membership Filter Design

In this section, we formulate the SDPs to be solved at each time-step for the SMF. These are essentially simplified versions of the results in Theorems 5.2.1 and 5.2.2. First, we state the result that summarizes the filtering problem at the correction step.

Theorem 6.2.1. *Consider the system in Eq.* (6.1) *under the Assumptions 6.1.1 and 6.1.2. Then, at each time-step* $k \in \mathbb{Z}_\star$, *upon receiving the measured output* $\boldsymbol{y}_k$ *with* $\boldsymbol{v}_k \in \mathcal{E}(\mathbf{0}_{\bar{v}}, \boldsymbol{R}_k)$ *and given* $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1})$, *the state* $\boldsymbol{x}_k$ *is contained in the optimal correction ellipsoid given by* $\mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$, *if there exist* $\boldsymbol{P}_{k|k} > 0$, $\boldsymbol{L}_k$, $\tau_i \geq 0,\ i = 1, 2$ *as solutions to the following SDP:*

$$\begin{aligned} &\min_{\boldsymbol{P}_{k|k}, \boldsymbol{L}_k, \tau_1, \tau_2} \quad \mathrm{trace(P_{k|k})} \\ &\text{subject to} \\ &\boldsymbol{P}_{k|k} > 0 \\ &\tau_i \geq 0,\ i = 1, 2 \\ &\begin{bmatrix} -\boldsymbol{P}_{k|k} & \boldsymbol{\Pi}_{k|k-1} \\ \\ \boldsymbol{\Pi}^{\mathrm{T}}_{k|k-1} & -\boldsymbol{\Theta}(\tau_1, \tau_2) \end{bmatrix} \leq 0 \end{aligned} \tag{6.6}$$

where $\boldsymbol{\Pi}_{k|k-1}$ *and* $\boldsymbol{\Theta}(\tau_1, \tau_2)$ *are given by*

$$\begin{aligned} \boldsymbol{\Pi}_{k|k-1} &= \Big[\mathbf{0}_{\bar{n}} \quad (\boldsymbol{E}_{k|k-1} - \boldsymbol{L}_k \boldsymbol{C}_k \boldsymbol{E}_{k|k-1}) \quad -\boldsymbol{L}_k \boldsymbol{D}_k\Big], \\ \boldsymbol{\Theta}(\tau_1, \tau_2) &= \mathrm{diag}\left(1 - \tau_1 - \tau_2, \tau_1 \boldsymbol{I}_{\bar{n}}, \tau_2 \boldsymbol{R}_k^{-1}\right) \end{aligned} \tag{6.7}$$

Furthermore, the center of the correction ellipsoid is given by the corrected state estimate in Eq. (6.2).

Proof. Using Eqs. (6.1), (6.2), and (6.3), we have

$$\begin{aligned}\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} &= (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) - \boldsymbol{L}_k(\boldsymbol{y}_k - \boldsymbol{C}_k\hat{\boldsymbol{x}}_{k|k-1}) \\ &= (\boldsymbol{E}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{C}_k\boldsymbol{E}_{k|k-1})\boldsymbol{z}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{D}_k\boldsymbol{v}_k\end{aligned} \tag{6.8}$$

Next, we define $\boldsymbol{\zeta} = \mathrm{col}[1, \boldsymbol{z}_{k|k-1}, \boldsymbol{v}_k]$. Therefore, Eq. (6.8) can be expressed in terms of $\boldsymbol{\zeta}$ as

$$\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = \boldsymbol{\Pi}_{k|k-1}\boldsymbol{\zeta} \tag{6.9}$$

where $\boldsymbol{\Pi}_{k|k-1}$ is as shown in Eq. (6.7). Now, $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ is given by

$$\boldsymbol{\zeta}^{\mathrm{T}}\Big[\boldsymbol{\Pi}^{\mathrm{T}}_{k|k-1}\boldsymbol{P}^{-1}_{k|k}\boldsymbol{\Pi}_{k|k-1} - \mathrm{diag}(1, \boldsymbol{O}_{\bar{n}}, \boldsymbol{O}_{\bar{v}})\Big]\boldsymbol{\zeta} \leq 0 \tag{6.10}$$

The unknowns in $\boldsymbol{\zeta}$ should satisfy the following inequalities:

$$\begin{cases}\boldsymbol{z}^{\mathrm{T}}_{k|k-1}\boldsymbol{z}_{k|k-1} - 1 \leq 0, \\ \boldsymbol{v}^{\mathrm{T}}_k\boldsymbol{R}^{-1}_k\boldsymbol{v}_k - 1 \leq 0,\end{cases} \tag{6.11}$$

which can be expressed in terms of $\boldsymbol{\zeta}$ as

$$\begin{cases}\boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-1, \boldsymbol{I}_{\bar{n}}, \boldsymbol{O}_{\bar{v}})\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-1, \boldsymbol{O}_{\bar{n}}, \boldsymbol{R}^{-1}_k)\boldsymbol{\zeta} \leq 0.\end{cases} \tag{6.12}$$

Next, the S-procedure (Lemma A.4.2) is applied to the inequalities in Eqs. (6.10) and (6.12). Thus, a sufficient condition such that the inequalities given in Eq. (6.12) imply the inequality in Eq. (6.10) to hold is that there exist $\tau_1 \geq 0, \tau_2 \geq 0$ such that the following is true:

$$\boldsymbol{\Pi}^{\mathrm{T}}_{k|k-1}\boldsymbol{P}^{-1}_{k|k}\boldsymbol{\Pi}_{k|k-1} - \mathrm{diag}(1, \boldsymbol{O}_{\bar{n}}, \boldsymbol{O}_{\bar{v}}) - \tau_1\mathrm{diag}(-1, \boldsymbol{I}_{\bar{n}}, \boldsymbol{O}_{\bar{v}}) - \tau_2\mathrm{diag}(-1, \boldsymbol{O}_{\bar{n}}, \boldsymbol{R}^{-1}_k) \leq 0$$

The above inequality can be expressed in a compact form as

$$\boldsymbol{\Pi}^{\mathrm{T}}_{k|k-1}\boldsymbol{P}^{-1}_{k|k}\boldsymbol{\Pi}_{k|k-1} - \boldsymbol{\Theta}(\tau_1, \tau_2) \leq 0 \tag{6.13}$$

where $\mathbf{\Theta}(\tau_1, \tau_2)$ is as shown in Eq. (6.7). Using the Schur complement (Lemma A.4.3), we express the inequality in Eq. (6.13) equivalently as

$$\begin{bmatrix} -\boldsymbol{P}_{k|k} & \mathbf{\Pi}_{k|k-1} \\ \\ \mathbf{\Pi}_{k|k-1}^{\mathrm{T}} & -\mathbf{\Theta}(\tau_1, \tau_2) \end{bmatrix} \leq 0 \tag{6.14}$$

Solving the inequality in Eq. (6.14) with $\boldsymbol{P}_{k|k} > 0,\ \tau_i \geq 0,\ i = 1, 2$ yields *a correction ellipsoid* containing the state $\boldsymbol{x}_k$. Then, the *optimal correction ellipsoid* is found by minimizing the trace of $\boldsymbol{P}_{k|k}$ subject to $\boldsymbol{P}_{k|k} > 0,\ \tau_i \geq 0,\ i = 1, 2$, and Eq. (6.14). This completes the proof. □

Next, we state the technical result for the prediction step.

Theorem 6.2.2. *Consider the system in Eq.* (6.1) *under the Assumption 6.1.2 with the state $\boldsymbol{x}_k$ in the correction ellipsoid $\mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ and $\boldsymbol{w}_k \in \mathcal{E}(\mathbf{0}_{\bar{w}}, \boldsymbol{Q}_k)$. Then, the successor state $\boldsymbol{x}_{k+1}$ belongs to the optimal prediction ellipsoid $\mathcal{E}(\hat{\boldsymbol{x}}_{k+1|k}, \boldsymbol{P}_{k+1|k})$, if there exist $\boldsymbol{P}_{k+1|k} > 0,\ \tau_i \geq 0,\ i = 3, 4$ as solutions to the following SDP:*

$$\begin{aligned} &\min_{\boldsymbol{P}_{k+1|k}, \tau_3, \tau_4} \quad \mathrm{trace}(\boldsymbol{P}_{k+1|k}) \\ &\text{subject to} \\ &\boldsymbol{P}_{k+1|k} > 0 \\ &\tau_i \geq 0,\ i = 3, 4 \\ &\begin{bmatrix} -\boldsymbol{P}_{k+1|k} & \mathbf{\Pi}_{k|k} \\ \\ \mathbf{\Pi}_{k|k}^{\mathrm{T}} & -\mathbf{\Psi}(\tau_3, \tau_4) \end{bmatrix} \leq 0 \end{aligned} \tag{6.15}$$

where $\mathbf{\Pi}_{k|k}$ and $\mathbf{\Psi}(\tau_3, \tau_4)$ are given by

$$\mathbf{\Pi}_{k|k} = \begin{bmatrix} \mathbf{0}_{\bar{n}} & \boldsymbol{A}_k \boldsymbol{E}_{k|k} & \boldsymbol{G}_k \end{bmatrix},$$

$$\mathbf{\Psi}(\tau_3, \tau_4) = \mathrm{diag}\ (1 - \tau_3 - \tau_4, \tau_3 \boldsymbol{I}_{\bar{n}}, \tau_4 \boldsymbol{Q}_k^{-1})$$

Furthermore, the center of the prediction ellipsoid is given by the predicted state estimate in Eq. (6.4)*.*

Proof. Follows directly from the proof of Theorem 6.2.1 and has been omitted. □

Interior point methods can be implemented to efficiently solve the SDPs in Eqs. (6.6) and (6.15) [169]. The recursive SMF algorithm is summarized in Algorithm 2.

Algorithm 2 The SMF Algorithm

```
1: (Initialization) Select a time-horizon $T_f$. Given the initial values $(\hat{\boldsymbol{x}}_0, \boldsymbol{P}_0)$, set $k = 0$, $\hat{\boldsymbol{x}}_{k|k-1} = \hat{\boldsymbol{x}}_0$, $\boldsymbol{E}_{k|k-1} = \boldsymbol{E}_0$ where $\boldsymbol{P}_0 = \boldsymbol{E}_0\boldsymbol{E}_0^{\mathrm{T}}$.
2: Find $\boldsymbol{P}_{k|k}$ and $\boldsymbol{L}_k$ by solving the SDP in Eq. (6.6).
3: Calculate $\hat{\boldsymbol{x}}_{k|k}$ using Eq. (6.2). Also, calculate $\boldsymbol{E}_{k|k}$ using $\boldsymbol{P}_{k|k} = \boldsymbol{E}_{k|k}\boldsymbol{E}_{k|k}^{\mathrm{T}}$.
4: Solve the SDP in Eq. (6.15) to obtain $\boldsymbol{P}_{k+1|k}$.
5: Calculate $\hat{\boldsymbol{x}}_{k+1|k}$ using Eq. (6.4). Compute $\boldsymbol{E}_{k+1|k}$ using $\boldsymbol{P}_{k+1|k} = \boldsymbol{E}_{k+1|k}\boldsymbol{E}_{k+1|k}^{\mathrm{T}}$.
6: If $k = T_f$ exit. Otherwise, set $k = k + 1$ and go to Step 2.
```

6.3 Leader-Follower Synchronization of Multi-Agent Systems

This section describes local control input synthesis for the leader-follower synchronization. Results presented in this section are based on the results given in [90] and, to be consistent, we have adopted some of the terminologies and notations used in [90].

6.3.1 Graph-Related Preliminaries [75]

Consider a multi-agent system consisting of N agents. The communication topology of the multi-agent system can be represented by a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ where $\mathscr{V} = \{1, 2, \ldots, N\}$ is a nonempty node set and $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ is an edge set of ordered pairs of nodes, called edges. Node i in the graph represents agent i. We consider simple, directed graphs in this chapter. The edge (i, j) in the edge set of a directed graph denotes that node j can obtain information from node i, but not necessarily *vice versa.* If an edge $(i, j) \in \mathscr{E}$, then node i is a neighbor of node j. The set of neighbors of node i is denoted as $\mathscr{N}_i$.

The adjacency matrix $\boldsymbol{\mathscr{A}} = [a_{ij}] \in \mathbb{R}^{N\times N}$ of a directed graph $(\mathscr{V}, \mathscr{E})$ is defined such that a_{ij} is a positive weight if $(j, i) \in \mathscr{E}$, and $a_{ij} = 0$ otherwise. The graph Laplacian matrix $\boldsymbol{\mathscr{L}}$ is defined as $\boldsymbol{\mathscr{L}} = \boldsymbol{\mathscr{D}} - \boldsymbol{\mathscr{A}}$ where $\boldsymbol{\mathscr{D}} = [d_{ij}] \in \mathbb{R}^{N\times N}$ is the in-degree matrix with $d_{ij} = 0, i \neq j$, and $d_{ii} = \sum_{j=1}^{N} a_{ij}, i = 1, 2, \ldots, N$. A directed path is a sequence of edges in a directed graph of the form $(i_1,\, i_2)$, $(i_2,\, i_3)$, $\ldots$. The graph $\mathscr{G}$ contains a (directed) spanning tree if there exists a node, called the root node, such that every other node in $\mathscr{V}$ can be connected by a directed path starting from that node.

6.3.2 Synchronization: Formulation and Results

We consider N agents connected via a directed graph and a leader. Agent i $(i = 1, 2, \ldots, N)$ is a dynamical system of the form

$$\begin{aligned} \boldsymbol{x}_{k+1}^{(i)} &= \boldsymbol{A}\boldsymbol{x}_k^{(i)} + \boldsymbol{B}\boldsymbol{u}_k^{(i)} + \boldsymbol{G}\boldsymbol{w}_k^{(i)}, \\ \boldsymbol{y}_k^{(i)} &= \boldsymbol{C}\boldsymbol{x}_k^{(i)} + \boldsymbol{D}\boldsymbol{v}_k^{(i)}, \quad k \in \mathbb{Z}_\star \end{aligned} \tag{6.16}$$

where $\boldsymbol{x}_k^{(i)} \in \mathbb{R}^n$, $\boldsymbol{u}_k^{(i)} \in \mathbb{R}^m$, $\boldsymbol{y}_k^{(i)} \in \mathbb{R}^p$, $\boldsymbol{w}_k^{(i)} \in \mathbb{R}^w$, $\boldsymbol{v}_k^{(i)} \in \mathbb{R}^v$ are the state, control input, measured output, input and output disturbances for agent i, respectively. Clearly, the system described by Eq. (6.16) is in the form of the system described by Eq. (6.1), with the time-varying matrices replaced by the constant matrices. Next, we modify Assumptions 6.1.1 & 6.1.2 and impose the following assumptions on the dynamics of agent i $(i = 1, 2, \ldots, N)$.

Assumption 6.3.1. *The initial state $\boldsymbol{x}_0^{(i)}$ is unknown. However, it satisfies $\boldsymbol{x}_0^{(i)} \in \mathcal{E}(\hat{\boldsymbol{x}}_0^{(i)}, \boldsymbol{P}_0^{(i)})$ where $\hat{\boldsymbol{x}}_0^{(i)}$ is a given initial estimate and $\boldsymbol{P}_0^{(i)}$ is known. Also, $|\boldsymbol{P}_0^{(i)}| \leq p_0$ holds with some $p_0 > 0$.*

Assumption 6.3.2. *$\boldsymbol{w}_k^{(i)}$ and $\boldsymbol{v}_k^{(i)}$ are unknown-but-bounded for all $k \in \mathbb{Z}_\star$. Also, $\boldsymbol{w}_k^{(i)} \in \mathcal{E}(\mathbf{0}_w, \boldsymbol{Q}_k^{(i)})$ and $\boldsymbol{v}_k^{(i)} \in \mathcal{E}(\mathbf{0}_v, \boldsymbol{R}_k^{(i)})$ for all $k \in \mathbb{Z}_\star$ where $\boldsymbol{Q}_k^{(i)}$, $\boldsymbol{R}_k^{(i)}$ are known with $|\boldsymbol{Q}_k^{(i)}| \leq \bar{q}$ and $|\boldsymbol{R}_k^{(i)}| \leq \bar{r}$ for all $k \in \mathbb{Z}_\star$ with some $\bar{q}, \bar{r} > 0$.*

Under this assumption, agent i ($i = 1, 2, \ldots, N$) employs the SMF in Algorithm 2 to estimate its own state. Now, we introduce the following assumption on the system matrices of the agents.

Assumption 6.3.3. *$\boldsymbol{B}$ is full column rank with the pair $(\boldsymbol{A}, \boldsymbol{B})$ stabilizable.*

We consider the leader to be a system of the form

$$\boldsymbol{x}_{k+1}^{(0)} = \boldsymbol{A}\boldsymbol{x}_k^{(0)}, \ \boldsymbol{y}_k^{(0)} = \boldsymbol{x}_k^{(0)}, \quad k \in \mathbb{Z}_\star \tag{6.17}$$

where $\boldsymbol{x}_k^{(0)} \in \mathbb{R}^n$ is the leader's state and $\boldsymbol{y}_k^{(0)}$ is the output. Note that the leader is a virtual system that generates the reference trajectory for the agents $i = 1, 2, \ldots, N$ to track. We define the local neighborhood tracking errors, using the corrected state estimates of the agents, as

$$\boldsymbol{\epsilon}_k^{(i)} = \sum_{j \in \mathcal{N}_i} a_{ij}(\hat{\boldsymbol{x}}_{k|k}^{(j)} - \hat{\boldsymbol{x}}_{k|k}^{(i)}) + g_i(\boldsymbol{x}_k^{(0)} - \hat{\boldsymbol{x}}_{k|k}^{(i)})$$

where $g_i \geq 0$ are the pinning gains, $\hat{\boldsymbol{x}}_{k|k}^{(i)}$ and $\hat{\boldsymbol{x}}_{k|k}^{(j)}$ are the corrected state estimates of agent i and j, respectively. If agent i is pinned to the leader, we take $g_i > 0$. Now, we choose the control input of agent i as [90]

$$\boldsymbol{u}_k^{(i)} = c(1 + d_{ii} + g_i)^{-1}\boldsymbol{K}\boldsymbol{\epsilon}_k^{(i)}$$

where $c > 0$ is a coupling gain and $\boldsymbol{K}$ is a control gain matrix to be discussed subsequently. Hence, the global dynamics of the N agents can be expressed as

$$\boldsymbol{x}_{k+1}^{(g)} = (\boldsymbol{I}_N \otimes \boldsymbol{A})\boldsymbol{x}_k^{(g)} + \boldsymbol{u}_k^{(g)} + (\boldsymbol{I}_N \otimes \boldsymbol{G})\boldsymbol{w}_k^{(g)}, \quad k \in \mathbb{Z}_\star \tag{6.18}$$

with $\boldsymbol{x}_k^{(g)} = \operatorname{col}[\boldsymbol{x}_k^{(1)}, \ldots, \boldsymbol{x}_k^{(N)}]$, $\boldsymbol{w}_k^{(g)} = \operatorname{col}[\boldsymbol{w}_k^{(1)}, \ldots, \boldsymbol{w}_k^{(N)}]$, and

$$\begin{aligned} \boldsymbol{u}_k^{(g)} = &- c(\boldsymbol{I}_N + \boldsymbol{\mathscr{D}} + \boldsymbol{\mathcal{G}})^{-1}(\boldsymbol{\mathscr{L}} + \boldsymbol{\mathcal{G}}) \otimes \boldsymbol{BK}\hat{\boldsymbol{x}}_{k|k}^{(g)} \\ &+ c(\boldsymbol{I}_N + \boldsymbol{\mathscr{D}} + \boldsymbol{\mathcal{G}})^{-1}(\boldsymbol{\mathscr{L}} + \boldsymbol{\mathcal{G}}) \otimes \boldsymbol{BK}\bar{\boldsymbol{x}}_k^{(0)} \end{aligned} \tag{6.19}$$

where $\boldsymbol{\mathcal{G}} = \operatorname{diag}(g_1, g_2, ..., g_N)$ is the matrix of pinning gains, $\hat{\boldsymbol{x}}_{k|k}^{(g)} = \operatorname{col}[\hat{\boldsymbol{x}}_{k|k}^{(1)}, \ldots, \hat{\boldsymbol{x}}_{k|k}^{(N)}]$, and $\bar{\boldsymbol{x}}_k^{(0)} = \left(\boldsymbol{1}_N \otimes \boldsymbol{x}_k^{(0)}\right)$. Note that the superscript (g) is utilized to denote the global variables. Now, using Eq. (6.5) for each agent's corrected state estimates, we can express $\boldsymbol{x}_k^{(g)}$ as

$$\boldsymbol{x}_k^{(g)} = \hat{\boldsymbol{x}}_{k|k}^{(g)} + \boldsymbol{E}_{k|k}^{(g)}\boldsymbol{z}_{k|k}^{(g)} \tag{6.20}$$

where $\boldsymbol{E}_{k|k}^{(g)} = \operatorname{diag}(\boldsymbol{E}_{k|k}^{(1)}, \ldots, \boldsymbol{E}_{k|k}^{(N)})$, $\boldsymbol{z}_{k|k}^{(g)} = \operatorname{col}[\boldsymbol{z}_{k|k}^{(1)}, \ldots, \boldsymbol{z}_{k|k}^{(N)}]$. Note that $\boldsymbol{E}_{k|k}^{(i)}\left(\boldsymbol{E}_{k|k}^{(i)}\right)^{\mathrm{T}} = \boldsymbol{P}_{k|k}^{(i)}$ where $\boldsymbol{P}_{k|k}^{(i)}$ is the correction ellipsoid shape matrix for agent i and $|\boldsymbol{z}_{k|k}^{(i)}| \leq 1$ for $i = 1, \ldots, N$. Our next assumption is regarding the interaction graph.

Assumption 6.3.4 ([90])**.** *The interaction graph contains a spanning tree with at least one nonzero pinning gain that connects the leader and the root node.*

The global disagreement error [90] is defined as $\boldsymbol{\delta}_k^{(g)} = \boldsymbol{x}_k^{(g)} - \bar{\boldsymbol{x}}_k^{(0)}$. Utilizing Eqs. (6.18)-(6.20), we express the global error system as

$$\boldsymbol{\delta}_{k+1}^{(g)} = \boldsymbol{A}_c\boldsymbol{\delta}_k^{(g)} + \boldsymbol{B}_c\boldsymbol{E}_{k|k}^{(g)}\boldsymbol{z}_{k|k}^{(g)} + (\boldsymbol{I}_N \otimes \boldsymbol{G})\boldsymbol{w}_k^{(g)}, \quad k \in \mathbb{Z}_\star \tag{6.21}$$

where

$$\boldsymbol{A}_c = \left[(\boldsymbol{I}_N \otimes \boldsymbol{A}) - c\boldsymbol{\Gamma} \otimes \boldsymbol{BK}\right], \ \boldsymbol{B}_c = c\boldsymbol{\Gamma} \otimes \boldsymbol{BK} \tag{6.22}$$

with $\boldsymbol{\Gamma} = (\boldsymbol{I}_N + \boldsymbol{\mathscr{D}} + \boldsymbol{\mathcal{G}})^{-1}(\boldsymbol{\mathscr{L}} + \boldsymbol{\mathcal{G}})$. Now, we recall the following technical result from [90].

Lemma 6.3.1 ([90]). *$\rho(\boldsymbol{A}_c) < 1$ iff $\rho(\boldsymbol{A} - c\Lambda_i\boldsymbol{BK}) < 1$ for all the eigenvalues $\Lambda_i, i = 1, 2, \ldots, N$ of $\boldsymbol{\Gamma}$.*

If the matrix $\boldsymbol{A}$ is unstable or marginally stable, then Lemma 6.3.1 requires Assumption 6.3.4 with the pair $(\boldsymbol{A}, \boldsymbol{B})$ stabilizable [90]. Using Theorem 2 in [90], c and $\boldsymbol{K}$ are chosen such that $\rho(\boldsymbol{A}_c) < 1$. To this end, we state the following result.

Lemma 6.3.2 ([90]). *Let Assumption 6.3.4 hold and let $\boldsymbol{\mathcal{P}}$ be a positive definite solution to the discrete-time Riccati-like equation*

$$\boldsymbol{A}^T\boldsymbol{\mathcal{P}}\boldsymbol{A} - \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{Q}} - \boldsymbol{A}^T\boldsymbol{\mathcal{P}}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{A} = \boldsymbol{O}_n \tag{6.23}$$

for some $\boldsymbol{\mathcal{Q}} > 0$. Define

$$r = [\sigma_{max}(\boldsymbol{\mathcal{Q}}^{-0.5}\boldsymbol{A}^T\boldsymbol{\mathcal{P}}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{A}\boldsymbol{\mathcal{Q}}^{-0.5})]^{-0.5}$$

Further, let there exist a $C(c_0, r_0)$ containing all the eigenvalues $\Lambda_i, i = 1, 2, \ldots, N$ of $\boldsymbol{\Gamma}$ such that $(r_0/c_0) < r$. Then, $\rho(\boldsymbol{A}_c) < 1$ for $\boldsymbol{K} = (\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\mathcal{P}}\boldsymbol{A}$ and $c = (1/c_0)$.

If $\boldsymbol{B}$ is full column rank, Eq. (6.23) has a positive definite solution $\boldsymbol{\mathcal{P}}$ only if the pair $(\boldsymbol{A}, \boldsymbol{B})$ is stabilizable [90]. In this regard, Assumption 6.3.3 is pertinent. Now, we state the main result (which involves the notion of input-to-state stability-see Appendix A) of this section in the next theorem.

Theorem 6.3.3. *Suppose the following conditions are satisfied: (i) Under Assumptions 6.3.1 and 6.3.2, agent i ($i = 1, 2, \ldots, N$) employs the SMF in Algorithm 2 to estimate its own state; (ii) Assumptions 6.3.3 and 6.3.4 hold; (iii) c and $\boldsymbol{K}$ are chosen using Lemma 6.3.2. Then, the global error system in Eq.* (6.21) *is input-to-state stable (ISS).*

Proof. The proof is inspired by Example 3.4 in [172]. Let us denote $\boldsymbol{e}_k^{(g)} = \mathrm{col}[\boldsymbol{e}_k^{(1)}, \ldots, \boldsymbol{e}_k^{(N)}]$ where $\boldsymbol{e}_k^{(i)} = \boldsymbol{x}_k^{(i)} - \hat{\boldsymbol{x}}_{k|k}^{(i)}$ are the state estimation errors of agent i at the correction steps. Now, using Eq. (6.20), we have $\boldsymbol{e}_k^{(g)} = \boldsymbol{x}_k^{(g)} - \hat{\boldsymbol{x}}_{k|k}^{(g)} = \boldsymbol{E}_{k|k}^{(g)} \boldsymbol{z}_{k|k}^{(g)}$. Similarly, let us denote $\boldsymbol{e}_0^{(g)} = \mathrm{col}[\boldsymbol{e}_0^{(1)}, \ldots, \boldsymbol{e}_0^{(N)}]$ where $\boldsymbol{e}_0^{(i)} = \boldsymbol{x}_0^{(i)} - \hat{\boldsymbol{x}}_0^{(i)}$ is the initial estimation error of agent i. Due to Assumption 6.3.1, we have

$$\boldsymbol{e}_0^{(g)} = \boldsymbol{E}_0^{(g)} \boldsymbol{z}_0^{(g)} \tag{6.24}$$

with $\boldsymbol{E}_0^{(g)} = \mathrm{diag}(\boldsymbol{E}_0^{(1)}, \ldots, \boldsymbol{E}_0^{(N)})$, $\boldsymbol{z}_0^{(g)} = \mathrm{col}[\boldsymbol{z}_0^{(1)}, \ldots, \boldsymbol{z}_0^{(N)}]$ where $\boldsymbol{E}_0^{(i)} \left(\boldsymbol{E}_0^{(i)}\right)^{\mathrm{T}} = \boldsymbol{P}_0^{(i)}$ and $|\boldsymbol{z}_0^{(i)}| \leq 1$ for $i = 1, \ldots, N$. Then, Eq. (6.21) becomes

$$\boldsymbol{\delta}_{k+1}^{(g)} = \boldsymbol{A}_c^{k+1} \boldsymbol{\delta}_0^{(g)} + \sum_{j=0}^{k} \boldsymbol{A}_c^j \boldsymbol{B}_c \boldsymbol{e}_{k-j}^{(g)} + \sum_{j=0}^{k} \boldsymbol{A}_c^j (\boldsymbol{I}_N \otimes \boldsymbol{G}) \boldsymbol{w}_{k-j}^{(g)}$$

where $\boldsymbol{e}^{(g)} : \mathbb{Z}_\star \to \mathbb{R}^{nN}$, $\boldsymbol{w}^{(g)} : \mathbb{Z}_\star \to \mathbb{R}^{wN}$ are the inputs. It is understood that $\boldsymbol{e}^{(g)} \in l_\infty^{nN}$ and $\boldsymbol{w}^{(g)} \in l_\infty^{wN}$. Due to the choices of c and $\boldsymbol{K}$ along with Assumptions 6.3.3 and 6.3.4, we have $\rho(\boldsymbol{A}_c) < 1$. Hence, there exist constants $\alpha > 0$ and $\mu \in [0, 1)$ such that $|\boldsymbol{A}_c^k| \leq \alpha \mu^k,\ k \in \mathbb{Z}_\star$ [172]. Then, the ISS property in Eq. (A.30) holds for the system in Eq. (6.21) with

$$\begin{aligned} &\beta(s, k) = \alpha \mu^k s,\ \gamma_1(s_1) = \sum_{j=0}^{\infty} \alpha \mu^j |\boldsymbol{B}_c| s_1 = \frac{\alpha |\boldsymbol{B}_c| s_1}{1 - \mu}, \\ &\gamma_2(s_2) = \sum_{j=0}^{\infty} \alpha \mu^j |\boldsymbol{G}| s_2 = \frac{\alpha |\boldsymbol{G}| s_2}{1 - \mu} \end{aligned} \tag{6.25}$$

where $|(\boldsymbol{I}_N \otimes \boldsymbol{G})| = |\boldsymbol{I}_N|\ |\boldsymbol{G}| = |\boldsymbol{G}|$ is utilized. Thus, along the trajectories of the system in Eq. (6.21), for each $k \in \mathbb{Z}_\star$, it holds that

$$|\boldsymbol{\delta}_k^{(g)}| \leq \beta(|\boldsymbol{\delta}_0^{(g)}|, k) + \gamma_1(||\boldsymbol{e}^{(g)}||) + \gamma_2(||\boldsymbol{w}^{(g)}||) \tag{6.26}$$

where the functions $\beta, \gamma_1, \gamma_2$ are as in Eq. (6.25) with $s = |\boldsymbol{\delta}_0^{(g)}|$, $s_1 = ||\boldsymbol{e}^{(g)}||$, $s_2 = ||\boldsymbol{w}^{(g)}||$. This completes the proof. □

Theorem 6.3.3 implies that the global disagreement error remains bounded under the proposed synchronization protocol.

Remark 6.3.4. *Objective of the SMF-based synchronization in [96] was to contain the states of the agents in a confidence ellipsoid which might not be small in general. Thus, the approach outlined in [96] may lead to conservative results where the states of the agents might not converge to a neighborhood of the leader's state trajectory. On the other hand, we have shown that, under appropriate conditions, the global error system is ISS with respect to the input disturbances and estimation errors. Since an ISS system admits the 'converging-input converging-state' property (see, [172, 173] for details), $|\boldsymbol{\delta}_k^{(g)}|$ would eventually converge to a neighborhood of zero as the estimation errors of the agents decrease. Thus, the agents would converge to a neighborhood of the leader's state trajectory. To this end, it is understood that $||\boldsymbol{w}^{(g)}||$ is relatively small (compared to $|\boldsymbol{\delta}_0^{(g)}|$ and $||\boldsymbol{e}^{(g)}||$) as the input disturbances satisfy Assumption 6.3.2.*

Next, we state the following result based on Theorem 6.3.3 where p_0 and $\bar{q}$ are as in Assumptions 6.3.1 and 6.3.2, respectively.

Corollary 6.3.4.1. *Under the conditions of Theorem 6.3.3, the normalized global disagreement error $\bar{\boldsymbol{\delta}}_k^{(g)}$ satisfies*

$$\lim_{k\to\infty} |\bar{\boldsymbol{\delta}}_k^{(g)}| \leq (|\boldsymbol{B}_c|\sqrt{p_0} + |\boldsymbol{G}|\sqrt{\bar{q}}) \tag{6.27}$$

with $\bar{\boldsymbol{\delta}}_k^{(g)} = \left(\boldsymbol{\delta}_k^{(g)}/\bar{\mu}\right)$ where $\bar{\mu} = \left(\alpha\sqrt{N}/(1-\mu)\right)$ and $\alpha > 0$, $\mu \in [0, 1)$ are such that $|\boldsymbol{A}_c^k| \leq \alpha\mu^k$ for all $k \in \mathbb{Z}_\star$.

Proof. Under the conditions of Theorem 6.3.3, the result in Eq. (6.26) holds. Then, let us rewrite Eq. (6.26) as

$$|\boldsymbol{\delta}_k^{(g)}| \leq \alpha\mu^k|\boldsymbol{\delta}_0^{(g)}| + (\alpha/(1-\mu))\left(|\boldsymbol{B}_c|\ ||\boldsymbol{e}^{(g)}|| + |\boldsymbol{G}|\ ||\boldsymbol{w}^{(g)}||\right)$$

Now, under the assumption that the SMFs of the agents are performing adequately, we can utilize Eq. (6.24) and take $||\boldsymbol{e}^{(g)}|| \leq |\boldsymbol{e}_0^{(g)}| \leq |\boldsymbol{E}_0^{(g)}|\ |\boldsymbol{z}_0^{(g)}|$. Using Assumption

6.3.1, we have $|\boldsymbol{E}_0^{(g)}| = \max(|\boldsymbol{E}_0^{(1)}|, \ldots, |\boldsymbol{E}_0^{(N)}|) \Rightarrow |\boldsymbol{E}_0^{(g)}| \leq \sqrt{p_0}$. Also, we have $|\boldsymbol{z}_0^{(g)}| \leq \sqrt{N}$. Therefore, we derive $||\boldsymbol{e}^{(g)}|| \leq \sqrt{p_0 N}$. Similarly, Assumption 6.3.2 leads to $||\boldsymbol{w}^{(g)}|| \leq \sqrt{\bar{q}N}$. Combining these, we calculate the following bound on $\boldsymbol{\delta}_k^{(g)}$

$$|\boldsymbol{\delta}_k^{(g)}| \leq \alpha\mu^k |\boldsymbol{\delta}_0^{(g)}| + \bar{\mu}\left(|\boldsymbol{B}_c|\sqrt{p_0} + |\boldsymbol{G}|\sqrt{\bar{q}}\right) \tag{6.28}$$

for each $k \in \mathbb{Z}_\star$ with $\bar{\mu} = \Big(\alpha\sqrt{N}/(1-\mu)\Big)$. Hence, the proof is completed by taking the limit in Eq. (6.27) and carrying out the normalization $\bar{\boldsymbol{\delta}}_k^{(g)} = \Big(\boldsymbol{\delta}_k^{(g)}/\bar{\mu}\Big)$. □

Remark 6.3.5. *The upper bound shown in Eq.* (6.28) *is monotonically decreasing. The estimate given in Eq.* (6.27) *is a conservative one as we have utilized* $||\boldsymbol{e}^{(g)}|| \leq \sqrt{p_0 N}$ *and* $||\boldsymbol{w}^{(g)}|| \leq \sqrt{\bar{q}N}$. *Also, the bound* $|\boldsymbol{R}_k^{(i)}| \leq \bar{r}$ *does not appear in Eqs.* (6.27) *and* (6.28) *as a result of utilizing* $||\boldsymbol{e}^{(g)}|| \leq |\boldsymbol{e}_0^{(g)}| \leq |\boldsymbol{E}_0^{(g)}|\ |\boldsymbol{z}_0^{(g)}|$. *However, the true value of* $\boldsymbol{e}_k^{(i)}$ *would depend on* $\boldsymbol{v}_k^{(i)}$ *and, thus, on* $\boldsymbol{R}_k^{(i)}$ *for all* $k \in \mathbb{Z}_\star$ *and all* $i = 1, 2, \ldots, N$.

Remark 6.3.6. *For a given multi-agent system (with the number of agents N, the matrices* $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{G}$, *and the interaction graph specified), we have* $|\boldsymbol{B}_c|$ *and* $|\boldsymbol{G}|$ *fixed once c and* $\boldsymbol{K}$ *are properly chosen using Lemma 6.3.2. Thus, the conservatism of the bound in Eq.* (6.27) *can be reduced if the available upper bounds (i.e.,* p_0 *and* $\bar{q}$*) are sufficiently small.*

6.4 Simulation Examples

Simulation examples are provided in this section to illustrate the effectiveness of the proposed SMF and SMF-based leader-follower synchronization protocol. All the simulations are carried out on a desktop computer with a 16.00 GB RAM and a 3.40 GHz Intel(R) Xeon(R) E-2124 G processor running MATLAB R2019a. The SDPs in Eqs. (6.6) and (6.15) are solved utilizing 'YALMIP' [170] with the 'SDPT3' solver in the MATLAB framework. Since the disturbances are only assumed to be unknown-but-bounded, different kinds of disturbance realizations are possible that satisfy the ellipsoidal assumptions (Assumptions 6.1.2 and 6.3.2). For example, periodic disturbances with time-varying or constant frequencies and amplitudes, random disturbances with each element being uniformly distributed in an interval, and so on.

6.4.1 Example-1

In this example, we illustrate the effectiveness of the proposed SMF algorithm and compare our results with the results obtained for the SMF in [174] (the discrete version). We choose a system governed by the Mathieu equation [174] for this example, i.e., the system given by

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= -\omega_0^2(1+\epsilon\sin\omega t)x_1 + w_d\end{aligned} \tag{6.29}$$

which is expressed in a compact form as

$$\dot{\boldsymbol{x}} = \begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{G}w_d \tag{6.30}$$

where

$$\boldsymbol{A}(t) = \begin{bmatrix} 0 & 1\\ -\omega_0^2(1+\epsilon\sin\omega t) & 0\end{bmatrix}, \quad \boldsymbol{G} = \begin{bmatrix}0\\1\end{bmatrix} \tag{6.31}$$

with w_d as the input disturbance. Utilizing zero-order hold (ZOH) with a sampling time Δt, the above system is put into an equivalent discrete-time form as

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}_k\boldsymbol{x}_k + \boldsymbol{G}_k w_k \tag{6.32}$$

where $\boldsymbol{x}_{(\cdot)} = [x_{1_{(\cdot)}} \quad x_{2_{(\cdot)}}]^{\mathrm{T}}$ and w_k is the input disturbance at the current time-step. The measured outputs are considered as $y_k = x_{1_k} + v_k$. Thus, we have $\boldsymbol{C}_k =$

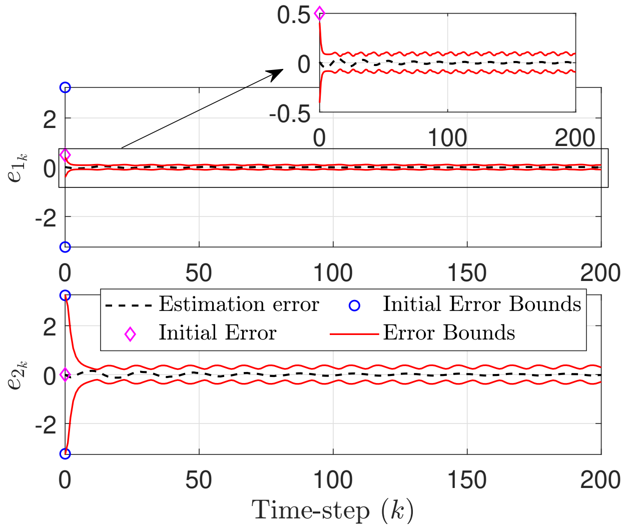


Figure 6.1: Estimation errors and corresponding error bounds for the proposed SMF (Example-1).

$[1 \quad 0]$, $\boldsymbol{D}_k = 1$. The system parameters, Δt, disturbances, disturbance ellipsoid shape matrices, and initial conditions chosen are as follows:

$$\begin{aligned} &\omega = 2\pi,\ \omega_0 = \pi,\ \epsilon = 0.3,\ \Delta t = 0.1 \text{ seconds} \\ &w_k = 0.05\sin(\omega t_k),\ v_k = w_k,\ \boldsymbol{Q}_k = 0.0025,\ \boldsymbol{R}_k = \boldsymbol{Q}_k, \\ &\boldsymbol{x}_0 = [0.5 \quad 0]^{\mathrm{T}},\ \hat{\boldsymbol{x}}_0 = \boldsymbol{0}_2,\ \boldsymbol{P}_0 = 10.5\boldsymbol{I}_2 \end{aligned} \tag{6.33}$$

With the above initial conditions, disturbances and disturbance ellipsoid shape matrices, Assumptions 6.1.1 and 6.1.2 are satisfied, and we implement the proposed SMF in Algorithm 2 with $T_f = 200$. The simulation results are shown in Figs. 6.1 and 6.2. The estimation errors and error bounds shown in Fig. 6.1 are corresponding to the correction steps. Thus, we have $\boldsymbol{e}_k = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} = [e_{1_k} \quad e_{2_k}]^{\mathrm{T}}$. As shown in Fig. 6.1, the estimation errors remain within the corresponding error bounds for the entire time-horizon considered. Thus, the true state is contained in the correction ellipsoids for the entire time-horizon too. The error bounds are time-varying for this example as the dynamical system considered here is time-varying. Further, the error bounds decrease significantly from the corresponding initial values, evidencing the ongoing optimization process for the SMF. The phase plane plot of the true states and corrected state estimates are shown in Fig. 6.2. Since the estimation errors are small (as shown in Fig. 6.1), the true state and corrected state estimate trajectories remain in a close neighborhood of each other. This is depicted in Fig. 6.2. Also, the zoomed-in plot in Fig. 6.2 shows that the SMF is able to bring the corrected state estimate in a neighborhood of the true state after the correction step at $k = 0$. This explains the small e_{1_k} at $k = 0$ compared to the initial error (see the zoomed-in plot in Fig. 6.1).

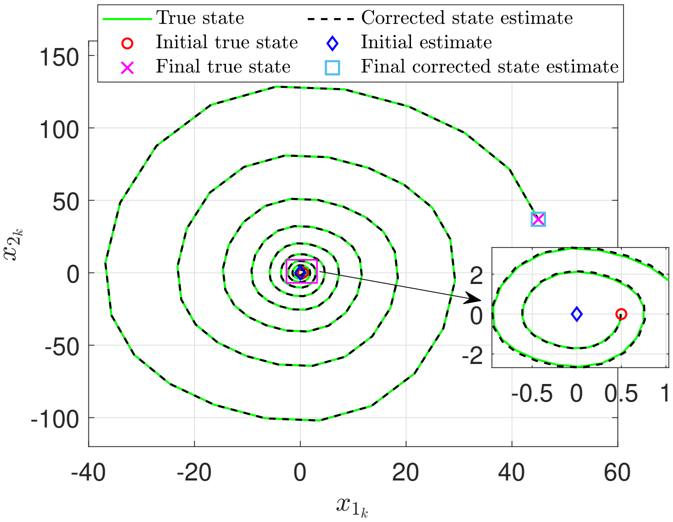


Figure 6.2: The true state and corrected state estimate trajectories in the phase plane (Example-1).

Table 6.1: Estimation error comparisons over $T = 201$ time-steps ($T_f = 200$)

Item	Proposed SMF	Balandin et al. ([174])
$\frac{1}{T}\sum \lvert \boldsymbol{e}_k \rvert$	0.0434	0.0477
$\frac{1}{T}\sum e_{1_k}^2$	0.0002	0.0015
$\frac{1}{T}\sum e_{2_k}^2$	0.00267	0.0028

Next, we compare the results of the proposed SMF with the ones corresponding to the SMF framework given in Balandin et al. ([174], the discrete version). The system parameters, Δt, disturbances, and initial conditons considered for both the frameworks are as shown in Eq. (6.33). However, the disturbance ellipsoid shape matrices for the framework in [174] are adopted from the example in section 3 in [174]. Whereas, for the proposed SMF, the disturbance ellipsoid shape matrices are as shown in Eq. (6.33). The difference in the disturbance ellipsoid shape matrices are due to the different kinds of ellipsoidal assumptions utilized in this chapter, compared to the ones in [174]. However, these values are chosen such that both sets of disturbance ellipsoid shape matrices are equivalent. The comparison results are shown in Figs. 6.3, 6.4. Fig. 6.3 shows the comparison in estimation error norms where estimation errors at the correction steps for the proposed SMF are depicted. The proposed SMF is able to reduce the error norm at $k = 0$ due to the initial correction step (see the zoomed-in plot in Fig. 6.3). After that, both the SMFs have qualitatively similar error norms. Table 6.1 illustrates quantitative comparisons of the estimation errors. Clearly, the proposed SMF outperforms the SMF in [174] in terms of the mean absolute error ($\frac{1}{T}\sum|\boldsymbol{e}_k|$) and mean squared errors ($\frac{1}{T}\sum e_{1_k}^2$, $\frac{1}{T}\sum e_{2_k}^2$).

The trace of the shape matrices, corresponding to the correction ellispoids of the proposed SMF and the estimation ellipsoids of the SMF in [174], are shown in Fig. 6.4. These results show that overall the correction ellipsoids of the proposed SMF are smaller in 'size' compared to the estimation ellipsoids of the SMF in [174]. Thus, the error bounds shown in Fig. 6.1 for the proposed SMF are tighter compared to the ones for the SMF in [174]. Also, the proposed SMF is able to reduce the 'size' of the correction ellipsoid at $k = 0$ due to the initial correction step, as shown in the zoomed-in plot in Fig. 6.4. Note that the proposed SMF employs a two-step filtering approach wherein two SDPs are solved during every filter recursion which results

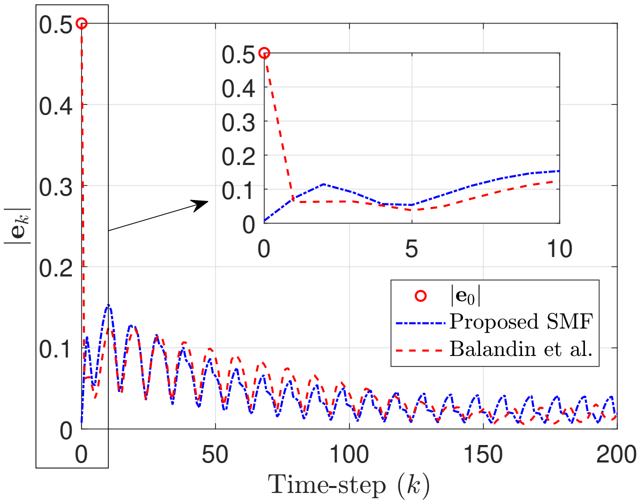


Figure 6.3: Comparison of the estimation error norms where $|\boldsymbol{e}_0| = |\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0|$ (Example-1).

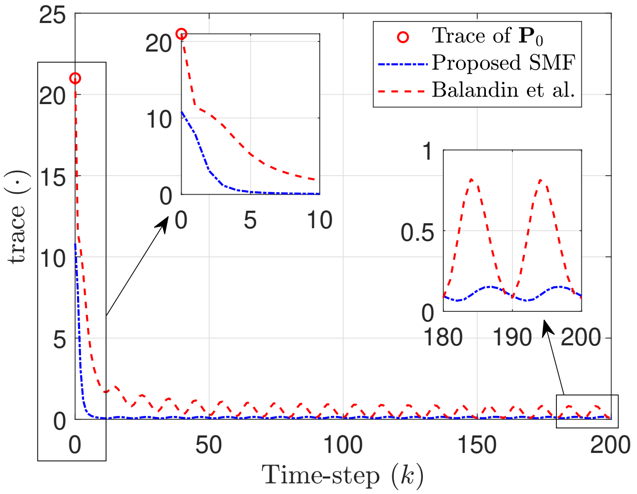


Figure 6.4: Comparison of the trace of the ellipsoid shape matrices (Example-1).

in optimal (minimum trace) correction and prediction ellipsoids (with the respective state estimates at the centers). On the other hand, the SMF in [174] employs a one-step filtering technique with a combined correction and prediction step. Thus, the optimization process for estimation happens only once during each recursion of the

SMF in [174]. This is likely the reason for better overall performance of the proposed SMF in this example.

6.4.2 Example-2

In this example, we illustrate results of the proposed SMF-based leader-follower synchronization protocol. We consider four agents, i.e., $N = 4$. Matrices related to the dynamics of the leader and the agents are

$$\boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \ \boldsymbol{B} = \boldsymbol{I}_2, \ \boldsymbol{C} = [1 \quad 0], \ \boldsymbol{D} = 1, \ \boldsymbol{G} = \boldsymbol{I}_2 \tag{6.34}$$

where $\boldsymbol{A}$ is marginally stable. Also, Assumption 6.3.3 is satisfied with the above choices of $\boldsymbol{A}$ and $\boldsymbol{B}$. Ellipsoidal parameters related to the SMFs of the agents are $\boldsymbol{P}_0^{(i)} = 2\boldsymbol{I}_2, \boldsymbol{Q}_k^{(i)} = 0.1\boldsymbol{I}_2, \ \boldsymbol{R}_k^{(i)} = 0.1$ for $i = 1, 2, 3, 4$. The initial state estimates of the agents are as follows: $\hat{\boldsymbol{x}}_0^{(1)} = [50 \quad -50]^{\mathrm{T}}$, $\hat{\boldsymbol{x}}_0^{(2)} = \hat{\boldsymbol{x}}_0^{(1)}$, $\hat{\boldsymbol{x}}_0^{(3)} = [-50 \quad 50]^{\mathrm{T}}$, $\hat{\boldsymbol{x}}_0^{(4)} = \hat{\boldsymbol{x}}_0^{(3)}$. The true initial state for the agents 1 and 2 ($\boldsymbol{x}_0^{(1)}$, $\boldsymbol{x}_0^{(2)}$) are chosen randomly (uniform distribution) between $[50 \quad -50]^{\mathrm{T}}$ and $[51 \quad -49]^{\mathrm{T}}$. Similarly, the true initial

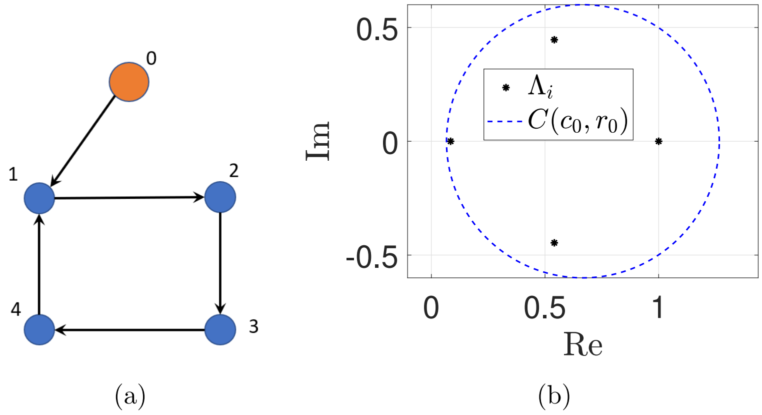


Figure 6.5: (a) The interaction graph and (b) eigenvalues of $\boldsymbol{\Gamma}$ ($\Lambda_i, \ i = 1, 2, 3, 4$) in the complex plane with $C(c_0, r_0)$ (Example-2).

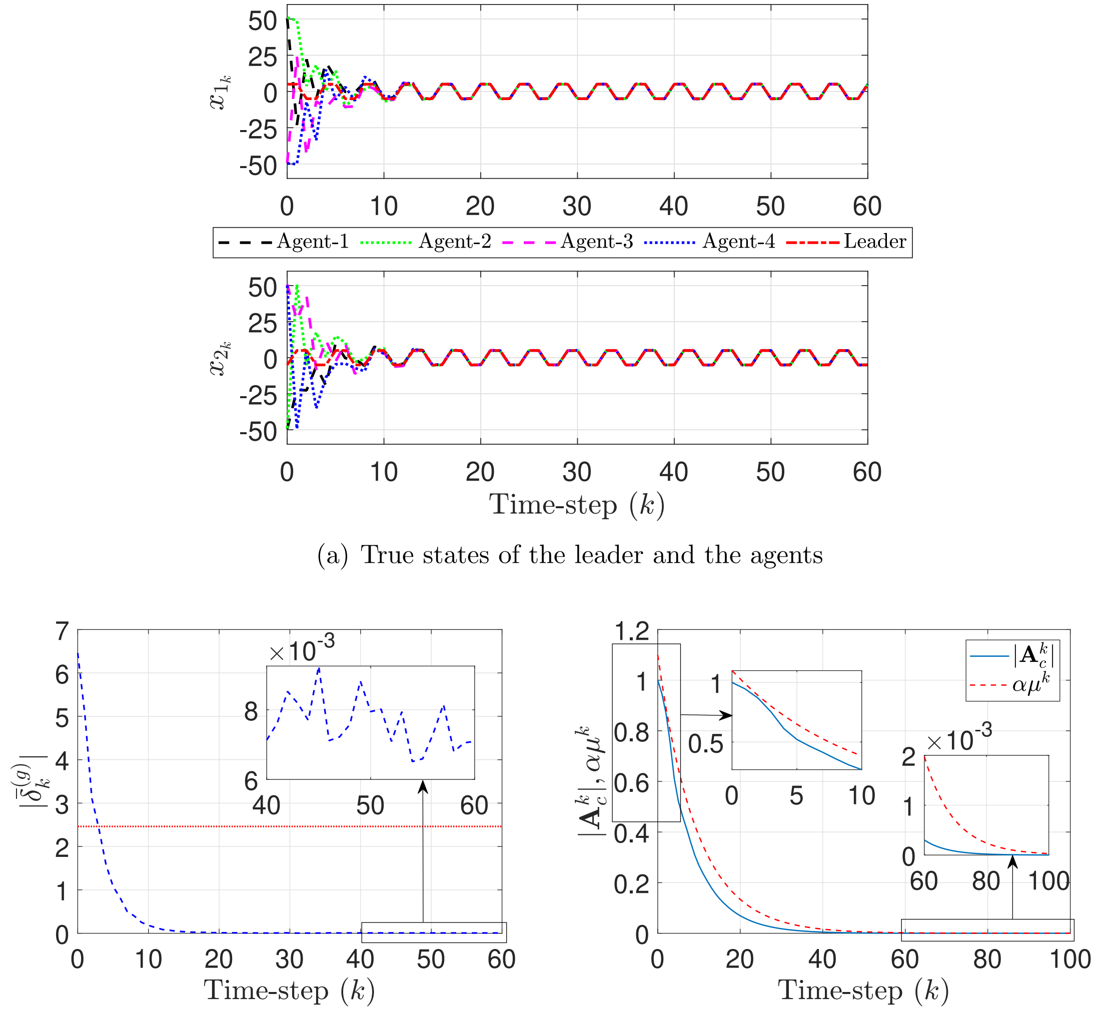
state for the agents 3 and 4 ($\boldsymbol{x}_0^{(3)}$, $\boldsymbol{x}_0^{(4)}$) are chosen randomly (uniform distribution) between $[-50 \quad 50]^{\mathrm{T}}$ and $[-49 \quad 51]^{\mathrm{T}}$. The input disturbances ($\boldsymbol{w}_k^{(i)}$, $i = 1, 2, 3, 4$) are chosen randomly (uniform distribution) between $-0.05\mathbf{1}_2$ and $0.05\mathbf{1}_2$, and the output disturbances ($\boldsymbol{v}_k^{(i)}$, $i = 1, 2, 3, 4$) are chosen randomly (uniform distribution) between -0.05 and 0.05. Thus, Assumptions 6.3.1 and 6.3.2 have been satisfied with the above parameters, initial conditions, and disturbance terms. The initial state of the leader is chosen as $\boldsymbol{x}_0^{(0)} = [5 \ -5]^{\mathrm{T}}$.

The interaction graph is shown in Fig. 6.5(a) and Assumption 6.3.4 holds for this interaction graph. Thus, we have

$$\boldsymbol{\mathscr{L}} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \tag{6.35}$$

$\boldsymbol{\mathcal{G}} = \mathrm{diag}(1, 0, 0, 0), \boldsymbol{\mathscr{D}} = \mathrm{diag}(1, 1, 1, 1)$. With regards to Lemma 6.3.2, we choose $\boldsymbol{\mathcal{Q}} = 0.1\boldsymbol{I}_2$, $c_0 = (2/3)$, $r_0 = 0.6$. Clearly, $C(c_0, r_0)$ contains all the eigenvalues of $\boldsymbol{\Gamma}$ as shown in Fig. 6.5(b). Also, we have $r = 1$ and $(r_0/c_0) = 0.9 < r$. Hence, the conditions for Lemma 6.3.2 are satisfied and we take $c = (1/c_0) = 1.5$, $\boldsymbol{K} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\mathcal{P}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\mathcal{P}}\boldsymbol{A}$.

The synchronization results are shown in Figs. 6.6(a) and 6.6(b). Figure 6.6(a) shows that the trajectories of the agents converge close to that of the leader. As a result, the normalized global disagreement error norm converges to a neighborhood of zero (Fig. 6.6(b)). The red dotted line in Fig. 6.6(b) denotes the conservative bound in Eq. (6.27) for which we have utilized $p_0 = 2$ and $\bar{q} = 0.1$. Also, for $\bar{\mu}$, we have taken $\alpha = 1.1$ and $\mu = 0.9$. For this choice of α and μ, $|\boldsymbol{A}_c^k| \leq \alpha\mu^k$ is satisfied, as



(a) True states of the leader and the agents

(b) Normalized global disagreement error norm

(c) $|\boldsymbol{A}_c^k|$ and the upper bound

Figure 6.6: Simulation results for the Example-2.

shown in Fig. 6.6(c). With the above values, the conservative upper bound is equal to 2.462, which is shown using the red dotted line in Fig. 6.6(b).

The estimation results corresponding to the SMFs of the agents are shown in Figs. 6.7, 6.8, 6.9, 6.10. The estimation errors (at the correction steps) for agent i's SMF are denoted by $\boldsymbol{e}_k^{(i)} = [e_{1_k}^{(i)} \quad e_{2_k}^{(i)}]^{\mathrm{T}}$ and the initial errors are denoted by $\boldsymbol{e}_0^{(i)} = \boldsymbol{x}_0^{(i)} - \hat{\boldsymbol{x}}_0^{(i)}$ $(i = 1, 2, 3, 4)$. Figs. 6.7, 6.8 show that the SMFs for the agents

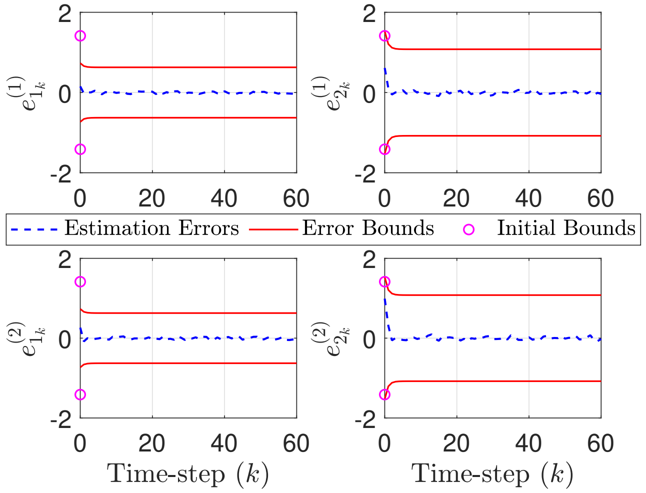


Figure 6.7: Estimation results for SMFs of agents 1 and 2 (Example-2).

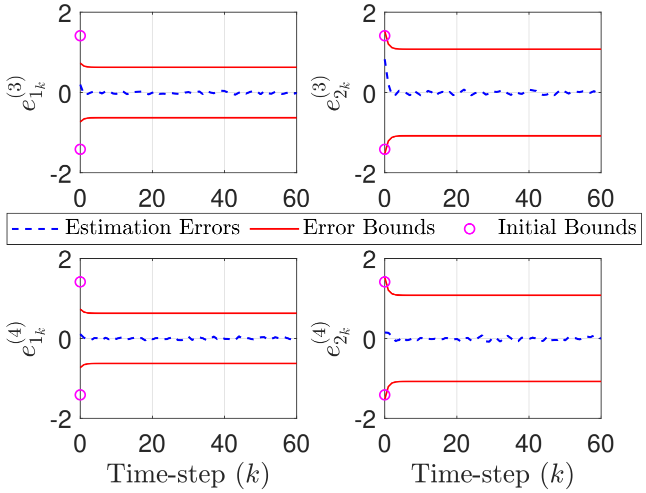


Figure 6.8: Estimation results for SMFs of agents 3 and 4 (Example-2).

perform adequately as the estimation errors remain in a neighborhood of zero and the error bounds decrease from the respective initial values. Also, the estimation errors are contained within the error bounds which mean the SMFs of the agents are able to contain the respective true states inside the respective correction ellipsoids. The estimation error norms, shown in Fig. 6.9, further illustrate the effectiveness of the SMFs and demonstrate that the SMFs are able to reduce the estimation errors from the respective initial values, starting from the correction step at $k = 0$. The results in Fig. 6.9 essentially verify our earlier use of $||\boldsymbol{e}^{(g)}|| \leq |\boldsymbol{e}_0^{(g)}|$ in deriving the conservative bound in Eq. (6.27).

The trace of correction ellipsoid shape matrices for the SMFs of the agents are shown in Fig. 6.10 where $\boldsymbol{P}_{k|k}^{(i)}$ $(i = 1, 2, 3, 4)$ denote the shape matrices of agent i's correction ellipsoids. Clearly, SMFs of the agents are able to reduce the trace from the initial values and construct optimal (minimum trace) correction ellipsoids at each time-step (starting from $k = 0$). Quantitatively, the trace of these shape matrices converge approximately to 1.5 (see Fig. 6.10), which is approximately a 2.667-fold decrease with respect to the initial trace of 4. The trends shown in Fig. 6.10 for all the agents are roughly the same as the same set of ellipsoidal parameters is utilized for the SMFs of all the agents and the agents have identical dynamics.

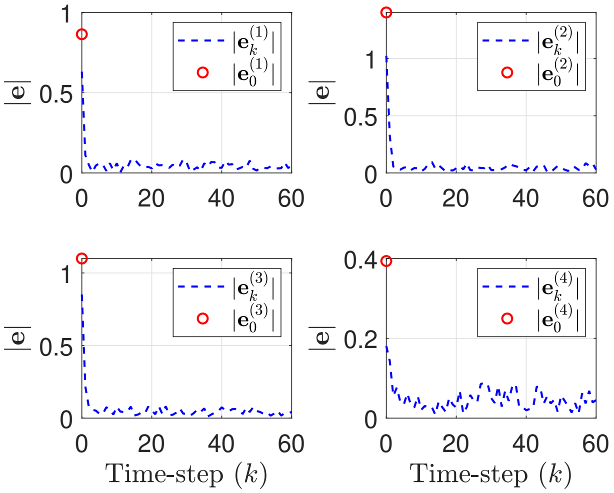


Figure 6.9: Estimation error norms for SMFs of the agents (Example-2).

Finally, consider this example with different values of $\boldsymbol{w}_k^{(i)}$, $\boldsymbol{v}_k^{(i)}$, $\boldsymbol{Q}_k^{(i)}$, $\boldsymbol{R}_k^{(i)}$ $(i = 1, 2, 3, 4)$ while keeping all other conditions and parameters unchanged. Now,

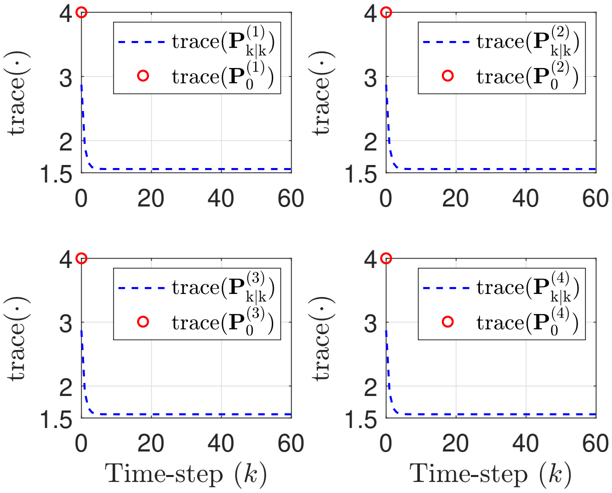


Figure 6.10: Trace of correction ellipsoid shape matrices for SMFs of the agents (Example-2).

let us allow for higher magnitudes of disturbances (with $\boldsymbol{Q}_k^{(i)}$, $\boldsymbol{R}_k^{(i)}$ properly chosen such that Assumption 6.3.2 is satisfied) and compare $|\bar{\boldsymbol{\delta}}_k|$ results with the one given in Fig. 6.6(b). Results of this study are given in Table 6.2 where the following two comparison metrics are used (with $T = T_f + 1$): (i) $\frac{1}{T}\sum_{k=0}^{T_f} |\bar{\boldsymbol{\delta}}_k|$: mean value of $|\bar{\boldsymbol{\delta}}_k|$; (ii) $\sqrt{\frac{1}{T}\sum_{k=0}^{T_f} |\bar{\boldsymbol{\delta}}_k|^2}$: root mean square value of $|\bar{\boldsymbol{\delta}}_k|$. Also, $\boldsymbol{w}_k^{(i)}$ are chosen randomly (uniform distribution) between $-\alpha_w \mathbf{1}_2$ and $\alpha_w \mathbf{1}_2$, and $\boldsymbol{v}_k^{(i)}$ are chosen randomly (uniform distribution) between $-\alpha_v$ and α_v. Thus, the first row in Table 6.2 corresponds to the result in Fig. 6.6(b). We observe that both the metrics in Table 6.2 are com-

Table 6.2: $|\bar{\boldsymbol{\delta}}_k|$ comparisons over $T = 61$ time-steps ($T_f = 60$)

Disturbance parameters	$\frac{1}{T}\sum_{k=0}^{T_f} \lvert\bar{\boldsymbol{\delta}}_k\rvert$	$\sqrt{\frac{1}{T}\sum_{k=0}^{T_f} \lvert\bar{\boldsymbol{\delta}}_k\rvert^2}$
$\alpha_w = \alpha_v = 0.05$, $\boldsymbol{Q}_k^{(i)} = 0.1\boldsymbol{I}_2$, $\boldsymbol{R}_k^{(i)} = 0.1$	0.3706	1.1985
$\alpha_w = \alpha_v = 0.5$, $\boldsymbol{Q}_k^{(i)} = \boldsymbol{I}_2$, $\boldsymbol{R}_k^{(i)} = 1$	0.4219	1.2052
$\alpha_w = \alpha_v = 1$, $\boldsymbol{Q}_k^{(i)} = 2\boldsymbol{I}_2$, $\boldsymbol{R}_k^{(i)} = 1$	0.4730	1.2124

parable among the three cases studied, despite the higher magnitudes of disturbances considered for the two cases in second and third rows of Table 6.2. Therefore, the $|\bar{\boldsymbol{\delta}}_k|$ trends for these two cases with higher disturbance magnitudes would be qualitatively similar to the one shown in Fig. 6.6(b).

6.5 Chapter Summary

A set-membership filtering-based leader-follower synchronization protocol for high-order discrete-time linear multi-agent systems has been put forward for which the global error system is shown to be input-to-state stable with respect to the input disturbances and estimation errors. A monotonically decreasing upper bound on the norm of the global disagreement error vector is calculated.

Chapter 7

Nonlinear Model Predictive Control and Collision Cone-Based Missile Guidance Algorithm[*]

In this chapter, a hybrid missile guidance algorithm is proposed that comprises two components: a point mass-based NMPC (referred to as PM-NMPC) and a collision cone-based NMPC (referred to as CC-NMPC). The PM-NMPC is utilized during that phase of the engagement when the distance between the missile and the target is large, and considers both the missile and the target to be point objects. The CC-NMPC is utilized during that phase of the engagement when the distance between the missile and the target is comparatively smaller, and takes into account the fact that the missile is carrying a warhead with a non-zero blast radius. The CC-NMPC models the engagement as that between a point object missile and a circular target. By combining these two approaches, the guidance algorithm is able to achieve interception with different impact angles.

The rest of the chapter is organized as follows. The planar kinematics of missile target engagement is described in Section 7.1. Also, Section 7.1 reviews the collision cone concept and then presents several considerations on the impact angle, from a collision cone viewpoint. Section 7.2 details the guidance algorithm formulation.

[*] The materials of this chapter have been published in D. Bhattacharjee, A. Chakravarthy, and K. Subbarao, "Nonlinear Model Predictive Control and Collision Cone-Based Missile Guidance Algorithm," *AIAA Journal of Guidance, Control, and Dynamics*, pp. 1-17, 2021, DOI: 10.2514/1.G005879 (reference [132]), and these are reproduced here with the permission of the co-authors.

Section 7.3 provides simulation results that illustrate the effectiveness of the proposed algorithm. Finally, Section 7.4 summarizes the findings of this chapter.

7.1 Engagement Kinematics, Collision Cone, and Impact Angle Magnitude

In this section, we discuss the governing kinematic equations, the collision cone approach, and the impact angle-related considerations.

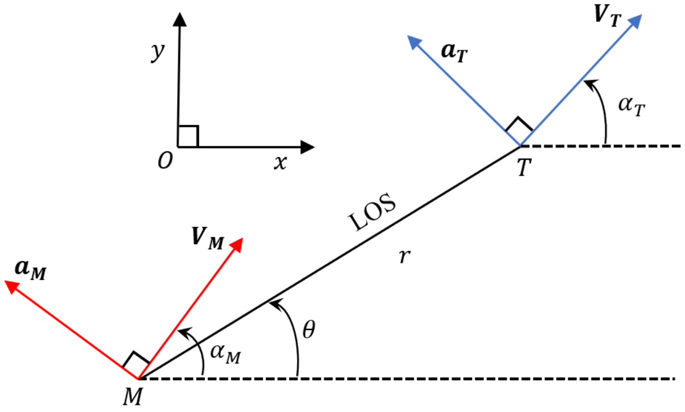


Figure 7.1: A schematic of the planar engagement geometry between the missile (M) and the target (T).

7.1.1 Engagement Kinematics

We assume a planar engagement scenario between the missile and the target. A schematic of the engagement geometry is shown in Fig. 7.1 where $x - O - y$ is an inertial frame of reference. The states in the relative frame of reference (whose origin is at the missile) are the range and bearing angle to the target (r and θ, respectively), the relative velocity components along and perpendicular to the LOS (V_r and V_θ, respectively). These relative velocity components are given by (see Fig. 7.1):

$$V_r = V_T \cos(\alpha_T - \theta) - V_M \cos(\alpha_M - \theta) \tag{7.1}$$

$$V_\theta = V_T \sin(\alpha_T - \theta) - V_M \sin(\alpha_M - \theta) \tag{7.2}$$

where V_M and V_T represent the missile and target speeds, respectively, while α_M and α_T represent their heading angles. The quantities a_M and a_T represent the missile and target accelerations, respectively, and these are assumed to act perpendicular to their respective velocity vectors. As a result, both V_M and V_T remain constant during the engagement. The kinematics of the engagement can be expressed in the following compact form:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})u + \boldsymbol{d}(t) \tag{7.3}$$

where $u = a_M \in \mathbb{R}$ is the control input (or latax), $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5) = (r, V_r, \theta, V_\theta, \alpha_M) \in \mathbb{R}^n$ $(n = 5)$ is the state vector, and $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^n$, $\boldsymbol{d} : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ are as follows:

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{x}) &= \begin{bmatrix} V_r & \frac{V_\theta^2}{r} & \frac{V_\theta}{r} & -\frac{V_r V_\theta}{r} & 0 \end{bmatrix}^{\mathrm{T}}, \\
\boldsymbol{g}(\boldsymbol{x}) &= \begin{bmatrix} 0 & \sin(\alpha_M - \theta) & 0 & -\cos(\alpha_M - \theta) & (1/V_M) \end{bmatrix}^{\mathrm{T}}, \\
\boldsymbol{d}(t) &= \begin{bmatrix} 0 & a_{T_r} & 0 & a_{T\theta} & 0 \end{bmatrix}^{\mathrm{T}}.
\end{aligned} \tag{7.4}$$

In the above, a_{T_r} and a_{T_θ} represent, respectively, the target acceleration components along the LOS, and perpendicular to the LOS. These are regarded as unknown bounded disturbances. We would like to mention that the terms latax and control input are used interchangeably, both referring to the lateral acceleration of the missile.

MPC typically requires the governing differential equations to be expressed in a discrete-time form. Accordingly, the kinematic equations (7.3) are expressed in the following equivalent discrete-time form using Euler discretization:

$$\boldsymbol{x}(k+1) = \boldsymbol{f}_d(\boldsymbol{x}(k)) + \boldsymbol{g}_d(\boldsymbol{x}(k))u(k) + \boldsymbol{d}_d(k) \tag{7.5}$$

where $\boldsymbol{f}_d(\boldsymbol{x}(k))$, $\boldsymbol{g}_d(\boldsymbol{x}(k))$, and $\boldsymbol{d}_d(k)$ are as follows:

$$\begin{gathered}
\boldsymbol{f}_d(\boldsymbol{x}(k)) = \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \\ x_5(k) \end{bmatrix} + \Delta t \begin{bmatrix} x_2(k) \\ x_4^2(k)/x_1(k) \\ x_4(k)/x_1(k) \\ -x_2(k)x_4(k)/x_1(k) \\ 0 \end{bmatrix}, \\
\boldsymbol{g}_d(\boldsymbol{x}(k)) = \Delta t \begin{bmatrix} 0 \\ \sin(x_5(k) - x_3(k)) \\ 0 \\ -\cos(x_5(k) - x_3(k)) \\ (1/V_M) \end{bmatrix}, \quad \boldsymbol{d}_d(k) = \Delta t \begin{bmatrix} 0 \\ a_{T_r}(k) \\ 0 \\ a_{T\theta}(k) \\ 0 \end{bmatrix}
\end{gathered} \tag{7.6}$$

with Δt as the discretization time-step and $\boldsymbol{x}(k) = (x_1(k), x_2(k), x_3(k), x_4(k), x_5(k)) = (r(k), V_r(k), \theta(k), V_\theta(k), \alpha_M(k))$.

7.1.2 Collision Cone and Impact Angle Magnitude

As stated earlier, when the missile is far from the target, the guidance algorithm employs the PM-NMPC formulation. When the missile gets sufficiently close to the target, the guidance algorithm switches from PM-NMPC to CC-NMPC. This switch occurs when the distance between the missile and the target falls below a threshold R_{switch}, that is, the CC-NMPC is engaged once $r(k) \leq R_{\text{switch}}$. The guidance objective during the CC-NMPC phase is to drive the missile inside a circle whose center is located at the target, and furthermore, to arrive at this circle with the impact angle lying within a pre-specified range. The radius of the circle is equal to the blast radius of the warhead carried by the missile. We refer to this circle as the blast circle.

A metric of the collision cone between a point object and a circular object is utilized. In discrete-time form, this is given by ([124, 126])

$$\xi(k) = \frac{r^2(k)V_\theta^2(k)}{V_\theta^2(k) + V_r^2(k)} - R_{\text{blast}}^2 \tag{7.7}$$

where R_{blast} is the blast radius. In this chapter, this function is referred to as the collision cone function. By satisfying the conditions $\xi(k) < 0, V_r(k) < 0$, it can be ensured that the velocity vector of the missile lies inside the collision cone to the blast circle. This is based on the idea that if the value of the predicted miss-distance between the missile and the target is less than R_{blast}, then the missile is on a course to intercept the blast circle. Note that interception or impact time in this chapter is to be understood as the time instant when the missile first intercepts the blast circle, that is, the earliest time instant at which $r(k) \leq R_{\text{blast}}$ occurs. We represent this time-step by k_f and denote $\xi(k_f) = \xi_f$.

We note that by choosing different values of ξ_f, we can cause the missile to intercept the circle at different intercept points. Different intercept points on the circle, in turn, lead to different impact angles. To visualize this, refer to Fig. 7.2 which shows an engagement between a point object and a moving circle. The trajectory of the point object for several intercept points on the circle is shown in Fig. 7.2(a). The differences between the heading angles of the point object and the circle are then shown in Fig. 7.2(b). The impact angle is evident from Fig. 7.2(b) as the angle at the final time (time of interception), and it is clear that the impact angle varies with the intercept point.

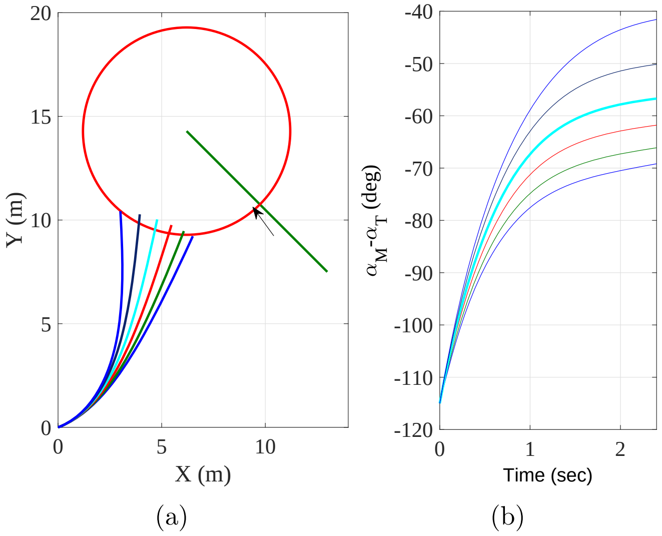


Figure 7.2: Different interception points and the corresponding impact angles.

In the missile guidance literature dealing with moving targets, we note that some papers (for example, [118]) define the impact angle as the missile heading at the final time, while some other papers (for example, [121, 122]) define it as the angle between the missile and target headings at final time. In this chapter, we adopt the latter definition of impact angle.

The influence of ξ_f on the impact angle is mathematically established as follows. For any time-step k with $\xi(k) = \xi_k$, we have the following:

$$\frac{r^2(k)V_\theta^2(k)}{V_\theta^2(k) + V_r^2(k)} - R_{\text{blast}}^2 = \xi_k \tag{7.8}$$

Define non-dimensional relative velocity components $\tilde{V}_\theta(k) \equiv V_\theta(k)/V_M$, $\tilde{V}_r(k) \equiv V_r(k)/V_M$. From Eqs. (7.1), (7.2), these are written as follows:

$$\tilde{V}_r(k) = \nu\cos(\alpha_T(k) - \theta(k)) - \cos(\alpha_M(k) - \theta(k)) \tag{7.9}$$

$$\tilde{V}_\theta(k) = \nu\sin(\alpha_T(k) - \theta(k)) - \sin(\alpha_M(k) - \theta(k)) \tag{7.10}$$

where $\nu \equiv V_T/V_M$ represents the speed ratio. Eq. (7.8) can then be written in terms of $\tilde{V}_\theta(k)$ and $\tilde{V}_r(k)$ as follows:

$$\frac{r^2(k)\tilde{V}_\theta^2(k)}{\tilde{V}_\theta^2(k) + \tilde{V}_r^2(k)} - R_{\text{blast}}^2 = \xi_k \tag{7.11}$$

We see that the non-dimensional total relative velocity term in Eq. (7.11) is as follows:

$$\tilde{V}_\theta^2(k) + \tilde{V}_r^2(k) = 1 + \nu^2 - 2\nu\cos(\alpha_M(k) - \alpha_T(k)) \tag{7.12}$$

In the above equation, note that $(\alpha_M(k) - \alpha_T(k))$ at the time of interception represents the impact angle. We denote the impact angle as ϕ_f and $\phi_f = (\alpha_M(k_f) - \alpha_T(k_f))$. Substituting Eq. (7.12) in Eq. (7.11), we get:

$$(\alpha_M(k) - \alpha_T(k)) = \cos^{-1}\left[\frac{1+\nu^2}{2\nu} - \frac{r^2(k)\tilde{V}_\theta^2(k)}{2\nu(\xi_k + R_{\text{blast}}^2)}\right] \tag{7.13}$$

At $k = k_f$, we have $r(k) = R_{\text{blast}}$ and $\xi_k = \xi_f$. Substituting these in the above, we get the impact angle magnitude at the time of interception as:

$$|\phi_f| = |(\alpha_M(k_f) - \alpha_T(k_f))| = \cos^{-1}\left[\frac{1+\nu^2}{2\nu} - \frac{R_{\text{blast}}^2\tilde{V}_\theta^2(k_f)}{2\nu(\xi_f + R_{\text{blast}}^2)}\right] \tag{7.14}$$

with the $\cos^{-1}(\cdot)$ restricted between 0 and 180 deg. In this work, our objective is to make ξ_f satisfy

$$-\epsilon_1 R_{\text{blast}}^2 \leq \xi_f \leq -\epsilon_2 R_{\text{blast}}^2 \tag{7.15}$$

where $\epsilon_1, \epsilon_2 \in (0, 1)$ and $\epsilon_1 > \epsilon_2$.

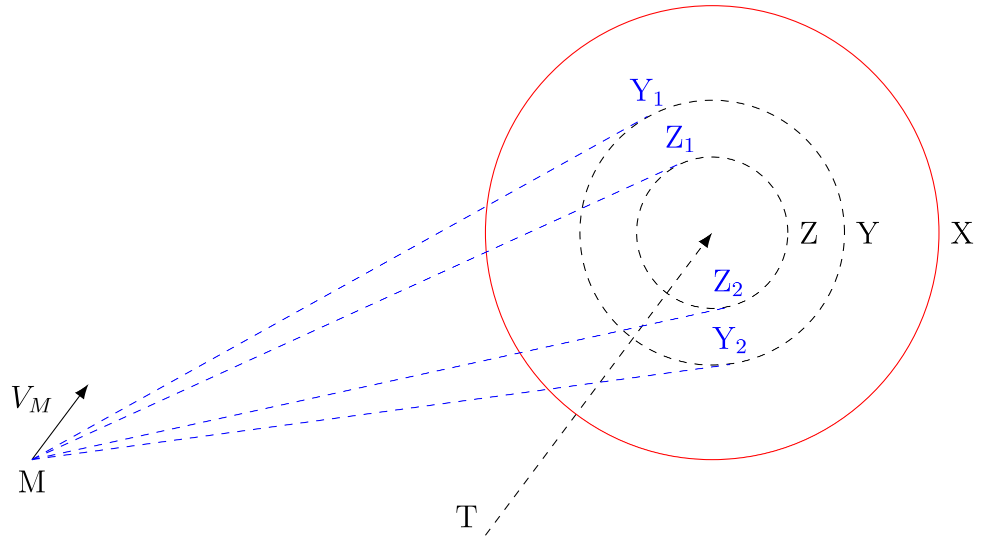


Figure 7.3: Schematic illustration of Eq. (7.15).

Eq. (7.15) can be interpreted by referring to Fig 7.3. Eq. (7.15) imposes a constraint which stipulates that the missile velocity vector is steered into one of the two sectors shown. In this figure, M and T represent the current positions of the missile and target, respectively, while X, Y and Z represent three concentric circles centered at the location of the target at the predicted time of interception. The radius of circle X is R_{blast}, while those of circles Y and Z are functions of ϵ_1 and ϵ_2, respectively. For $\epsilon_2 = 0$, circle Y is identical with circle X. As ϵ_2 increases, the radius of Y becomes progressively smaller, and as $\epsilon_2 \to 1$, the radius of Y tends to zero. A corresponding set of comments hold for circle Z as well. Thus, by appropriate choice of ϵ_1 and ϵ_2, it is possible to adjust the radii of Y and Z. The lines MY_1, MZ_1, MY_2, MZ_2 are tangents to circles Y and Z. Satisfying Eq. (7.15) ensures that the missile velocity vector, at the time of intercepting the circle, lies in either sector Y_1MZ_1 or sector Y_2MZ_2. These two sectors correspond to opposite signs of $V_\theta(k_f)$. Thus, by adjusting the values of ϵ_1 and ϵ_2, the angle of the sector in which the velocity vector of M resides, can be adjusted. This in turn adjusts the sector of the impact angle, as discussed subsequently. Next, we state the following useful result.

Lemma 7.1.1. *ξ_f satisfies Eq.* (7.15) *if and only if*

$$-|V_\theta(k_f)|\left(\frac{\epsilon_1}{1-\epsilon_1}\right)^{1/2} \leq V_r(k_f) \leq -|V_\theta(k_f)|\left(\frac{\epsilon_2}{1-\epsilon_2}\right)^{1/2}. \tag{7.16}$$

Proof. For ξ_f to satisfy Eq. (7.15), we need the following to hold:

$$-\epsilon_1 R_{\text{blast}}^2 \leq \frac{R_{\text{blast}}^2 V_\theta^2(k_f)}{V_\theta^2(k_f)+V_r^2(k_f)} - R_{\text{blast}}^2 \leq -\epsilon_2 R_{\text{blast}}^2. \tag{7.17}$$

Let us analyze the upper bound $\xi_f \leq -\epsilon_2 R_{\text{blast}}^2$ first. This can be expressed as

$$\frac{1}{1+\left(\frac{V_r(k_f)}{V_\theta(k_f)}\right)^2} - 1 \leq -\epsilon_2 \;\Rightarrow \left(\frac{V_r(k_f)}{V_\theta(k_f)}\right)^2 \geq \frac{1}{(1-\epsilon_2)} - 1. \tag{7.18}$$

Similarly, analyzing the lower bound $-\epsilon_1 R_{\text{blast}}^2 \leq \xi_f$ yields

$$\left(\frac{V_r(k_f)}{V_\theta(k_f)}\right)^2 \leq \frac{1}{(1-\epsilon_1)} - 1. \tag{7.19}$$

Combining these two results, we conclude that the required condition on $V_r(k_f)$ is given by

$$|V_\theta(k_f)|\left(\frac{\epsilon_2}{1-\epsilon_2}\right)^{1/2} \leq |V_r(k_f)| \leq |V_\theta(k_f)|\left(\frac{\epsilon_1}{1-\epsilon_1}\right)^{1/2}. \tag{7.20}$$

However, we require $V_r(\cdot)$ to be negative for achieving interception. Hence, the above condition is modified as

$$-|V_\theta(k_f)|\left(\frac{\epsilon_1}{1-\epsilon_1}\right)^{1/2} \leq V_r(k_f) \leq -|V_\theta(k_f)|\left(\frac{\epsilon_2}{1-\epsilon_2}\right)^{1/2}. \tag{7.21}$$

This completes the proof. □

Note that, by definition, $|\tilde{V}_\theta(k)| \leq (1+\nu)$. Since it is typically true that $\nu < 1$, we can see that $|\tilde{V}_\theta(k)| < 2$. Specifically, we deduce the range of $|\tilde{V}_\theta(k_f)|$ at the time of interception in the following result:

Lemma 7.1.2. *Let the following conditions hold: (i) ξ_f satisfies Eq.* (7.15)*; (ii) $\nu < 1$; (iii) For a given ν, the values of ϵ_1 and ϵ_2 are chosen such that $\epsilon_1 < \nu_c\epsilon_2 - \nu_c + 1$, where $\nu_c = \left(\frac{1-\nu}{1+\nu}\right)^2$. Then, $|\tilde{V}_\theta(k_f)|$ should lie in the range:*

$$(1-\nu)\sqrt{(1-\epsilon_2)} \leq |\tilde{V}_\theta(k_f)| \leq (1+\nu)\sqrt{(1-\epsilon_1)}. \tag{7.22}$$

Proof. Let Θ represent the argument of the $\cos^{-1}$ term in Eq. (7.14), that is, $\Theta = \frac{1+\nu^2}{2\nu} - \frac{R^2_{\text{blast}}\tilde{V}^2_\theta(k_f)}{2\nu(\xi_f + R^2_{\text{blast}})}$. Since ξ_f satisfies Eq. (7.15), we have

$$\begin{aligned}&(1-\epsilon_1)R^2_{\text{blast}} \leq (\xi_f + R^2_{\text{blast}}) \leq (1-\epsilon_2)R^2_{\text{blast}}\\ &\Rightarrow -\frac{1}{(1-\epsilon_1)R^2_{\text{blast}}} \leq -\frac{1}{(\xi_f + R^2_{\text{blast}})} \leq -\frac{1}{(1-\epsilon_2)R^2_{\text{blast}}}.\end{aligned} \tag{7.23}$$

With this, Θ satisfies

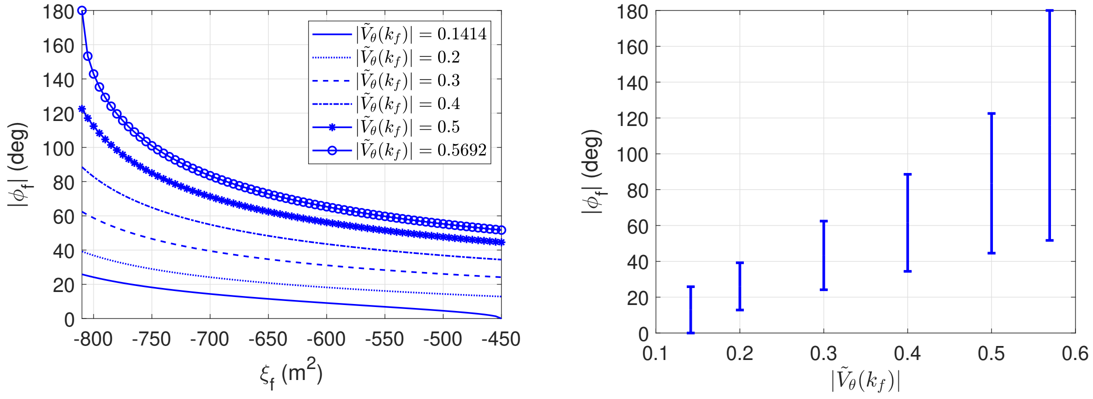
$$\frac{1+\nu^2}{2\nu} - \frac{\tilde{V}^2_\theta(k_f)}{2\nu(1-\epsilon_1)} \leq \Theta \leq \frac{1+\nu^2}{2\nu} - \frac{\tilde{V}^2_\theta(k_f)}{2\nu(1-\epsilon_2)}. \tag{7.24}$$

Thus, the necessary and sufficient conditions for the $\cos^{-1}(\Theta)$ to be a real-valued quantity, are

$$\begin{aligned}\frac{1+\nu^2}{2\nu} - \frac{\tilde{V}^2_\theta(k_f)}{2\nu(1-\epsilon_1)} &\geq -1,\\ \frac{1+\nu^2}{2\nu} - \frac{\tilde{V}^2_\theta(k_f)}{2\nu(1-\epsilon_2)} &\leq 1.\end{aligned} \tag{7.25}$$

Assuming $\nu < 1$, the first condition in the above equation leads to $|\tilde{V}_\theta(k_f)| \leq (1+\nu)\sqrt{(1-\epsilon_1)}$ and the second one leads to $|\tilde{V}_\theta(k_f)| \geq (1-\nu)\sqrt{(1-\epsilon_2)}$. For these bounds to be consistent for a choice of ϵ_1, ϵ_2 with $\nu < 1$ given, we need $(1+\nu)\sqrt{(1-\epsilon_1)} > (1-\nu)\sqrt{(1-\epsilon_2)}$. Squaring both sides of the inequality and carrying out simple algebraic manipulations, we derive $\epsilon_1 < \nu_c\epsilon_2 - \nu_c + 1$ with $\nu_c = \left(\frac{1-\nu}{1+\nu}\right)^2$. This completes the proof. □

Note that the result in Lemma 7.1.2 can be extended to the case of $\nu \geq 1$ by making straightforward modifications. However, practical engagement scenarios generally do not correspond to $\nu \geq 1$. Fig. 7.4(a) shows the trend of impact angle magnitudes computed using Eq. (7.14) for different values of ξ_f. For the results in Fig. 7.4(a), we have chosen $R_{\text{blast}} = 30$ m, $\nu = 0.8$, $\epsilon_1 = 0.9$, and $\epsilon_2 = 0.5$ which satisfy the conditions in Lemma 7.1.2. Therefore, using Lemma 7.1.2, we have $|\tilde{V}_\theta(k_f)| \in [0.1414, 0.5692]$. As shown in Fig. 7.4(a), impact angle magnitudes from 0 to 180 deg are achievable. Also, for all $|\tilde{V}_\theta(k_f)| \in [0.1414, 0.5692]$, there exists an impact angle magnitude for every ξ_f in the range. This result essentially verifies the theoretical result in Lemma 7.1.2.



(a) Variation in $|\phi_f|$ with ξ_f

(b) Ranges of $|\phi_f|$ for different $|\tilde{V}_\theta(k_f)|$

Figure 7.4: Results for the impact angle magnitude calculated using Eq. (7.14).

Fig. 7.4(b) shows the range of impact angle magnitudes achieved for every $|\tilde{V}_\theta(k_f)|$. It is interesting to note that a higher $|\tilde{V}_\theta(k_f)|$ corresponds to a wider range of impact angle magnitudes. Furthermore, a higher $|\tilde{V}_\theta(k_f)|$ results in higher magnitudes of impact angles and vice versa. Note that, for a given $\nu < 1$ and choices of ϵ_1, ϵ_2 that satisfy the condition in Lemma 7.1.2, plots similar to Figs. 7.4(a) and 7.4(b) can be generated to investigate the variation in the impact angle magnitude with varying $|\tilde{V}_\theta(k_f)|$ and ξ_f. Performing such an analysis would help one select the desired $|\tilde{V}_\theta(k_f)|$

and desired range for ξ_f so that the impact angle at the time of interception lies within a specified range.

7.2 Guidance Algorithm

The NMPC formulations for the missile guidance algorithm are elaborated in this section. In the first subsection, the prediction form utilized for both the NMPC formulations is provided. Then, the PM-NMPC is discussed in the second subsection, followed by the CC-NMPC in the third subsection. The NMPC formulations are converted into QPs, which are solved at each time-step. Both the QPs are shown to be strictly convex quadratic programs (SCQPs) (see, for example, [175]).

7.2.1 Prediction Form for the NMPC Formulations

Since the target acceleration components are treated as unknown bounded disturbances to the kinematics, we neglect the disturbance term in the prediction structure for the NMPC problems. Therefore, the discrete-time system in Eq. (7.5), without the disturbance term, is expressed in the 'referenced predictive form' [98] as

$$\boldsymbol{x}(k+j|k) = \boldsymbol{f}_d(\boldsymbol{x}(k+j-1|k)) + \boldsymbol{g}_d(\boldsymbol{x}(k+j-1|k)) \left[u(k+j-2|k) + \Delta u(k+j-1|k)\right] \tag{7.26}$$

where $(k+j|k)$ means that the current time-step is k and the distance from the current time-step is j [98]. The increment (positive or negative) in the control input is given by $\Delta u(k+j-1|k) = u(k+j-1|k) - u(k+j-2|k)$. Without loss of generality, we consider $N_p = N_c = N$ where N_p and N_c are the prediction and control horizons, respectively. With that, the standard prediction form of Eq. (7.26), required for the NMPC formulations, is given by

$$\boldsymbol{X}_k = \boldsymbol{F}_k + \boldsymbol{G}_k \Delta \boldsymbol{U}_k + \boldsymbol{g}_k \tag{7.27}$$

with

$$\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{x}(k+1|k) \\ \boldsymbol{x}(k+2|k) \\ \vdots \\ \boldsymbol{x}(k+N|k) \end{bmatrix}, \ \boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{f}_d(\boldsymbol{x}(k|k)) \\ \boldsymbol{f}_d(\boldsymbol{x}(k+1|k)) \\ \vdots \\ \boldsymbol{f}_d(\boldsymbol{x}(k+N-1|k)) \end{bmatrix},$$

$$\boldsymbol{g}_k = \begin{bmatrix} \boldsymbol{g}_d(\boldsymbol{x}(k|k))u(k-1|k-1) \\ \boldsymbol{g}_d(\boldsymbol{x}(k+1|k))u(k-1|k-1) \\ \vdots \\ \boldsymbol{g}_d(\boldsymbol{x}(k+N-1|k))u(k-1|k-1) \end{bmatrix}, \ \boldsymbol{\Delta U}_k = \begin{bmatrix} \Delta u(k|k) \\ \Delta u(k+1|k) \\ \vdots \\ \Delta u(k+N-1|k) \end{bmatrix}, \qquad (7.28)$$

$$\boldsymbol{G}_k = \begin{bmatrix} \boldsymbol{g}_d(\boldsymbol{x}(k|k)) & \boldsymbol{0}_n & \cdots & \boldsymbol{0}_n \\ \boldsymbol{g}_d(\boldsymbol{x}(k+1|k)) & \boldsymbol{g}_d(\boldsymbol{x}(k+1|k)) & \cdots & \boldsymbol{0}_n \\ \vdots & \vdots & \ddots & \boldsymbol{0}_n \\ \boldsymbol{g}_d(\boldsymbol{x}(k+N-1|k)) & \boldsymbol{g}_d(\boldsymbol{x}(k+N-1|k)) & \cdots & \boldsymbol{g}_d(\boldsymbol{x}(k+N-1|k)) \end{bmatrix}$$

where $u(k-1|k-1) = u(k-1|k)$ is the control input of the last time-step, $\boldsymbol{X}_k \in \mathbb{R}^{nN\times 1}$, $\boldsymbol{F}_k \in \mathbb{R}^{nN\times 1}$, $\boldsymbol{G}_k \in \mathbb{R}^{nN\times N}$, $\boldsymbol{\Delta U}_k \in \mathbb{R}^{N\times 1}$, and $\boldsymbol{g}_k \in \mathbb{R}^{nN\times 1}$. Note that $\Delta\boldsymbol{U}_k$ is (or is part of) the decision vector for the optimization problems to be solved for the NMPC formulations and the matrices $\boldsymbol{F}_k, \boldsymbol{G}_k, \boldsymbol{g}_k$ cannot be calculated until the problem is solved. A remedy to this issue was given in [101] where the predicted states of the previous time-step were utilized to calculate the above matrices. This is the approach adopted in the present formulation. Although not explicitly mentioned in [101], this is similar, in spirit, to the approach utilized in sequential quadratic programming [176]. Therefore, the prediction form in Eq. (7.27) is reformulated as

$$\boldsymbol{X}_k = \boldsymbol{F}_{k-1} + \boldsymbol{G}_{k-1}\Delta\boldsymbol{U}_k + \boldsymbol{g}_{k-1} \qquad (7.29)$$

where the matrices $\boldsymbol{F}_{k-1}, \boldsymbol{G}_{k-1}, \boldsymbol{g}_{k-1}$ are given by

$$\boldsymbol{F}_{k-1} = \begin{bmatrix} \boldsymbol{f}_d(\boldsymbol{x}(k-1|k-1)) \\ \boldsymbol{f}_d(\boldsymbol{x}(k|k-1)) \\ \vdots \\ \boldsymbol{f}_d(\boldsymbol{x}(k+N-2|k-1)) \end{bmatrix}, \ \boldsymbol{g}_{k-1} = \begin{bmatrix} \boldsymbol{g}_d(\boldsymbol{x}(k-1|k-1))u(k-1|k-1) \\ \boldsymbol{g}_d(\boldsymbol{x}(k|k-1))u(k-1|k-1) \\ \vdots \\ \boldsymbol{g}_d(\boldsymbol{x}(k+N-2|k-1))u(k-1|k-1) \end{bmatrix},$$

$$\boldsymbol{G}_{k-1} = \begin{bmatrix} \boldsymbol{g}_d(\boldsymbol{x}(k-1|k-1)) & \boldsymbol{0}_n & \cdots & \boldsymbol{0}_n \\ \boldsymbol{g}_d(\boldsymbol{x}(k|k-1)) & \boldsymbol{g}_d(\boldsymbol{x}(k|k-1)) & \cdots & \boldsymbol{0}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{g}_d(\boldsymbol{x}(k+N-2|k-1)) & \boldsymbol{g}_d(\boldsymbol{x}(k+N-2|k-1)) & \cdots & \boldsymbol{g}_d(\boldsymbol{x}(k+N-2|k-1)) \end{bmatrix}. \tag{7.30}$$

These reformulated matrices are utilized for both the PM-NMPC and the CC-NMPC. Also, full state measurements are assumed, that is, $\boldsymbol{y}(k|k) = \boldsymbol{y}(k) = \boldsymbol{x}(k|k) = \boldsymbol{x}(k)$ which are utilized to store the predicted states and calculate the matrices $\boldsymbol{F}_{k-1}, \boldsymbol{G}_{k-1}, \boldsymbol{g}_{k-1}$ for the NMPC formulations at the next time-step $k+1$ (see Remark 7.2.2). Also, several other key decisions for the guidance algorithm are taken based on these measurements (see Algorithm 3). Note that the sampling time for measurements is equal to the discretization time-step Δt.

7.2.2 Point Mass-Based NMPC (PM-NMPC)

A preliminary version of the PM-NMPC was proposed in [135], and is presented here. The cost function is chosen so as to penalize deviations of V_θ from a reference value V_{θ_d}. Toward this end, the output of interest is expressed in a 'referenced predictive form' [98] as $y_c(k+j|k) = \boldsymbol{C}\boldsymbol{x}(k+j|k)$, where $\boldsymbol{C} = [0 \quad 0 \quad 0 \quad 1 \quad 0]$. Similarly, $y_d(k+j|k)$ denotes the desired value for $y_c(k+j|k)$. Additionally, the magnitude of the control increments (positive or negative) are to be minimized so as to reduce the latax requirements. With these, the quadratic cost function, for time-step k, is as follows:

$$\begin{aligned} J_{1_k} =& \sum_{j=1}^{N} \left(y_c(k+j|k) - y_d(k+j|k)\right)^{\mathrm{T}} q \left(y_c(k+j|k) - y_d(k+j|k)\right) \\ &+ \sum_{j=0}^{N-1} \Delta u^{\mathrm{T}}(k+j|k) r \Delta u(k+j|k), \\ =& \left(\boldsymbol{Y}_k - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right)^{\mathrm{T}} \bar{\boldsymbol{Q}} \left(\boldsymbol{Y}_k - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right) + \boldsymbol{\Delta U}_k^{\mathrm{T}} \bar{\boldsymbol{R}} \boldsymbol{\Delta U}_k \end{aligned} \tag{7.31}$$

where

$$\begin{aligned} \boldsymbol{Y}_k &= \begin{bmatrix} y_c(k+1|k) & y_c(k+2|k) & \cdots & y_c(k+N|k) \end{bmatrix}^{\mathrm{T}}, \\ \boldsymbol{Y}_{\boldsymbol{d}_{1_k}} &= \begin{bmatrix} y_d(k+1|k) & y_d(k+2|k) & \cdots & y_d(k+N|k) \end{bmatrix}^{\mathrm{T}}, \\ \bar{\boldsymbol{Q}} &= q\boldsymbol{I}_N, \ \bar{\boldsymbol{R}} = r\boldsymbol{I}_N \end{aligned}$$

with $\boldsymbol{Y}_k \in \mathbb{R}^{N\times 1}$, $\boldsymbol{Y}_{\boldsymbol{d}_{1_k}} \in \mathbb{R}^{N\times 1}$, $q > 0$, $r > 0$, $\bar{\boldsymbol{Q}} \in \mathbb{R}^{N\times N}$, $\bar{\boldsymbol{R}} \in \mathbb{R}^{N\times N}$. Now, $\boldsymbol{Y}_k$ can be expressed in a compact form as $\boldsymbol{Y}_k = \bar{\boldsymbol{C}}\boldsymbol{X}_k$ where $\bar{\boldsymbol{C}} = \operatorname{diag}(\boldsymbol{C}, \boldsymbol{C}, \cdots, \boldsymbol{C}) = \boldsymbol{I}_N \otimes \boldsymbol{C} \in \mathbb{R}^{N\times nN}$ and $\boldsymbol{X}_k$ is as shown in Eq. (7.29). With these, the cost function J_{1_k} in Eq. (7.31) can be expressed as

$$\begin{aligned} J_{1_k} &= \left(\bar{\boldsymbol{C}}\boldsymbol{X}_k - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right)^{\mathrm{T}} \bar{\boldsymbol{Q}} \left(\bar{\boldsymbol{C}}\boldsymbol{X}_k - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right) + \boldsymbol{\Delta U}_k^{\mathrm{T}} \bar{\boldsymbol{R}} \boldsymbol{\Delta U}_k, \\ &= \left(\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{G}_{k-1}\Delta\boldsymbol{U}_k + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right)^{\mathrm{T}} \bar{\boldsymbol{Q}} \left(\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{G}_{k-1}\Delta\boldsymbol{U}_k + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}\right) \\ &\quad + \boldsymbol{\Delta U}_k^{\mathrm{T}} \bar{\boldsymbol{R}} \boldsymbol{\Delta U}_k, \\ &= (\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}})^{\mathrm{T}} \bar{\boldsymbol{Q}} (\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}) \\ &\quad + \boldsymbol{\Delta U}_k^{\mathrm{T}} (\boldsymbol{G}_{k-1}^{\mathrm{T}} \bar{\boldsymbol{C}}^{\mathrm{T}} \bar{\boldsymbol{Q}} \bar{\boldsymbol{C}} \boldsymbol{G}_{k-1} + \bar{\boldsymbol{R}}) \boldsymbol{\Delta U}_k \\ &\quad + 2(\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}})^{\mathrm{T}} \bar{\boldsymbol{Q}} \bar{\boldsymbol{C}} \boldsymbol{G}_{k-1} \boldsymbol{\Delta U}_k. \end{aligned} \tag{7.32}$$

We introduce the following notations to express the cost function in a compact form:

$$\boldsymbol{H}_k = (\boldsymbol{G}_{k-1}^{\mathrm{T}}\bar{\boldsymbol{C}}^{\mathrm{T}}\bar{\boldsymbol{Q}}\bar{\boldsymbol{C}}\boldsymbol{G}_{k-1} + \bar{\boldsymbol{R}}),\ \boldsymbol{f}_{1_k} = 2\left((\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}})^{\mathrm{T}}\bar{\boldsymbol{Q}}\bar{\boldsymbol{C}}\boldsymbol{G}_{k-1}\right)^{\mathrm{T}},$$
$$J_{c_{1_k}} = (\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}})^{\mathrm{T}}\bar{\boldsymbol{Q}}(\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{1_k}}) \tag{7.33}$$

where $\boldsymbol{H}_k \in \mathbb{R}^{N\times N}$, $\boldsymbol{f}_{1_k} \in \mathbb{R}^{N\times 1}$, $J_{c_{1_k}} \in \mathbb{R}$. Therefore, the cost function can be expressed in the following compact form:

$$J_{1_k} = J_{c_{1_k}} + \boldsymbol{\Delta U}_k^{\mathrm{T}}\boldsymbol{H}_k\boldsymbol{\Delta U}_k + \boldsymbol{f}_{1_k}^{\mathrm{T}}\boldsymbol{\Delta U}_k. \tag{7.34}$$

Box constraints are imposed on the input (latax) and input rate (change in latax) magnitudes. The control vector at each time-step is constrained as follows:

$$\boldsymbol{U}_{\min} \le \boldsymbol{U}_k \le \boldsymbol{U}_{\max} \Rightarrow \boldsymbol{U}_{\min} \le \boldsymbol{U}_{k-1} + \boldsymbol{I}_{\mathrm{lt}}\boldsymbol{\Delta U}_k \le \boldsymbol{U}_{\max} \tag{7.35}$$

where $\boldsymbol{U}_{k-1} \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{I}_{\mathrm{lt}} \in \mathbb{R}^{N\times N}$ are as follows:

$$\boldsymbol{U}_{k-1} = u(k-1|k-1)\mathbf{1}_N,\ \boldsymbol{I}_{\mathrm{lt}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{7.36}$$

The changes in the control input (latax) are constrained as

$$\boldsymbol{\Delta U}_{\min} \le \boldsymbol{\Delta U}_k \le \boldsymbol{\Delta U}_{\max}. \tag{7.37}$$

We require $V_r(\cdot)$ to remain (or become) negative for a successful interception. In the NMPC framework, this is introduced as a terminal constraint, i.e., the terminal state in the prediction is constrained so that the corresponding V_r is negative. This is given by

$$\boldsymbol{E}_1\boldsymbol{X}_k \le \boldsymbol{V}_{\boldsymbol{r}_d} \tag{7.38}$$

where $\boldsymbol{E}_1 \in \mathbb{R}^{N\times nN}$ and $\boldsymbol{V}_{\boldsymbol{r}_d} \in \mathbb{R}^N$ are given by

$$\boldsymbol{E}_1 = \begin{bmatrix} \mathbf{0}_n^{\mathrm{T}} & \cdots & \mathbf{0}_n^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_n^{\mathrm{T}} & \cdots & \boldsymbol{e}_1 \end{bmatrix}, \ \boldsymbol{V}_{\boldsymbol{r}_d} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ V_{r_d} \end{bmatrix}. \tag{7.39}$$

In the above, $\boldsymbol{e}_1 = [0 \quad 1 \quad 0 \quad 0 \quad 0]$ and $V_{r_d} < 0$. Substituting $\boldsymbol{X}_k$ from Eq. (7.29) in Eq. (7.38), we get:

$$\boldsymbol{E}_1(\boldsymbol{F}_{k-1} + \boldsymbol{G}_{k-1}\Delta\boldsymbol{U}_k + \boldsymbol{g}_{k-1}) \leq \boldsymbol{V}_{\boldsymbol{r}_d} \ \Rightarrow \boldsymbol{E}_1\boldsymbol{G}_{k-1}\boldsymbol{\Delta U}_k \leq \boldsymbol{V}_{\boldsymbol{r}_d} - \boldsymbol{E}_1(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}). \tag{7.40}$$

The constraints in Eqs. (7.35), (7.37), (7.40) are expressed in a compact form as $\boldsymbol{A}_{\mathbf{1}_k}\boldsymbol{\Delta U}_k \leq \boldsymbol{b}_{\mathbf{1}_k}$ where $\boldsymbol{A}_{\mathbf{1}_k} \in \mathbb{R}^{5N\times N}$ and $\boldsymbol{b}_{\mathbf{1}_k} \in \mathbb{R}^{5N}$ are given in Eq. (7.41).

$$\boldsymbol{A}_{\mathbf{1}_k} = \begin{bmatrix} \boldsymbol{I}_N \\ -\boldsymbol{I}_N \\ \boldsymbol{I}_{\mathrm{lt}} \\ -\boldsymbol{I}_{\mathrm{lt}} \\ \boldsymbol{E}_1\boldsymbol{G}_{k-1} \end{bmatrix}, \ \boldsymbol{b}_{\mathbf{1}_k} = \begin{bmatrix} \boldsymbol{\Delta U}_{\max} \\ -\boldsymbol{\Delta U}_{\min} \\ \boldsymbol{U}_{\max} - \boldsymbol{U}_{k-1} \\ -\boldsymbol{U}_{\min} + \boldsymbol{U}_{k-1} \\ \boldsymbol{V}_{\boldsymbol{r}_d} - \boldsymbol{E}_1(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \end{bmatrix}. \tag{7.41}$$

Remark 7.2.1. *Note that the cost function developed for the present formulation is different from the one proposed in [101]. The cost function chosen in [101] is suitable for the state regulation problem, whereas we have developed a cost function that allows tracking of a desired V_{θ_d}. Also, the terminal constraint ensures that $V_r \leq V_{r_d} < 0$. Therefore, if one were to implement the PM-NMPC for the entire engagement (as in [135]), the well-known necessary and sufficient conditions for target interception can be achieved by setting $V_{\theta_d} = 0$. Alternatively, for a less aggressive guidance strategy, the desired trajectory can be set as $V_{\theta_d} = c\sqrt{r}$ [177] where $c > 0$ is a constant.*

On the other hand, if all states are regulated to the origin, as in [101], there might arise an engagement scenario where V_r is regulated to zero faster than r, and in such cases, the missile might not be able to intercept the target.

With the above formulations of the cost function and constraints, the PM-NMPC requires the solution to the following QP at each time-step:

$$\begin{aligned} &\min_{\Delta \boldsymbol{U}_k} J_{c_{1_k}} + \Delta \boldsymbol{U}_k^{\mathrm{T}} \boldsymbol{H}_k \Delta \boldsymbol{U}_k + \boldsymbol{f}_{1_k}^{\mathrm{T}} \Delta \boldsymbol{U}_k, \\ &\text{subject to} \quad \boldsymbol{A}_{\mathbf{1}_k} \Delta \boldsymbol{U}_k \leq \boldsymbol{b}_{\mathbf{1}_k}. \end{aligned} \tag{7.42}$$

Remark 7.2.2. *Let us assume that the solution to the QP in Eq.* (7.42) *at time-step k is given by $\boldsymbol{\Delta U}_k^*$. Then, the latax for that time-step ($u(k|k)$) is calculated as the first element of the vector $\boldsymbol{U}_k = \boldsymbol{U}_{k-1} + \boldsymbol{I}_{lt} \boldsymbol{\Delta U}_k^*$. Utilizing $\boldsymbol{U}_k$ and the measurement at the current time-step $\boldsymbol{x}(k|k)$, we calculate and store the states as $\boldsymbol{X}_s(k) = \Big[\boldsymbol{x}^{\mathrm{T}}(k|k) \quad \boldsymbol{x}^{\mathrm{T}}(k+1|k) \cdots \boldsymbol{x}^{\mathrm{T}}(k+N-1|k)\Big]^{\mathrm{T}}$ where $\boldsymbol{x}(k+j|k) = \boldsymbol{f}_d(\boldsymbol{x}(k+j-1|k)) + \boldsymbol{g}_d(\boldsymbol{x}(k+j-1|k))u(k+j-1|k)$ with $u(k+j-1|k)$ as the j-th element of the vector $\boldsymbol{U}_k$ for $j = 1, 2, \cdots, N-1$. $\boldsymbol{X}_s(k)$ is then utilized at the next time-step ($k+1$) to calculate the matrices $\boldsymbol{F}_{k-1}, \boldsymbol{G}_{k-1}, \boldsymbol{g}_{k-1}$. Further, $\boldsymbol{U}_k$ is reformulated as $\boldsymbol{U}_k = u(k|k)\mathbf{1}_N$ where $u(k|k)$ is the current latax value. These calculations are carried out recursively at each time-step (see Algorithm 3).*

It is not always possible to satisfy the hard terminal constraint for V_r in Eq. (7.40). In order to avoid constraint infeasibility issues, the hard constraint is relaxed and is replaced with a soft constraint. This is achieved by utilizing a slack variable-based approach [178,179]. Toward this end, we relax the inequality in Eq. (7.40) with $\boldsymbol{E}_1 \boldsymbol{G}_{k-1} \boldsymbol{\Delta U}_k + [\mathbf{0}_{N-1}^{\mathrm{T}} \quad -\gamma_k]^{\mathrm{T}} \leq \boldsymbol{V}_{\boldsymbol{r}_d} - \boldsymbol{E}_1(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1})$ where $\gamma_k \geq 0$ is the slack variable. Also, we introduce penalty functions for the slack variable γ_k [178, 179]. Hence, the QP in Eq. (7.42) is modified as

$$\begin{aligned}
&\min_{\Delta\boldsymbol{U}_k,\gamma_k} \; J_{c_{1_k}} + \Delta\boldsymbol{U}_k^{\mathrm{T}}\boldsymbol{H}_k\Delta\boldsymbol{U}_k + \boldsymbol{f}_{1_k}^{\mathrm{T}}\Delta\boldsymbol{U}_k + r_1\gamma_k^2 + r_2\gamma_k,\\
&\text{subject to} \begin{bmatrix} \boldsymbol{I}_N \\ -\boldsymbol{I}_N \\ \boldsymbol{I}_{\mathrm{lt}} \\ -\boldsymbol{I}_{\mathrm{lt}} \end{bmatrix} \Delta\boldsymbol{U}_k \leq \begin{bmatrix} \boldsymbol{\Delta U}_{\max} \\ -\boldsymbol{\Delta U}_{\min} \\ \boldsymbol{U}_{\max} - \boldsymbol{U}_{k-1} \\ -\boldsymbol{U}_{\min} + \boldsymbol{U}_{k-1} \end{bmatrix},\\
&\qquad \boldsymbol{E}_1\boldsymbol{G}_{k-1}\boldsymbol{\Delta U}_k + [\boldsymbol{0}_{N-1}^{\mathrm{T}} \quad -\gamma_k]^{\mathrm{T}} \leq \boldsymbol{V}_{\boldsymbol{r}_d} - \boldsymbol{E}_1(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}),\\
&\qquad \gamma_k \geq 0
\end{aligned} \tag{7.43}$$

where $r_1, r_2 > 0$. The decision variable for the modified QP is $\boldsymbol{\zeta}_k = \begin{bmatrix} \Delta\boldsymbol{U}_k^{\mathrm{T}} & \gamma_k \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{N+1}$. The cost function and the constraints are expressed in terms of $\boldsymbol{\zeta}_k$ as

$$\begin{aligned}
J_{2_k} &= J_{c_{1_k}} + \boldsymbol{\zeta}_k^{\mathrm{T}}\tilde{\boldsymbol{H}}_{\boldsymbol{1}_k}\boldsymbol{\zeta}_k + \tilde{\boldsymbol{f}}_{\boldsymbol{1}_k}^{\mathrm{T}}\boldsymbol{\zeta}_k,\\
\tilde{\boldsymbol{A}}_{\boldsymbol{1}_k}\boldsymbol{\zeta}_k &\leq \tilde{\boldsymbol{b}}_{\boldsymbol{1}_k}
\end{aligned} \tag{7.44}$$

where the relevant matrices are shown in Eq. (7.45) with $\bar{\boldsymbol{0}} = [\boldsymbol{0}_{N-1}^{\mathrm{T}} \quad -1]^{\mathrm{T}}$.

$$\tilde{\boldsymbol{H}}_{\boldsymbol{1}_k} = \mathrm{diag}\left(\boldsymbol{H}_k, r_1\right), \; \tilde{\boldsymbol{f}}_{\boldsymbol{1}_k} = \begin{bmatrix} \boldsymbol{f}_{1_k} \\ r_2 \end{bmatrix}, \; \tilde{\boldsymbol{A}}_{\boldsymbol{1}_k} = \begin{bmatrix} \boldsymbol{I}_N & \boldsymbol{0}_N \\ -\boldsymbol{I}_N & \boldsymbol{0}_N \\ \boldsymbol{I}_{\mathrm{lt}} & \boldsymbol{0}_N \\ -\boldsymbol{I}_{\mathrm{lt}} & \boldsymbol{0}_N \\ \boldsymbol{E}_1\boldsymbol{G}_{k-1} & \bar{\boldsymbol{0}} \\ \boldsymbol{0}_N^{\mathrm{T}} & -1 \end{bmatrix}, \; \tilde{\boldsymbol{b}}_{\boldsymbol{1}_k} = \begin{bmatrix} \boldsymbol{b}_{\boldsymbol{1}_k} \\ 0 \end{bmatrix}. \tag{7.45}$$

Finally, the soft-constraint-based QP for the PM-NMPC can be summarized as

$$\begin{aligned}
&\min_{\boldsymbol{\zeta}_k} \; J_{c_{1_k}} + \boldsymbol{\zeta}_k^{\mathrm{T}}\tilde{\boldsymbol{H}}_{\boldsymbol{1}_k}\boldsymbol{\zeta}_k + \tilde{\boldsymbol{f}}_{\boldsymbol{1}_k}^{\mathrm{T}}\boldsymbol{\zeta}_k,\\
&\text{subject to } \tilde{\boldsymbol{A}}_{\boldsymbol{1}_k}\boldsymbol{\zeta}_k \leq \tilde{\boldsymbol{b}}_{\boldsymbol{1}_k}.
\end{aligned} \tag{7.46}$$

Remark 7.2.3. *The QP in Eq.* (7.46) *is feasible for all k and guarantees the important feature of recursive feasibility in the MPC framework. Furthermore, if r_2 is sufficiently large and a feasible solution to the hard-constraint-based QP in Eq.* (7.42) *exists, the optimal solution to the soft-constraint-based QP in Eq.* (7.46) *corresponds to that of the hard-constraint-based QP in Eq.* (7.42) *[178]. Thus, we only need to solve the QP in Eq.* (7.46) *for the PM-NMPC.*

Next, we state the useful result regarding the QP to be solved for the PM-NMPC.

Theorem 7.2.1. *The QP in Eq.* (7.46) *is an SCQP $\forall k$.*

Proof. The feasible set for the optimization problem in Eq. (7.46) is a polytope (or polyhedron) for all k. Hence, the feasible set is convex for all k [180]. Next, let us consider the Hessian of the cost function J_{1_k} which can be rewritten as $\boldsymbol{H}_k = \bar{\boldsymbol{H}}_{k-1} + \bar{\boldsymbol{R}}$ where $\bar{\boldsymbol{H}}_{k-1} = \boldsymbol{G}_{k-1}^{\mathrm{T}} \bar{\boldsymbol{C}}^{\mathrm{T}} \bar{\boldsymbol{Q}} \bar{\boldsymbol{C}} \boldsymbol{G}_{k-1}$. Clearly, $\bar{\boldsymbol{H}}_{k-1}$ is positive semi-definite (at least). Thus, $\boldsymbol{H}_k$ is positive definite since $\bar{\boldsymbol{R}}$ is positive definite. Therefore, we conclude that the cost function J_{1_k} is strictly convex for all k. Now, due to the positive definiteness of $\boldsymbol{H}_k$ and $r_1 > 0$ (by choice), it follows that the Hessian $\tilde{\boldsymbol{H}}_{\mathbf{1}_k}$ is positive definite for all k. Thus, the cost function J_{2_k} is strictly convex for all k. With that, the corresponding QP in Eq. (7.46) is an SCQP for all k. This completes the proof. □

For the remainder of the chapter, we refer to the QP in Eq. (7.46) as SCQP-1.

7.2.3 Collision Cone-Based NMPC (CC-NMPC)

The derivation of CC-NMPC is similar to that of the PM-NMPC except for the fact that CC-NMPC has additional reliance on Lemmas 7.1.1 and 7.1.2. We elaborate the design process in the following enumerated list:

1. Assuming the conditions given in Lemma 7.1.2 hold, choose a desired $V_{\theta_d}(k_f) = V_M \tilde{V}_{\theta_d}(k_f)$ using Lemma 7.1.2. Since the PM-NMPC already has a cost function that allows tracking in the V_θ, we utilize the same cost function for the CC-NMPC which is given in Eq. (7.34). However, as the desired values to be tracked in the CC-NMPC are, in general, not the same as for PM-NMPC, we replace $\boldsymbol{Y}_{\boldsymbol{d}_{1_k}}$ with $\boldsymbol{Y}_{\boldsymbol{d}_{2_k}}$. Thus, the cost function is given by

$$J_{3_k} = J_{c2_k} + \boldsymbol{\Delta U}_k^{\mathrm{T}} \boldsymbol{H}_k \boldsymbol{\Delta U}_k + \boldsymbol{f}_{2_k}^{\mathrm{T}} \boldsymbol{\Delta U}_k \tag{7.47}$$

   where $J_{c2_k} = (\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{2_k}})^{\mathrm{T}} \bar{\boldsymbol{Q}} (\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{2_k}})$ and $\boldsymbol{f}_{2_k} = 2\left((\bar{\boldsymbol{C}}(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) - \boldsymbol{Y}_{\boldsymbol{d}_{2_k}})^{\mathrm{T}} \bar{\boldsymbol{Q}} \bar{\boldsymbol{C}} \boldsymbol{G}_{k-1}\right)^{\mathrm{T}}$.

2. Based on ϵ_1, ϵ_2 and the $V_{\theta_d}(k_f)$ chosen, we derive the desired range of $V_r(k_f)$ using Lemma 7.1.1. In order to enforce the V_r to lie within the desired range mentioned above, we constrain the V_r for the entire prediction horizon N. Accordingly, we modify the constraint in Eq. (7.40) as

$$\boldsymbol{E}_2 \boldsymbol{G}_{k-1} \boldsymbol{\Delta U}_k \leq \boldsymbol{V}_{\boldsymbol{r}_{d_2}} - \boldsymbol{E}_2(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \tag{7.48}$$

   where $\boldsymbol{V}_{\boldsymbol{r}_{d_2}} = V_{r_2} \mathbf{1}_N$ with $V_{r_2} = -|V_{\theta_d}(k_f)| \left(\frac{\epsilon_2}{1-\epsilon_2}\right)^{1/2}$ and $\boldsymbol{E}_2 = \boldsymbol{I}_N \otimes \boldsymbol{e}_1$ with $\boldsymbol{e}_1 = [0 \quad 1 \quad 0 \quad 0 \quad 0]$. Similarly, we introduce

$$\begin{aligned} &\boldsymbol{E}_2 \boldsymbol{G}_{k-1} \boldsymbol{\Delta U}_k \geq \boldsymbol{V}_{\boldsymbol{r}_{d_1}} - \boldsymbol{E}_2(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \\ &\Rightarrow -\boldsymbol{E}_2 \boldsymbol{G}_{k-1} \boldsymbol{\Delta U}_k \leq -\boldsymbol{V}_{\boldsymbol{r}_{d_1}} + \boldsymbol{E}_2(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \end{aligned} \tag{7.49}$$

   where $\boldsymbol{V}_{\boldsymbol{r}_{d_1}} = V_{r_1} \mathbf{1}_N$ with $V_{r_1} = -|V_{\theta_d}(k_f)| \left(\frac{\epsilon_1}{1-\epsilon_1}\right)^{1/2}$.

3. Incorporate the constraints on the latax and the change in latax, in the same way as done in PM-NMPC.

Thus, combining the steps 2 and 3 outlined above, the constraints for the CC-NMPC are as shown in Eq. (7.50). Finally, the CC-NMPC requires solution to the

$$\boldsymbol{A}_{\boldsymbol{2}_k} = \begin{bmatrix} \boldsymbol{I}_N \\ -\boldsymbol{I}_N \\ \boldsymbol{I}_{\mathrm{lt}} \\ -\boldsymbol{I}_{\mathrm{lt}} \\ \boldsymbol{E}_2\boldsymbol{G}_{k-1} \\ -\boldsymbol{E}_2\boldsymbol{G}_{k-1} \end{bmatrix}, \ \boldsymbol{b}_{\boldsymbol{2}_k} = \begin{bmatrix} \boldsymbol{\Delta U}_{\max} \\ -\boldsymbol{\Delta U}_{\min} \\ \boldsymbol{U}_{\max} - \boldsymbol{U}_{k-1} \\ -\boldsymbol{U}_{\min} + \boldsymbol{U}_{k-1} \\ \boldsymbol{V}_{\boldsymbol{r}_{d_2}} - \boldsymbol{E}_2(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \\ -\boldsymbol{V}_{\boldsymbol{r}_{d_1}} + \boldsymbol{E}_2(\boldsymbol{F}_{k-1} + \boldsymbol{g}_{k-1}) \end{bmatrix}. \tag{7.50}$$

following hard-constraint-based QP at each time-step:

$$\begin{aligned} &\min_{\Delta \boldsymbol{U}_k} J_{c2_k} + \Delta \boldsymbol{U}_k^{\mathrm{T}} \boldsymbol{H}_k \Delta \boldsymbol{U}_k + \boldsymbol{f}_{2_k}^{\mathrm{T}} \Delta \boldsymbol{U}_k, \\ &\text{subject to} \quad \boldsymbol{A}_{\boldsymbol{2}_k} \Delta \boldsymbol{U}_k \leq \boldsymbol{b}_{\boldsymbol{2}_k}. \end{aligned} \tag{7.51}$$

Similar to the PM-NMPC, the hard constraints on V_r are relaxed using slack variables to avoid infeasibility issues. The corresponding soft-constraint-based QP is given by

$$\begin{aligned} &\min_{\boldsymbol{\zeta}_k} \ J_{c2_k} + \boldsymbol{\zeta}_k^{\mathrm{T}} \tilde{\boldsymbol{H}}_{\boldsymbol{2}_k} \boldsymbol{\zeta}_k + \tilde{\boldsymbol{f}}_{\boldsymbol{2}_k}^{\mathrm{T}} \boldsymbol{\zeta}_k, \\ &\text{subject to } \tilde{\boldsymbol{A}}_{\boldsymbol{2}_k} \boldsymbol{\zeta}_k \leq \tilde{\boldsymbol{b}}_{\boldsymbol{2}_k} \end{aligned} \tag{7.52}$$

where $\boldsymbol{\zeta}_k = \begin{bmatrix} \Delta \boldsymbol{U}_k^{\mathrm{T}} & \boldsymbol{\gamma}_{1_k}^{\mathrm{T}} & \boldsymbol{\gamma}_{2_k}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{3N}$ and the other relevant matrices are shown in Eq. (7.53) with $\boldsymbol{O}_1 = \boldsymbol{O}_{N \times 2N}$, $\boldsymbol{O}_2 = [-\boldsymbol{I}_N \quad \boldsymbol{O}_{N \times N}]$, and $\boldsymbol{O}_3 = [\boldsymbol{O}_{N \times N} \quad -\boldsymbol{I}_N]$. Based on the discussion in Remark 7.2.3, we only need to solve the soft-constraint-based QP in Eq. (7.52) for the CC-NMPC.

We choose to specify $\boldsymbol{Y}_{\boldsymbol{d}_{2_k}}$ such that the reference value for tracking asymptotically converges to $V_{\theta_d}(k_f)$ as $r(k) \to R_{\mathrm{blast}}$. To this end, we choose $\boldsymbol{Y}_{\boldsymbol{d}_{2_k}} = \left(\frac{R_{\mathrm{blast}}}{r(k)}\right)^{\beta} V_{\theta_d}(k_f) \mathbf{1}_N$ where $\beta \in (0, 1]$ is a tuning parameter. Therefore, we will have $V_\theta(k) \to V_{\theta_d}(k_f)$ as $r(k) \to R_{\mathrm{blast}}$. Next, we state the following useful result regarding the QP in Eq. (7.52).

Theorem 7.2.2. *The QP in Eq.* (7.52) *is an SCQP* $\forall k$.

$$\tilde{\boldsymbol{H}}_{\boldsymbol{2}_k} = \mathrm{diag}\left(\boldsymbol{H}_k, r_1\boldsymbol{I}_{2N}\right), \ \tilde{\boldsymbol{f}}_{\boldsymbol{2}_k} = \begin{bmatrix} \boldsymbol{f}_{2_k} \\ r_2\mathbf{1}_{2N} \end{bmatrix}, \ \tilde{\boldsymbol{A}}_{\boldsymbol{2}_k} = \begin{bmatrix} \boldsymbol{I}_N & \boldsymbol{O}_1 \\ -\boldsymbol{I}_N & \boldsymbol{O}_1 \\ \boldsymbol{I}_{\mathrm{lt}} & \boldsymbol{O}_1 \\ -\boldsymbol{I}_{\mathrm{lt}} & \boldsymbol{O}_1 \\ \boldsymbol{E}_2\boldsymbol{G}_{k-1} & \boldsymbol{O}_2 \\ -\boldsymbol{E}_2\boldsymbol{G}_{k-1} & \boldsymbol{O}_3 \\ \boldsymbol{O}_{N\times N} & \boldsymbol{O}_2 \\ \boldsymbol{O}_{N\times N} & \boldsymbol{O}_3 \end{bmatrix}, \ \tilde{\boldsymbol{b}}_{\boldsymbol{2}_k} = \begin{bmatrix} \boldsymbol{b}_{\boldsymbol{2}_k} \\ \mathbf{0}_{2N} \end{bmatrix}. \tag{7.53}$$

Proof. The proof follows directly from that of Theorem 7.2.1. □

For the remainder of the chapter, we refer to the QP in Eq. (7.52) as SCQP-2.

Remark 7.2.4. *For the hard constraints on the magnitudes of latax and the changes in latax, one can simply choose* $\boldsymbol{\Delta U}_{\max} = \Delta u_{\max}\mathbf{1}_N$, $\boldsymbol{\Delta U}_{\min} = \Delta u_{\min}\mathbf{1}_N$, $\boldsymbol{U}_{\max} = u_{\max}\mathbf{1}_N$, $\boldsymbol{U}_{\min} = u_{\min}\mathbf{1}_N$.

Note that the discussion in Remark 7.2.2 is applicable to the CC-NMPC as well (see Algorithm 3). The proposed missile guidance algorithm is summarized in Algorithm 3.

Remark 7.2.5. *Since the proposed guidance algorithm requires solving SCQPs, we do not need to provide an initial guess for the solution at each time-step. This is advantageous compared to the method utilized in [118] where an initial guess is required for the solution. Also, due to the strict convexity property, the optimal solution for the SCQPs can be efficiently obtained using existing convex optimization tools and solvers.*

Remark 7.2.6. *The proposed algorithm treats the target acceleration as an unknown but bounded disturbance to the kinematics. Additionally, the proposed algorithm neither requires the target acceleration nor does it attempt to estimate the same. Thus, we cannot expect that the NMPC formulation will achieve 'offset-free' tracking (see,*

Algorithm 3 Missile guidance algorithm

```
1:  (Initialization of PM-NMPC) Choose the parameters involved. Set $k = 0$, $u(k-1|k-1) = 0$, $\boldsymbol{U}_{k-1} = \boldsymbol{0}_N$ and $\boldsymbol{X}_s(k-1) = \boldsymbol{1}_N \otimes \boldsymbol{x}_0$.
2:  while $x_1(k) \geq R_{\text{switch}}$ do
3:      Calculate the matrices $\boldsymbol{F}_{k-1}, \boldsymbol{G}_{k-1}, \boldsymbol{g}_{k-1}$ utilizing $\boldsymbol{X}_s(k-1)$ and $u(k-1|k-1)$ (Eq. (7.30)).
4:      Solve SCQP-1 to get $\boldsymbol{\Delta U}_k^\star$.
5:      Calculate $\boldsymbol{U}_k = \boldsymbol{U}_{k-1} + \boldsymbol{I}_{\text{lt}}\boldsymbol{\Delta U}_k^\star$ and apply the first element of $\boldsymbol{U}_k$ as the current latax value.
6:      Utilizing $\boldsymbol{U}_k$, calculate and store $\boldsymbol{X}_s(k)$.
7:      Reformulate $\boldsymbol{U}_k$ as $\boldsymbol{U}_k = u(k|k)\boldsymbol{1}_N$ where $u(k|k)$ is the current latax value.
8:      Set $k = k + 1$ and go to Step 2.
9:  end while
10: (Initialization of CC-NMPC) Reformulate $\boldsymbol{U}_{k-1}$ and $\boldsymbol{X}_s(k-1)$, if required.
11: while $x_1(k) \geq R_{\text{blast}}$ do
12:     Calculate the matrices $\boldsymbol{F}_{k-1}, \boldsymbol{G}_{k-1}, \boldsymbol{g}_{k-1}$ utilizing $\boldsymbol{X}_s(k-1)$ and $u(k-1|k-1)$ (Eq. (7.30)).
13:     Solve SCQP-2 to get $\boldsymbol{\Delta U}_k^\star$.
14:     Calculate $\boldsymbol{U}_k = \boldsymbol{U}_{k-1} + \boldsymbol{I}_{\text{lt}}\boldsymbol{\Delta U}_k^\star$ and apply the first element of $\boldsymbol{U}_k$ as the current latax value.
15:     Utilizing $\boldsymbol{U}_k$, calculate and store $\boldsymbol{X}_s(k)$.
16:     Reformulate $\boldsymbol{U}_k$ as $\boldsymbol{U}_k = u(k|k)\boldsymbol{1}_N$ where $u(k|k)$ is the current latax value.
17:     Set $k = k + 1$ and go to Step 11.
18: end while
```

for example, [181]) for a maneuvering target with $a_T \neq 0$. However, by choosing the NMPC parameters properly, we can keep the tracking errors small (see Section 7.3).

$$\begin{aligned}
r_0 &= 20\text{ km},\ V_M = 350\text{ m/s},\ V_T = 250\text{ m/s},\ q = 1,\ r = 0.01,\ r_1 = 100,\\
r_2 &= 200,\ \Delta t = 0.01\text{ s},\ R_{\text{switch}} = 2\text{ km},\ R_{\text{blast}} = 30\text{ m},\ \epsilon_1 = 0.9,\ \epsilon_2 = 0.5,\\
u_{\max} &= 50g,\ u_{\min} = -u_{\max},\ \beta = 0.4,\ |V_{\theta_d}(k_f)| = 80\text{ m/s}.
\end{aligned} \tag{7.54}$$

7.3 Simulation Results

In this section, we present simulation results for the hybrid NMPC guidance algorithm. We consider maneuvering targets with (a) constant acceleration and (b) time-varying acceleration. All the simulations are carried out on a Dell XPS 13 laptop with a 8.00 GB RAM and a 1.60-1.80 GHz Intel(R) Core(TM) i5-8250U processor running MATLAB R2019b. To this end, 'quadprog' has been utilized to solve the SCQPs. We choose the vectors $\boldsymbol{U}_{\max}$, $\boldsymbol{U}_{\min}$, $\boldsymbol{\Delta U}_{\max}$, $\boldsymbol{\Delta U}_{\min}$ as discussed in Remark 7.2.4. Unless otherwise mentioned, we have utilized the parameter values shown in Eq. (7.54) for the simulations where g is the acceleration due to gravity (in m/s^2). The choices of V_M, V_T, ϵ_1, and ϵ_2 are such that the conditions in Lemma 7.1.2 are satisfied. Also, for both the NMPC formulations, we take $N = 30$. For the PM-NMPC, we set $V_{r_d} = 0.1V_{r_0}$ if $V_{r_0} < 0$ and $V_{r_d} = -0.1V_{r_0}$ otherwise. For the PM-NMPC, $\boldsymbol{Y}_{\boldsymbol{d}_{1_k}} = \mathbf{0}_N$ is chosen. Also, we set $\Delta u_{\max} = 0.01u_{\max}$, $\Delta u_{\min} = 0.01u_{\min}$ for the PM-NMPC and $\Delta u_{\max} = 0.1u_{\max}$, $\Delta u_{\min} = 0.1u_{\min}$ for the CC-NMPC. Moreover, a first-order actuator is included (for both the NMPC formulations), which is given by

$$\frac{a_M}{a_{M_c}} = \frac{1}{\tau s + 1} \tag{7.55}$$

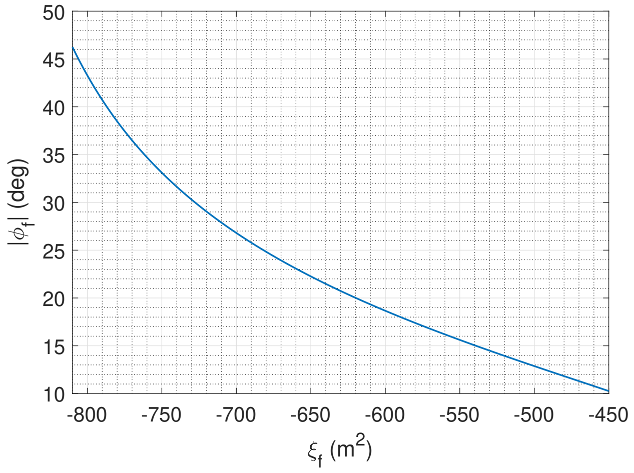


Figure 7.5: The variation in the impact angle magnitude with ξ_f for $V_M = 350$ m/s, $\mathrm{V_T} = 250$ m/s, $\mathrm{R_{blast}} = 30$ m, $\epsilon_1 = 0.9$, $\epsilon_2 = 0.5$, $|\mathrm{V}_{\theta_\mathrm{d}}(\mathrm{k_f})| = 80$ m/s.

where s is the Laplace variable, a_{M_c} is the commanded latax (generated by the NMPC formulations), and $\tau = 0.05$ s. Note that similar values for τ (the actuator time constant) have been utilized in the literature (see, for example [182, 183]). For each engagement scenario in the following subsections, Case-1 and Case-2 correspond to $V_{\theta_d}(k_f) = 80$ m/s and $V_{\theta_d}(k_f) = -80$ m/s, respectively. For the values of the relevant parameters chosen, the variation in the magnitude of achievable impact angles, calculated using Eq. (7.14), is shown in Fig. 7.5. Moreover, using Lemma 7.1.1, the desired range of V_r for the CC-NMPC is given by $V_r \in [-240, -80]$ m/s. Also note, the results corresponding to the PM-NMPC for the aforementioned two cases would be the same and would only vary for different engagement scenarios. In all the following results, the vertical lines (-.-.-) represent the time when the algorithm switches from PM-NMPC to CC-NMPC. Additionally, the horizontal dashed lines (- - -) in the latax plots represent the latax constraints, the horizontal dashed-dot lines (-.-.-) in the collision cone function plots mark the desired range for ξ_f (i.e., $\xi_f \in [-810, -450]$

m^2), and the solid horizontal lines (—) in the V_r plots mark the desired range for V_r (i.e., $V_r \in [-240, -80]$ m/s).

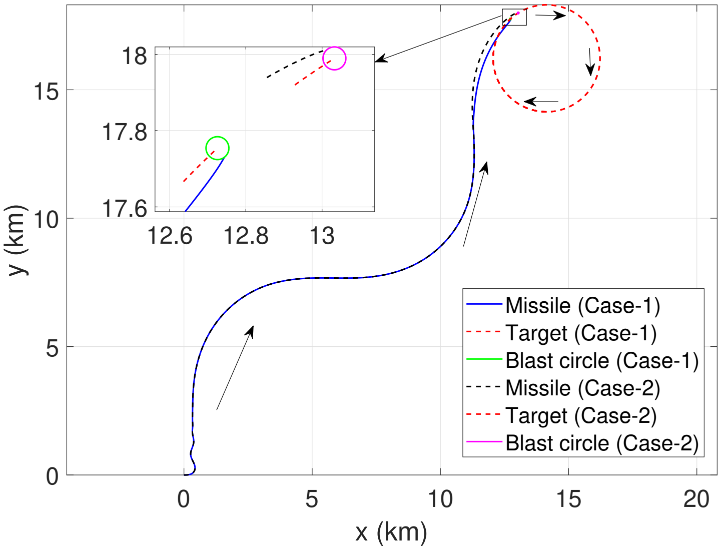
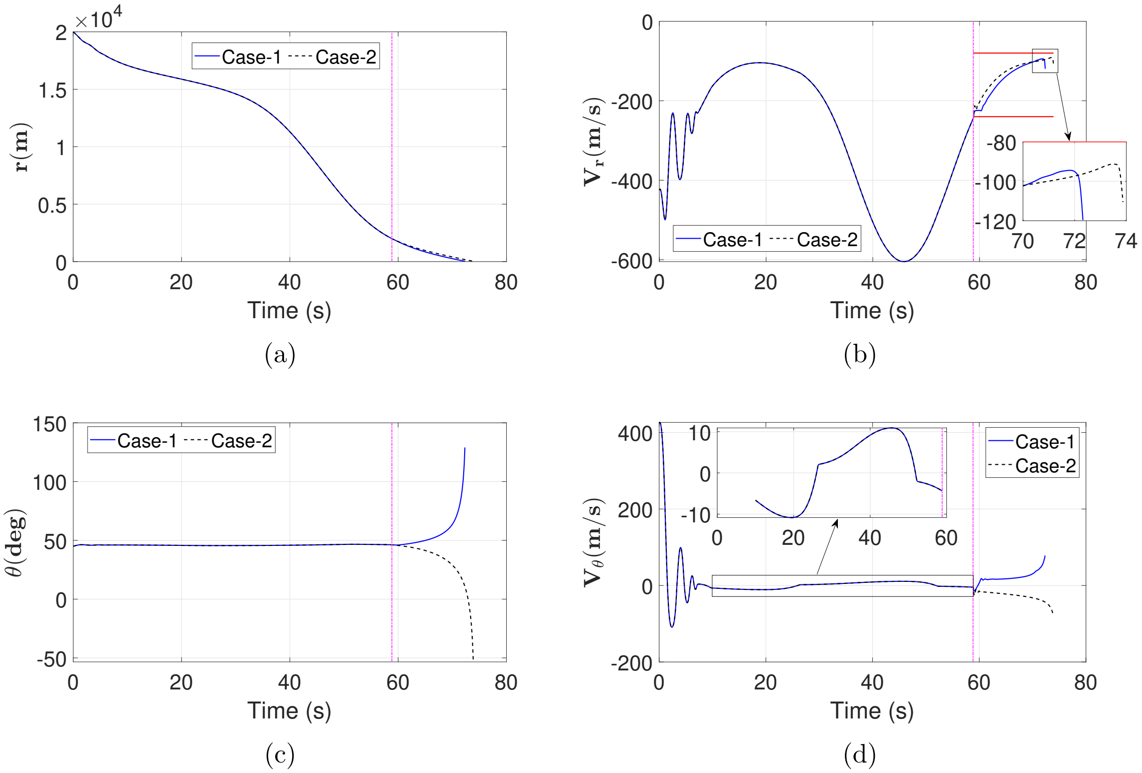


Figure 7.6: Missile and target trajectories for the target with constant acceleration.

7.3.1 Target With Constant Acceleration

We first consider a target moving with constant acceleration of $-3g$. The initial engagement geometry is given by $\theta_0 = 45$ deg, $\alpha_{T_0} = 180$ deg, $\alpha_{M_0} = 0$ deg. The simulation results are shown in Figs. 7.6, 7.7, 7.8, 7.9, 7.10. The missile successfully intercepts the target in both the cases, as shown in Fig. 7.6. The states are shown in Fig. 7.7. Interestingly, we observe that once V_θ converges to a neighborhood of zero, both V_r and V_θ follow approximate periodic trajectories for rest of the PM-NMPC-based guidance phase (Figs. 7.7(b) and 7.7(d)). Also, the results in Fig. 7.7(b) show that the CC-NMPC is able to initially drive the V_r to be within the desired range and keep it there for both the cases.



(a)

(b)

(c)

(d)

Figure 7.7: States for the target with constant acceleration.

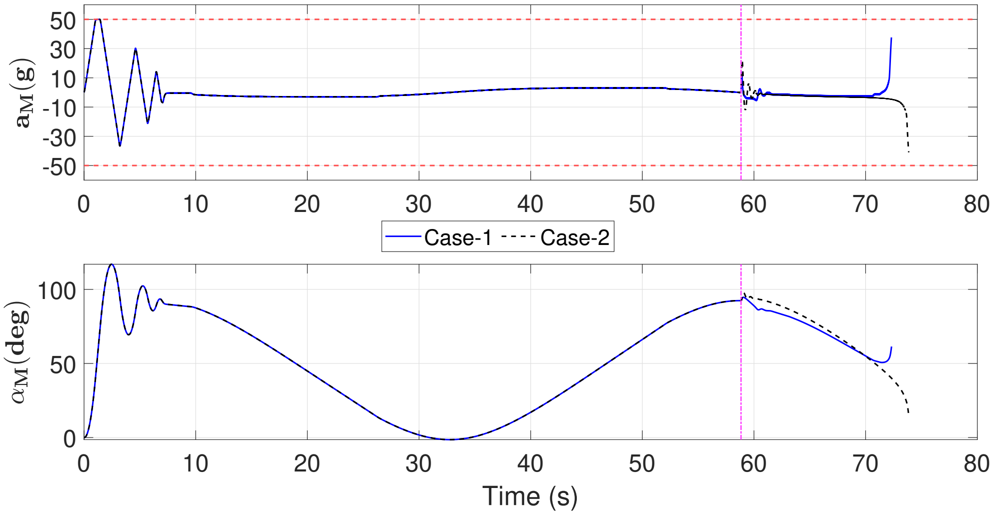


Figure 7.8: Missile latax and heading angle for the target with constant acceleration.

The latax profiles for the missile are shown in Fig. 7.8 along with the corresponding missile heading angle profiles. Based on the results shown in Fig. 7.8, we can conclude that the constraints are satisfied for both the cases. Maximum latax is initially applied to bring the V_θ close to zero. After that, the latax values converge close to zero for the latter part of the PM-NMPC-based guidance. In fact, we observe that the latax actually shows an approximate periodic behavior with a small amplitude and a large time period. A direct consequence of this can be observed in the missile heading angle shown in Fig. 7.8. The periodic-like nature of the latax gets manifested in the variations of V_r and V_θ, as mentioned above and as shown in Figs. 7.7(b) and 7.7(d). The V_θ tracking results for the CC-NMPC are shown in Fig. 7.9. Based on these results, we deduce that the tracking right after the switch to CC-NMPC is not adequate (especially for Case-1) as the V_r at the time of switch does not lie in the desired range. Thus, the algorithm prioritizes on minimizing the slack variables and drive the V_r within the desired range. Once V_r is driven inside the desired range, the tracking improves and the missile achieves the interception with tracking errors of approximately 1% in both the cases.

The variations in the angle $(\alpha_M - \alpha_T)$ are plotted in Fig. 7.10 and we conclude that the missile achieves impact angles of 18.66 and -17.82 deg for Case-1 and Case-2, respectively. The corresponding theoretical values of impact angle magnitudes, calculated using Eq. (7.14), are 18.31 and 17.19 deg for Case-1 and Case-2, respectively. The variations in the collision cone function are shown in Fig. 7.10. The collision cone function becomes large (positive values) as V_θ becomes higher in magnitude, i.e, the target moves out of the collision cone. However, at the time of interception, the collision cone function is driven within the desired range as shown in Fig. 7.10. Based on all these results, we conclude that the algorithm is successful in achieving all the guidance objectives.

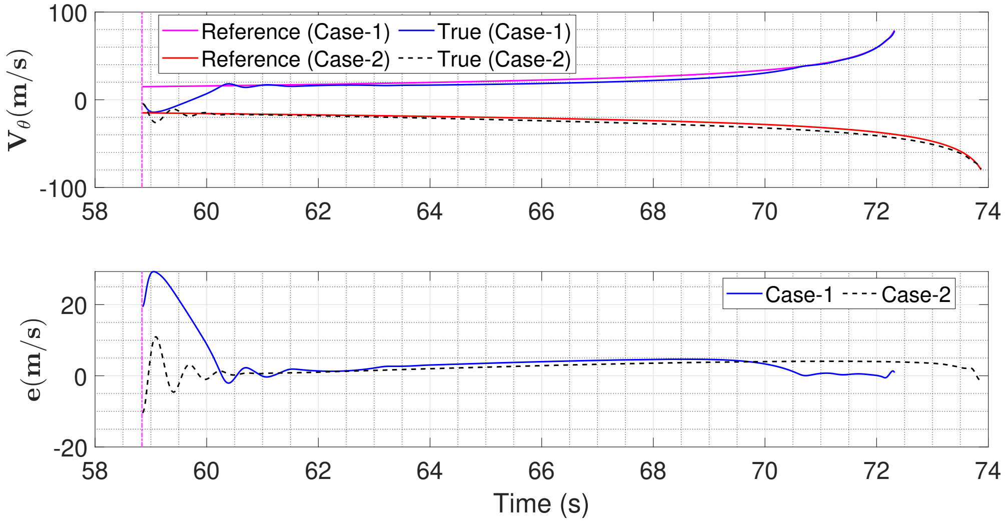


Figure 7.9: Tracking in V_θ and the tracking error (CC-NMPC) for the target with constant acceleration.

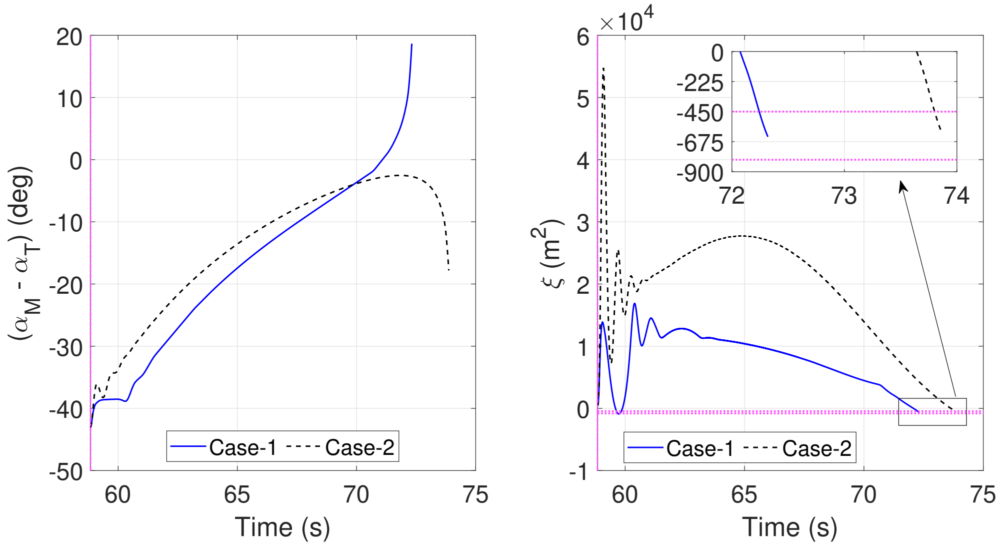


Figure 7.10: The angle $(\alpha_M - \alpha_T)$ and the collision cone function (ξ) for the target with constant acceleration.

7.3.2 Target With Time-Varying Acceleration

In this subsection we consider targets with time-varying accelerations. The following scenarios are considered:

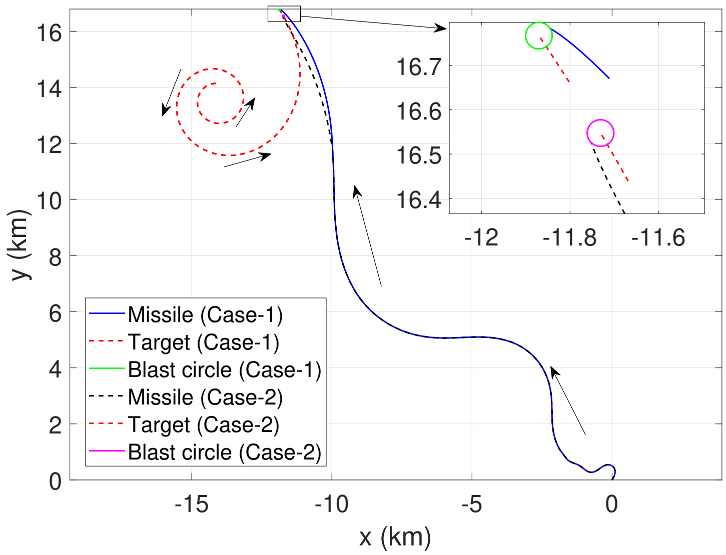
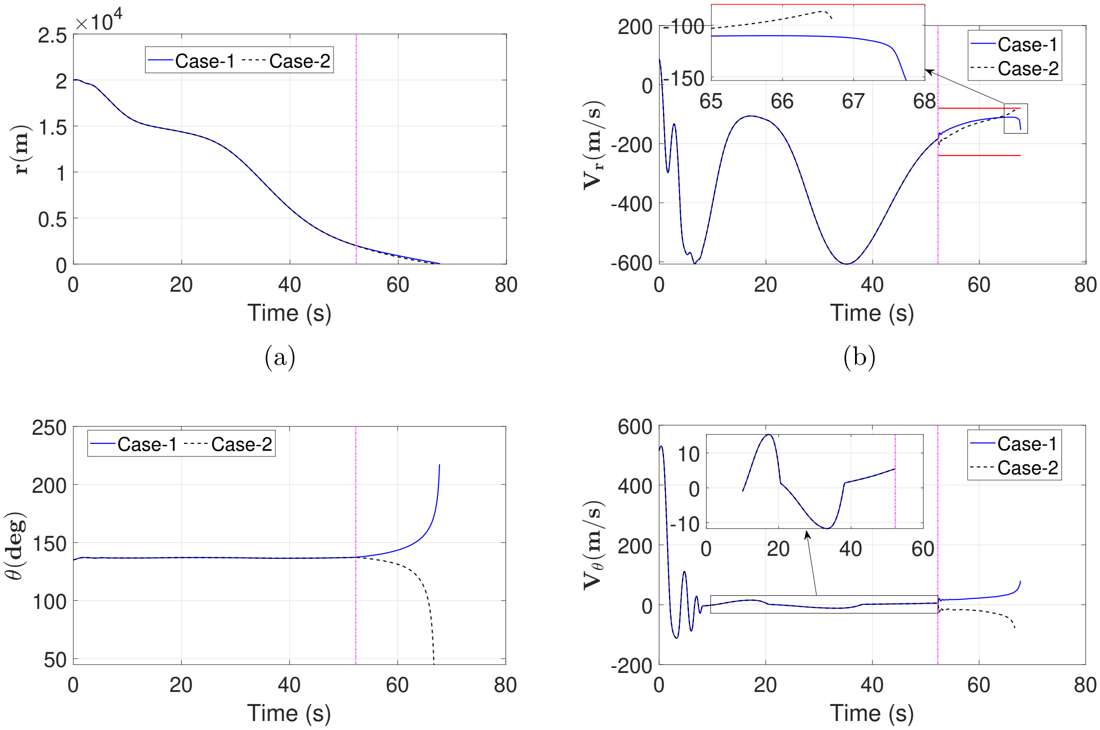


Figure 7.11: Missile and target trajectories for the target with exponential acceleration profile.



(a)

(b)

(c)

(d)

Figure 7.12: States for the target with exponential acceleration profile.

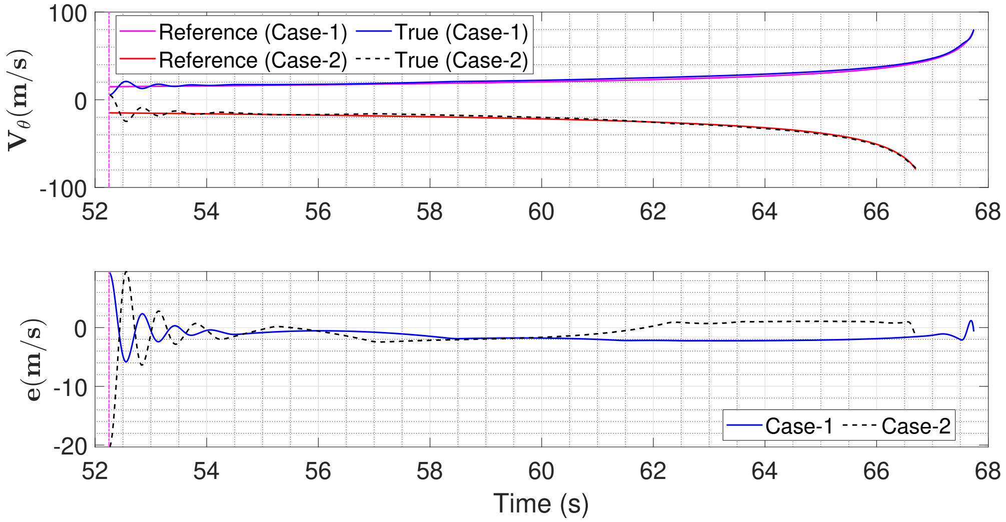


Figure 7.13: Tracking in V_θ and the tracking error (CC-NMPC) for the target with exponential acceleration profile.

7.3.2.1 Scenario-1

For this engagement scenario, we consider a target moving with an acceleration profile given by $a_T = 10g\exp(-0.03t)$. The initial engagement geometry is given by $\theta_0 = 135$ deg, $\alpha_{T_0} = 180$ deg, $\alpha_{M_0} = 60$ deg. Note that $V_{r_0} > 0$ for this initial engagement geometry. The simulation results are shown in Figs. 7.11, 7.12, 7.13, 7.14, 7.15. As shown in Fig. 7.11, interception is achieved by the missile in both the cases. The states are shown in Fig. 7.12 and we observe the approximate periodic behaviors in V_r and V_θ (during the PM-NMPC-based guidance phase) that are qualitatively similar to last scenario. On the other hand, unlike the last scenario, V_r is within the desired range when the algorithm switches from PM-NMPC to CC-NMPC (see Fig. 7.12(b)). Moreover, the V_r is successfully constrained within this desired range for the remainder of the engagement in both cases. As a result, the tracking in V_θ for CC-NMPC is better compared to the last scenario. Fig. 7.13 shows that V_θ

converges close to the commanded values with some initial transients, after the switch to CC-NMPC.

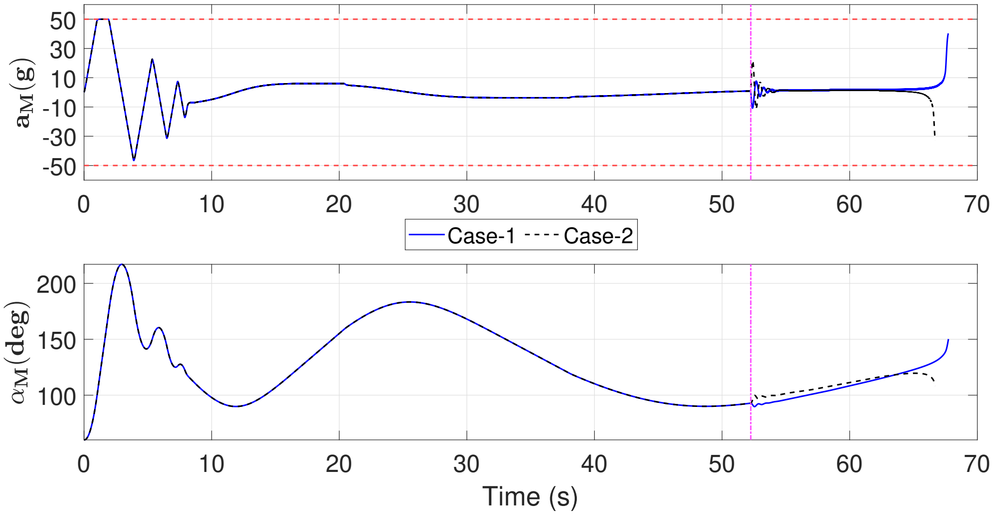


Figure 7.14: Missile latax and heading angle for the target with exponential acceleration profile.

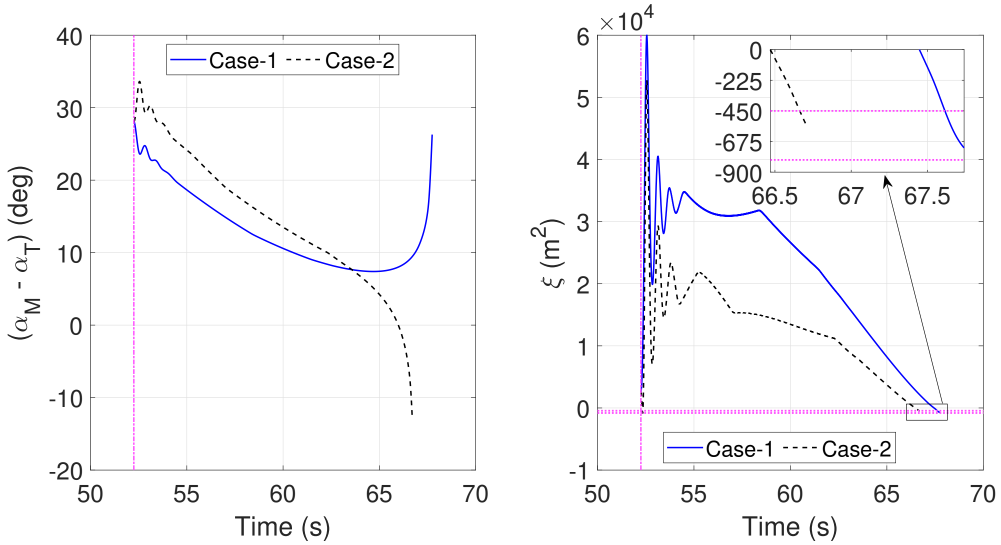


Figure 7.15: The angle $(\alpha_M - \alpha_T)$ and the collision cone function (ξ) for the target with exponential acceleration profile.

The latax profiles for the missile are shown in Fig. 7.14 along with the corresponding missile heading angle profiles. Based on the results shown in Fig. 7.14, we can conclude that the constraints are satisfied in both the cases. The latax profiles are qualitatively similar to ones in the last scenario (cf. Fig. 7.8) and the explanations for these trends are similar as well. Note, in addition to making V_θ get close to zero, the PM-NMPC is also able to make $V_r < 0$ initially. Further, the periodic-like behavior in the latax during the PM-NMPC-based guidance phase can be observed from the results shown in Fig. 7.14. The effects of this behavior can be seen in the missile heading angle, V_r, and V_θ variations (see Figs. 7.14, 7.12(b), 7.12(d)).

The variations in the the angle $(\alpha_M - \alpha_T)$ are plotted in Fig. 7.15 and we deduce that the impact angles are 26.29 and -12.5 deg for Case-1 and Case-2, respectively. The corresponding theoretical values of impact angle magnitudes, calculated using Eq. (7.14), are 26.65 and 12.3 deg for Case-1 and Case-2, respectively. The variations in the collision cone function are depicted in Fig. 7.15. As in the last scenario, the collision cone function is driven within the desired range for both the cases at the time of interception, as shown in Fig. 7.15. Finally, based on all these results, we conclude that the algorithm is successful in achieving all the guidance objectives for this scenario.

7.3.2.2 Scenario-2

Next, we consider a target moving with a sinusoidal acceleration profile given by $a_T = 7.5g\sin(0.25t)$. The initial geometry is given by $\theta_0 = 45$ deg, $\alpha_{T_0} = 30$ deg, $\alpha_{M_0} = 0$ deg. The simulation results are shown in Figs. 7.16, 7.17, 7.18, 7.19, 7.20. As shown in Fig. 7.16, the interception is successful in both the cases. The states are shown in Fig. 7.17 and more pronounced periodic behaviors in V_r and V_θ (during the PM-NMPC-based guidance phase) are observed, compared to the last

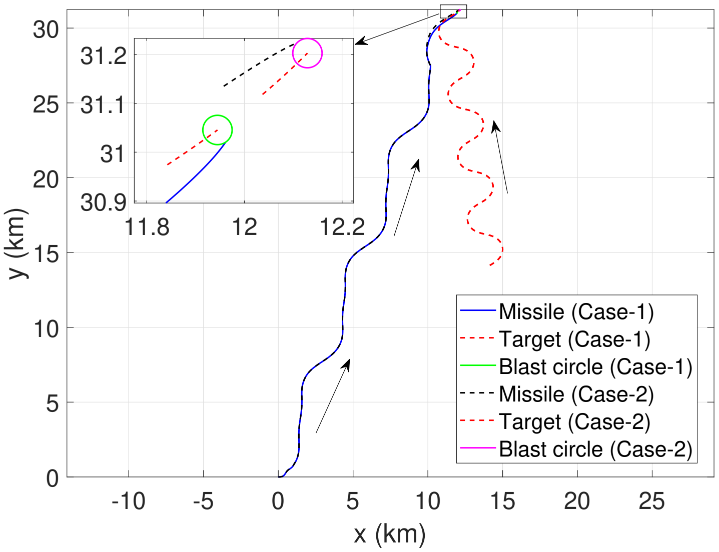
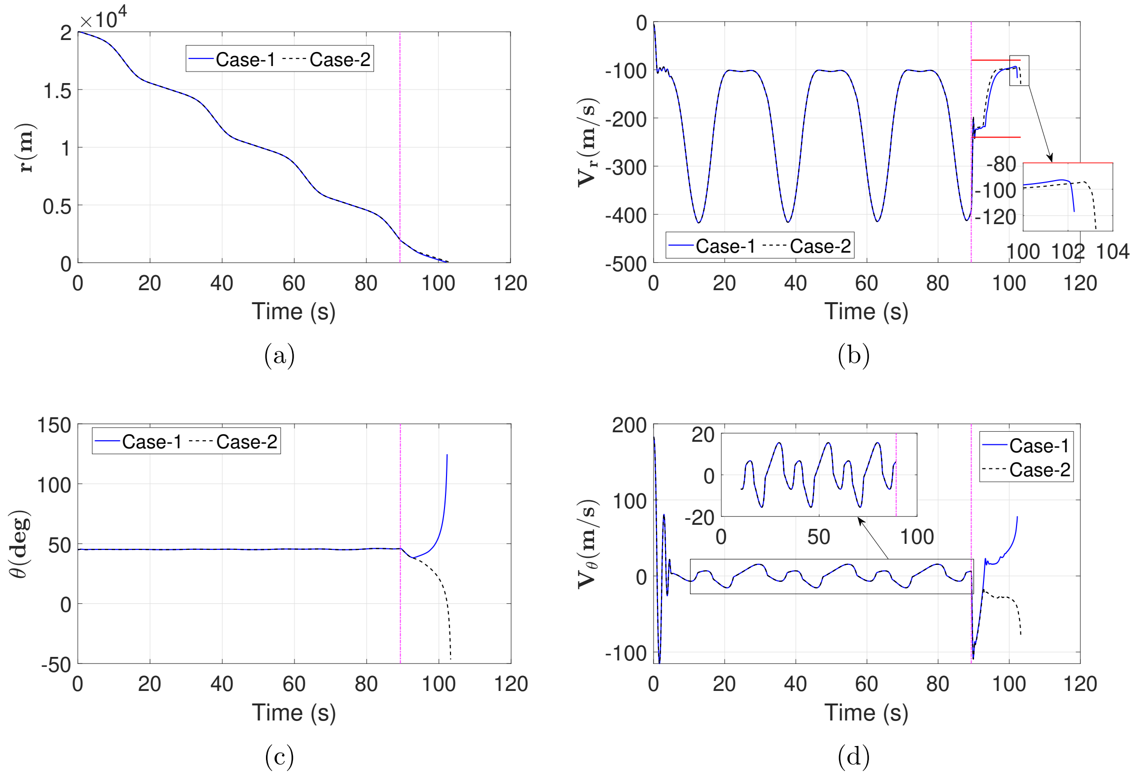


Figure 7.16: Missile and target trajectories for the target with sinusoidal acceleration (Scenario-2).



(a)

(b)

(c)

(d)

Figure 7.17: States for the target with sinusoidal acceleration (Scenario-2).

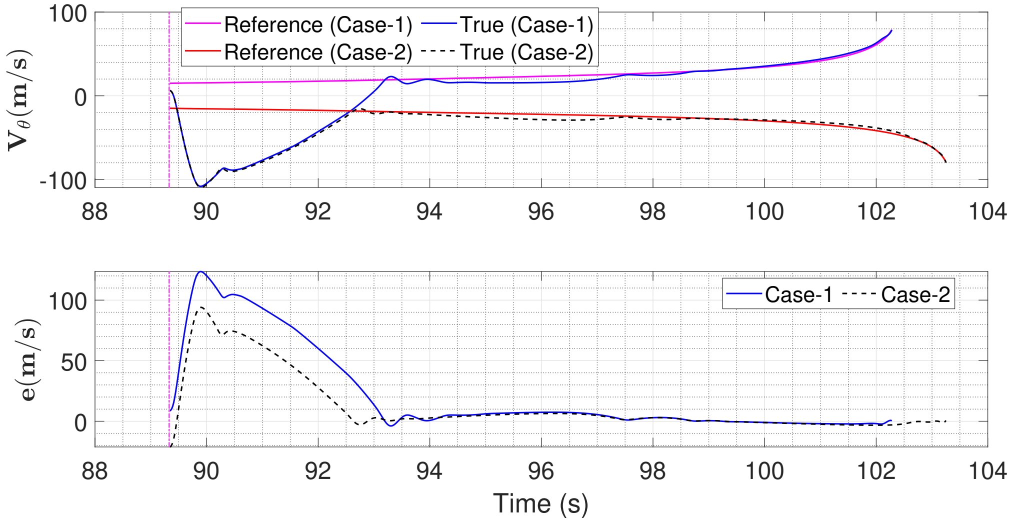


Figure 7.18: Tracking in V_θ and the tracking error (CC-NMPC) for the target with sinusoidal acceleration (Scenario-2).

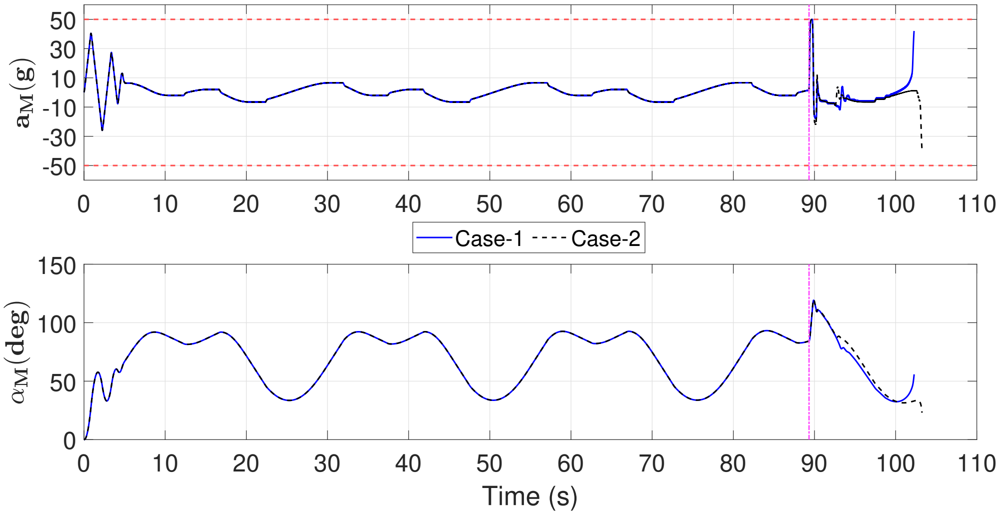


Figure 7.19: Missile latax and heading angle for the target with sinusoidal acceleration (Scenario-2).

two scenarios. Also, unlike the last two scenarios, V_r is well outside the desired range at the time of switch (see Fig. 7.17(b)). As a result, the algorithm primarily focuses on driving V_r to the desired range and the tracking in V_θ gets affected (as shown in Fig. 7.18). However, as soon as the V_r is driven inside the desired range, the tracking

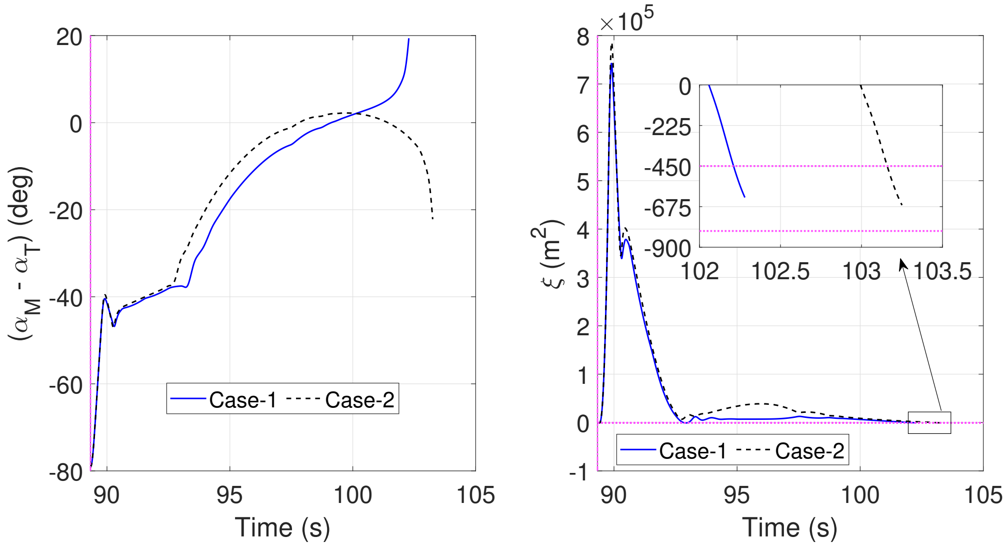


Figure 7.20: The angle $(\alpha_M - \alpha_T)$ and the collision cone function (ξ) for the target with sinusoidal acceleration (Scenario-2).

in V_θ improves and the tracking errors at the time of interception are approximately 1% in both the cases. Also, the algorithm is able to keep V_r within the desired range for the remainder of the engagement in both cases.

The time histories of latax for both the cases are shown in Fig. 7.19. This shows that constraint satisfaction has been achieved for this engagement scenario too. After applying the maximum latax initially, the latax shows a periodic-like behavior for the latter part of the PM-NMPC-based guidance, similar to the last two scenarios. For this scenario, however, the time period is shorter. The effects of this behavior can be clearly seen on the variations in α_M, V_θ, and V_r. Further, the periodicity in the trajectories of V_θ and V_r (especially V_θ) is more prominent for this scenario as the target acceleration is periodic (sinusoidal) as well.

The variations in the the angle $(\alpha_M - \alpha_T)$ are plotted in Fig. 7.20 and we deduce that the impact angles are 19.39 and -22.18 deg for Case-1 and Case-2, respectively. The corresponding theoretical values of impact angle magnitudes, calculated using Eq.

(7.14), are 19.59 and 23.51 deg for Case-1 and Case-2, respectively. The variations in the collision cone function are depicted in Fig. 7.20. The algorithm is successful in keeping the collision cone function value within the desired range at the time of interception in both the cases. Finally, based on all these results, we conclude that all the guidance objectives have been satisfied for this scenario.

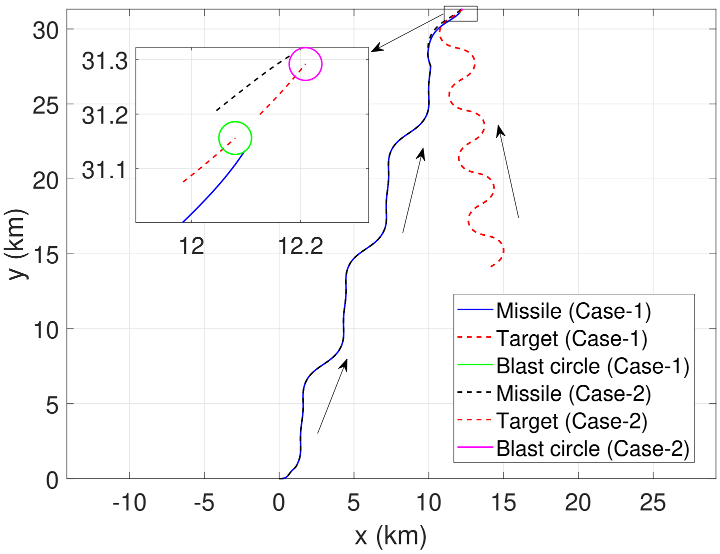
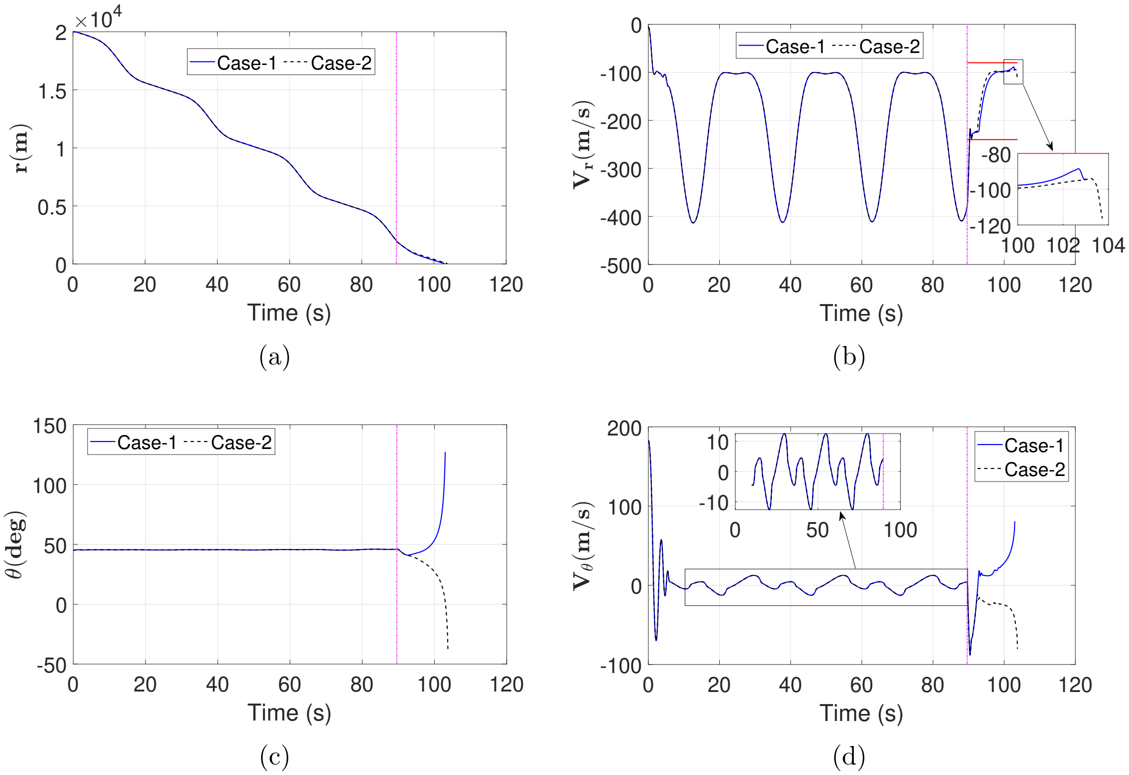


Figure 7.21: Missile and target trajectories for the target with sinusoidal acceleration (Scenario-3).

7.3.2.3 Scenario-3

For this, we consider the same target acceleration profile and initial engagement geometry as in the last scenario. However, the latax constraints are modified as $u_{\max} = 25g$, $u_{\min} = -u_{\max}$. Thus, the maximum allowable latax magnitude and magnitude of the changes in latax have been reduced by half. The simulation results for this scenario are shown in Figs. 7.21, 7.22, 7.23, 7.24, 7.25, for which β is readjusted to 0.45. As shown in Fig. 7.21, the interception is successful in both the cases. Overall,



(a)

(b)

(c)

(d)

Figure 7.22: States for the target with sinusoidal acceleration (Scenario-3).

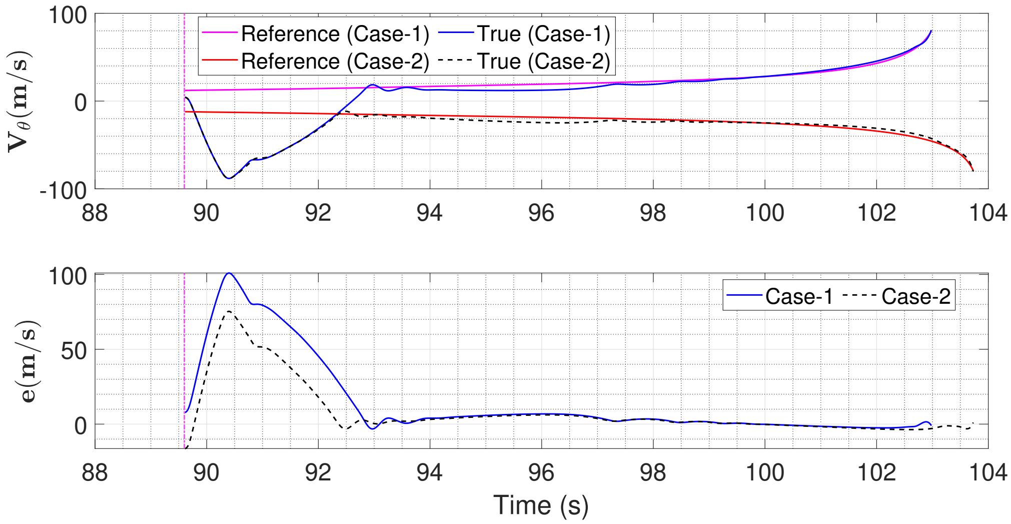


Figure 7.23: Tracking in V_θ and the tracking error (CC-NMPC) for the target with sinusoidal acceleration (Scenario-3).

all the trends observed here are qualitatively similar to the trends observed in the last scenario (Scenario-2). Further, constraint satisfaction (latax) has been achieved for this engagement scenario as well.

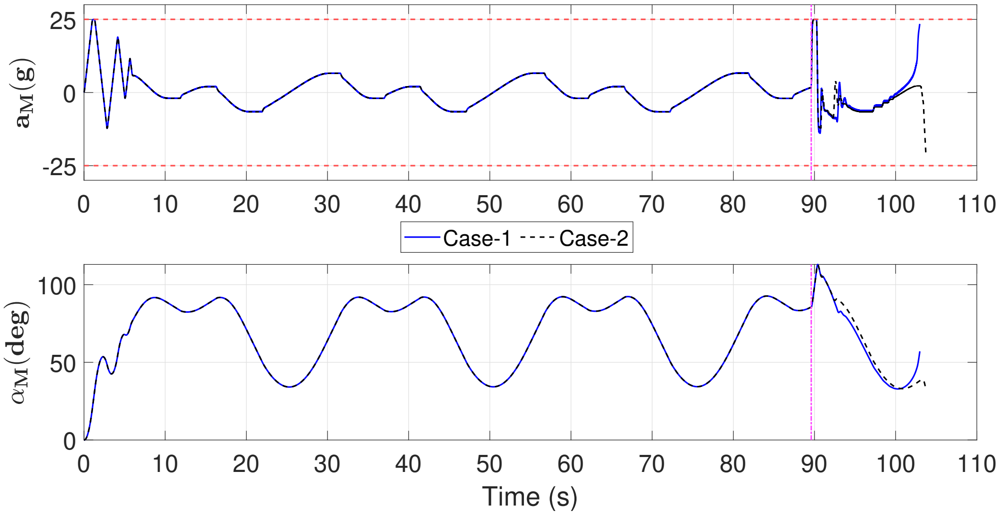


Figure 7.24: Missile latax and heading angle for the target with sinusoidal acceleration (Scenario-3).

The variations in the the angle $(\alpha_M - \alpha_T)$ are plotted in Fig. 7.25 and we deduce that the impact angles are 14.64 and -18.81 deg for Case-1 and Case-2, respectively. The corresponding theoretical magnitudes of impact angles, calculated using Eq. (7.14), are 15.29 and 18.03 deg for Case-1 and Case-2, respectively. The variations in the collision cone function are depicted in Fig. 7.25. The algorithm is successful in keeping the collision cone function value within the desired range at the time of interception. Again, all the guidance objectives have been satisfied for this scenario as well.

Remark 7.3.1. *With the NMPC-related parameters as given in this section, computation time for each SCQP is approximately between 0.006-0.007 s (on average), which*

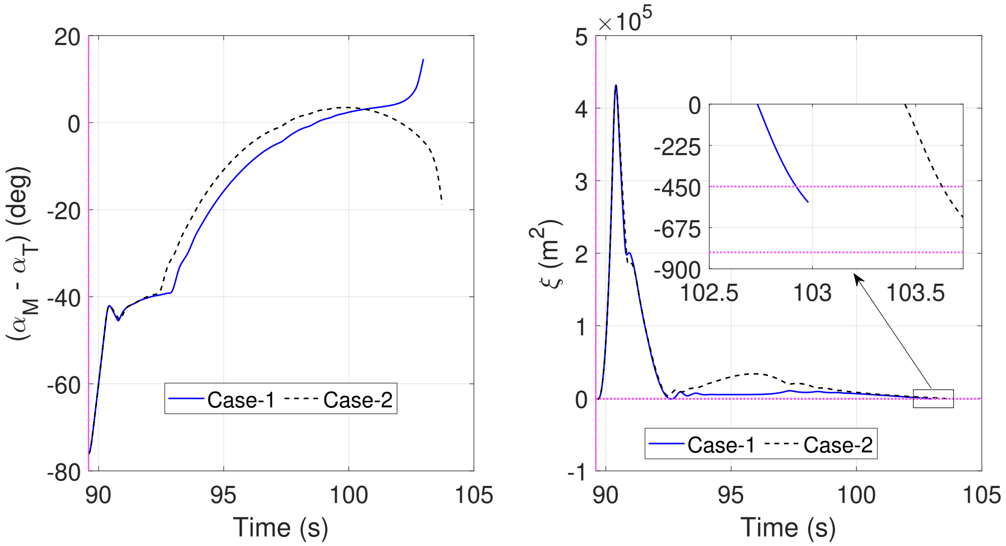


Figure 7.25: The angle $(\alpha_M - \alpha_T)$ and the collision cone function (ξ) for the target with sinusoidal acceleration (Scenario-3).

is less than the discretization time-step utilized for the simulation results $(\Delta t = 0.01$ *s). Based on this, we conclude that the computational performance of the proposed algorithm is adequate.*

7.4 Chapter Summary

In this chapter, an NMPC and collision cone-based hybrid missile guidance algorithm is proposed for achieving interception between a missile and a target, with an impact angle constraint. The proposed algorithm includes two components: PM-NMPC and CC-NMPC. The PM-NMPC is employed during the initial phase of the engagement (when the distance between the missile and the target is large), while the CC-NMPC is employed during the second phase of the engagement. It is shown that the impact angle is related to the value of the collision cone function at the time of interception, and this is used to develop conditions that enable interception with the target, while ensuring that the impact angle lies within a pre-defined range. The

NMPC problems are converted into QPs to be solved at each time-step. Both the QPs are shown to be strictly convex, and therefore easily solvable. Detailed simulation results that demonstrate the effectiveness of the proposed formulation, are presented.

Chapter 8

Closing Remarks

8.1 Conclusions

Online frameworks for performance optimization, estimation, and control of dynamical systems have been proposed in this dissertation, along with the application of one such framework in the context of avian-scale flapping. These frameworks involve set-theoretic concepts for their design and/or result in outcomes that have set-theoretic interpretations. Among these, two extremum seeking control schemes are proposed. The proposed schemes attenuate the steady-state oscillations and have the capability to converge (in practical asymptotic sense) to the global optimum, bypassing local extrema in the process. Then, a semi-analytical model for avian-scale (or bird-scale) forward flapping flight is proposed and, using the results of this model, an explanation of the unique range of Strouhal numbers used in bird-scale cruising flight is provided from a flight mechanics viewpoint. Further, a hypothesis, concerning the possible in-flight mechanism employed by birds to converge to the aforementioned unique (and optimal) Strouhal number range, is proposed. The model of flapping and extremum seeking control are combined to verify the hypothesis in simulations. Note that the proposed model (of flapping) and hypothesis can be useful for developing efficient bird-scale flapping aerial vehicles.

Apart from these, a novel nonlinear set-membership filter, termed SDC-SMF, is constructed by means of state dependent coefficient parameterization and Vetter calculus. It has been demonstrated that the SDC-SMF outperforms another existing set-membership filter for the state estimation of a nonlinear system governed by the

Van der Pol equation, despite the former being computationally cheaper compared to the latter. A linear version of SDC-SMF is utilized for the multi-agent leader-follower synchronization problem. The synchronization protocol, given appropriate conditions on the system matrices of the agents, the Riccati design, and the interaction graph are satisfied, makes the global error system ISS, i.e., the global disagreement error remains bounded. More importantly, due to the 'converging-input converging-state' property rendered by ISS systems, the disagreement error diminishes as the estimation errors reduce. Utilizing available bounds on the uncertainties, an upper bound for the global disagreement error norm is derived. It serves as a conservative performance measure of the protocol. Finally, a hybrid guidance algorithm that does not require target acceleration information for a successful capture has been developed based on nonlinear model predictive control and collision cone theory. Due to the use of model predictive control, constraints on the magnitude and rate of change of missile acceleration are taken into consideration explicitly. Also, collision cone theory enables us to incorporate appropriate constraints on the radial component of relative velocity vector to ensure target capture with a predefined range of impact angles. Detailed simulations are included to show successful capture for a variety of challenging initial engagement geometries and target acceleration profiles.

8.2 Future Directions

Some of the future directions of the research included in this dissertation are as described in the following subsections.

8.2.1 Extremum Seeking Control

The proposed schemes for ESC can be applied to a diverse range of problems where the system to be optimized is hard to model mathematically and steady-state oscillations are not permitted. Also, possible extension of the proposed ESC frameworks to ensure finite-time convergence would be a worthwhile pursuit.

8.2.2 Avian-Scale Flapping Flight

The proposed formulation can be utilized to perform a parametric analysis with the phase angle between plunging and twisting motions as the parameter. Another possible extension can be to allow the flow separation point to be dynamic during the flapping cycle and study the resulting effects. Effects of aeroelasticity and application of control theory would also be worth investigating.

8.2.3 Set-Membership Filtering

The task of assessing theoretical properties of the SDC-SMF for systems with control inputs acting through a (possibly non-square) state dependent matrix can be pursued. In addition, one can perform a detailed study comparing performances of the proposed framework with those of other nonlinear estimation approaches.

8.2.4 Multi-Agent Synchronization With Set-Membership Filtering

Extending the proposed synchronization formulation to discrete-time nonlinear dynamical systems and switching network topologies will be novel contributions. Also, a more fundamental shift involves extending the results in this dissertation by considering a control input for the leader or the leader to be any bounded reference trajectory.

8.2.5 Missile Guidance Using Model Predictive Control

The present formulation can be extended to the three dimensional engagement geometries. An elaborate numerical analysis, with different initial engagement geometries and/or with different target maneuverabilities, can be performed to characterize the sets of initial engagement geometries and target maneuverabilities for which capture is guaranteed with all the guidance objectives satisfied.

APPENDIX A

Mathematical Preliminaries

In this appendix, we outline some of the important mathematical preliminaries required in the context of this dissertation.

A.1 Systems with Two Time Scales and Singular Perturbation Theory

Let us consider a system which involves some of its states evolving faster (in time) compared to the rest due to the existence of a small parameter, i.e., there are two time scales, the separation between which is characterized by a small parameter, associated with the time evolution of the system states. This scenario is frequently encountered in engineering problems (see, for example, [184,185]). Singular perturbation theory offers a systematic approach for analyzing systems admitting this kind of behavior by decomposing the original system into two different limiting systems (or reduced-order subsystems) with distinct time scales. The term 'singular' essentially captures the fundamental shift in the nature of governing equations as the parameter characterizing the time scale separation (sometimes called the perturbation parameter [185]) is set equal to zero. A concise description of the singular perturbation approach is summarized in the following subsections.

A.1.1 Standard/Classical Singular Perturbation

In this subsection, we adopt the singular perturbation setup given in [186]. Let us consider a nonlinear dynamical system given by

$$\begin{aligned} \dot{\boldsymbol{x}}_1 &= \boldsymbol{f}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}, \epsilon) \\ \epsilon\dot{\boldsymbol{x}}_2 &= \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}, \epsilon) \end{aligned} \tag{A.1}$$

where $\epsilon > 0$ is a small parameter (the perturbation parameter), $\boldsymbol{x}_1 \in \mathbb{R}^{n_1}$, $\boldsymbol{x}_2 \in \mathbb{R}^{n_2}$ are the state vectors, and $\boldsymbol{d} \in \mathbb{R}^m$ is an input vector (representing system parameters, exogeneous disturbances and/or tracking signals [186]). In the above, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$

satisfy some smoothness properties (for example, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are locally Lipschitz on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m \times [0, \bar{\epsilon})$ for some $\bar{\epsilon} > 0$ [186]). The decomposition of the above system is achieved by setting $\epsilon = 0$ and expressing the above system as

$$\begin{aligned} \dot{\boldsymbol{x}}_1 &= \boldsymbol{f}_1(\boldsymbol{x}_1, \boldsymbol{x}_{2_s}, \boldsymbol{d}, 0) \\ \boldsymbol{0}_{n_2} &= \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_{2_s}, \boldsymbol{d}, 0) \end{aligned} \tag{A.2}$$

where $\boldsymbol{x}_{2_s}$ is a quasi-steady description of the fast evolving state $\boldsymbol{x}_2$. Note the drastic change in the dynamical properties of the system resulting from setting $\epsilon = 0$ as the differential equation $\epsilon\dot{\boldsymbol{x}}_2 = \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}, \epsilon)$ degenerates into the algebraic or transcendental equation $\boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_{2_s}, \boldsymbol{d}, 0) = \boldsymbol{0}_{n_2}$. Now, the system (A.2) is in *standard form* if the following assumption is satisfied.

Assumption A.1.1 ([186]). *The algebraic equation $\boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_{2_s}, \boldsymbol{d}, 0) = \boldsymbol{0}_{n_2}$ has a unique root*

$$\boldsymbol{x}_{2_s} = \boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{d}) \tag{A.3}$$

where $\boldsymbol{h} : \mathbb{R}^{n_1} \times \mathbb{R}^m \to \mathbb{R}^{n_2}$ and its partial derivatives are all locally Lipschitz.

Thus, we have the following limiting system

$$\dot{\boldsymbol{x}}_1 = \boldsymbol{f}_1(\boldsymbol{x}_1, \boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{d}), \boldsymbol{d}, 0) \tag{A.4}$$

which is termed the *reduced system* or the *slow subsystem.* On the other hand, the *fast subsystem* or the *boundary layer system* is given by

$$\frac{d\boldsymbol{z}}{dt'} = \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{d}) + \boldsymbol{z}, \boldsymbol{d}, 0) \tag{A.5}$$

where $t' = \frac{t}{\epsilon}$ is the fast time scale and $\boldsymbol{z} = \boldsymbol{x}_2 - \boldsymbol{h}(\boldsymbol{x}_1, \boldsymbol{d})$. Note that $\boldsymbol{x}_1$ and $\boldsymbol{d}$ are treated as constant vectors in the boundary layer system. Then, the remaining task involves assessing stability properties of the original system (A.1) from the stability properties admitted by the limiting systems (see, for example, [184–186]).

A.1.2 Generalized Singular Perturbation

Adopting the generalized singular perturbation setup given in [187], let us consider a system of the form

$$\begin{aligned}\dot{\boldsymbol{x}}_1 &= \boldsymbol{f}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}_1(t), \boldsymbol{d}_2(t), \epsilon)\\ \dot{\boldsymbol{x}}_2 &= \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}_1(t), \boldsymbol{d}_2(t), \epsilon)\end{aligned} \tag{A.6}$$

where ϵ denotes a small positive parameter, $\boldsymbol{x}_1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{x}_2 \in \mathbb{R}^{n_2}$ are the vectors of slowly evolving states and fast evolving states, respectively. Note that $\boldsymbol{d}_1(t)$ and $\boldsymbol{d}_2(t)$ (which are functions of time and take values in $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$, respectively) denote the slowly varying and fast varying disturbances, respectively. Also in the above, both $\dot{\boldsymbol{x}}_1$ vanishes and the changes in $\boldsymbol{d}_1(t)$ over a finite time interval decrease to zero as $\epsilon \to 0$ [187, Section III]. By selecting $\epsilon = 0$ in this setting, we obtain the boundary layer system as

$$\begin{aligned}\dot{\boldsymbol{x}}_1 &= \boldsymbol{0}_{n_1}\\ \dot{\boldsymbol{x}}_2 &= \boldsymbol{f}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{d}_1, \boldsymbol{d}_2(t), 0)\\ \dot{\boldsymbol{d}}_1 &= \boldsymbol{0}_{m_1}.\end{aligned} \tag{A.7}$$

Also, the average or reduced system is defined as

$$\dot{\boldsymbol{x}}_1 = \epsilon \boldsymbol{f}_{1_{\mathrm{av}}}(\boldsymbol{x}_1, \boldsymbol{d}_1(t), \boldsymbol{d}_2(t), \boldsymbol{e}(t)) \tag{A.8}$$

where $\boldsymbol{f}_{1_{\mathrm{av}}}(\cdot,\cdot,\cdot,\cdot)$ is an admissible average as dictated by Definition 1 in [187] and $\boldsymbol{e}(t)$ is a fictitious disturbance signal which allows different possibilities to exist (e.g., the average depends upon the initial condition of the boundary layer system [187]). With that, generalized input-to-state stability estimates (or bounds) are assumed on the trajectories of the boundary layer and average systems. The main analysis then involves characterizing the input-to-state stability bounds on the trajectories of the original system and the underlying conditions for which such bounds hold (see, [187, Sections III & IV] for details).

A.2 Method of Averaging

We consider the method of averaging described in [185]. Consider a system given by

$$\dot{\boldsymbol{x}} = \epsilon \boldsymbol{f}(t, \boldsymbol{x}, \epsilon) \tag{A.9}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $\epsilon > 0$ is a small parameter, and $\boldsymbol{f}(t, \boldsymbol{x}, \epsilon)$ is periodic in t with a time period $T > 0$, i.e., we have

$$\boldsymbol{f}(t + T, \boldsymbol{x}, \epsilon) = \boldsymbol{f}(t, \boldsymbol{x}, \epsilon), \ \forall (t, \boldsymbol{x}, \epsilon) \in \mathbb{R}_{\geq 0} \times \mathbb{D} \times [0, \epsilon_0] \tag{A.10}$$

for some $\epsilon_0 > 0$ and domain $\mathbb{D} \subset \mathbb{R}^n$. The method of averaging involves defining an autonomous 'average system', characterized by an average of $\boldsymbol{f}(t, \boldsymbol{x}, \epsilon)$ at $\epsilon = 0$ [185]. Thus, the autonomous average system takes the following form:

$$\dot{\boldsymbol{x}} = \epsilon \boldsymbol{f}_{\mathrm{av}}(\boldsymbol{x}) \tag{A.11}$$

where

$$\boldsymbol{f}_{\mathrm{av}}(\boldsymbol{x}) = \frac{1}{T} \int_0^T \boldsymbol{f}(\tau, \boldsymbol{x}, 0) d\tau. \tag{A.12}$$

Subsequently, the behavior of the non-autonomous system (A.9) is analyzed through the behavior of the autonomous average system (A.11). In this regard, we state the following result:

Theorem A.2.1 ([185])**.** *Let $\boldsymbol{f}(t, \boldsymbol{x}, \epsilon)$ and its partial derivatives with respect to $(\boldsymbol{x}, \epsilon)$ up to the second order be continuous and bounded for $(t, \boldsymbol{x}, \epsilon) \in \mathbb{R}_{\geq 0} \times \mathbb{D}_0 \times [0, \epsilon_0]$ where $\epsilon_0 > 0$ and $\mathbb{D}_0$ is a compact set satisfying $\mathbb{D}_0 \subset \mathbb{D}$ with $\mathbb{D} \subset \mathbb{R}^n$ as a domain. Suppose $\boldsymbol{f}(t, \boldsymbol{x}, \epsilon)$ is periodic in t with a time period $T > 0$ and ϵ is a positive parameter. If $\boldsymbol{x}_\star \in \mathbb{D}$ is an exponentially stable equilibrium point of the average system* (A.11)*, then there exists a positive constant $\epsilon^\star$ such that for all $\epsilon \in (0, \epsilon^\star)$,* (A.9) *has a unique, exponentially stable T-periodic solution in an $O(\epsilon)$ neighborhood of $\boldsymbol{x}_\star$.*

A.3 Matrix Calculus (Vetter Calculus)

We adopt the matrix calculus formalism by Vetter in [165]. Specifically, we make use of the matrix Taylor expansion given in [165]. To that end, the derivative matrices of a matrix-valued function $\boldsymbol{M}(\boldsymbol{X}) = [m_{ij}] \in \mathbb{R}^{p\times q}$ (where $\boldsymbol{X} = [x_{ij}] \in \mathbb{R}^{s\times t}$ and m_{ij} are functions of x_{ij}) with respect to x_{ij} and $\boldsymbol{X}$ are given by [165]

$$\mathcal{D}_{x_{ij}}\boldsymbol{M}(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial m_{11}}{\partial x_{ij}} & \frac{\partial m_{12}}{\partial x_{ij}} & \cdots \\ \frac{\partial m_{21}}{\partial x_{ij}} & \frac{\partial m_{22}}{\partial x_{ij}} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}, \; \mathcal{D}_{\boldsymbol{X}}\boldsymbol{M}(\boldsymbol{X}) = \begin{bmatrix} \mathcal{D}_{x_{11}}\boldsymbol{A} & \mathcal{D}_{x_{12}}\boldsymbol{A} & \cdots \\ \mathcal{D}_{x_{21}}\boldsymbol{A} & \mathcal{D}_{x_{22}}\boldsymbol{A} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}. \tag{A.13}$$

Now, let us consider a matrix-valued function $\boldsymbol{M}(\boldsymbol{x}) = [m_{ij}] \in \mathbb{R}^{p\times q}$ of a vector $\boldsymbol{x} \in \mathbb{R}^n$. Then, the Taylor expansion of $\boldsymbol{M}(\boldsymbol{x})$ about $\bar{\boldsymbol{x}}$ is given by

$$\boldsymbol{M}(\boldsymbol{x}) = \boldsymbol{M}(\bar{\boldsymbol{x}}) + \sum_{i=1}^{N} \frac{1}{i!}\left(\mathcal{D}^i_{\boldsymbol{x}^{\mathrm{T}i}}\boldsymbol{M}(\bar{\boldsymbol{x}})\right)\left((\boldsymbol{x}-\bar{\boldsymbol{x}})^{\times i} \otimes \boldsymbol{I}_q\right) + \boldsymbol{R}_{N+1}(\bar{\boldsymbol{x}}, \boldsymbol{x}) \tag{A.14}$$

where $\mathcal{D}^i_{\boldsymbol{x}^{\mathrm{T}i}}\boldsymbol{M}(\bar{\boldsymbol{x}})$ are the derivative matrices (with the i-th derivative calculated as $\mathcal{D}^i_{\boldsymbol{x}^{\mathrm{T}i}}\boldsymbol{M}(\boldsymbol{x}) = \mathcal{D}_{\boldsymbol{x}^{\mathrm{T}}}\left(\mathcal{D}_{\boldsymbol{x}^{\mathrm{T}}}\cdots\left(\mathcal{D}^i_{\boldsymbol{x}^{\mathrm{T}i}}\boldsymbol{M}(\boldsymbol{x})\right)\right))$ evaluated at $\boldsymbol{x} = \bar{\boldsymbol{x}}$, $(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\times i}$ is the i-th Kronecker power of $(\boldsymbol{x}-\bar{\boldsymbol{x}})$ (i.e., $(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\times i} = (\boldsymbol{x}-\bar{\boldsymbol{x}}) \otimes (\boldsymbol{x}-\bar{\boldsymbol{x}}) \otimes \cdots \otimes (\boldsymbol{x}-\bar{\boldsymbol{x}})$ (i factors)), and $\boldsymbol{R}_{N+1}(\bar{\boldsymbol{x}}, \boldsymbol{x})$ is the remainder term (matrix) [165]. For example, consider the case of $p = q = n = 2$ with $\boldsymbol{x} = [x_1 \quad x_2]^{\mathrm{T}}$ and $\boldsymbol{\delta x} = \boldsymbol{x} - \bar{\boldsymbol{x}} = [\delta x_1 \quad \delta x_2]^{\mathrm{T}}$. Then, we have the following expansion:

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} \bar{m}_{11} & \bar{m}_{12} \\ \bar{m}_{21} & \bar{m}_{22} \end{bmatrix} + \begin{bmatrix} \mathcal{D}_{x_1}m_{11} & \mathcal{D}_{x_1}m_{12} & \mathcal{D}_{x_2}m_{11} & \mathcal{D}_{x_2}m_{12} \\ \mathcal{D}_{x_1}m_{21} & \mathcal{D}_{x_1}m_{22} & \mathcal{D}_{x_2}m_{21} & \mathcal{D}_{x_2}m_{22} \end{bmatrix} \begin{bmatrix} \delta x_1 & 0 \\ 0 & \delta x_1 \\ \delta x_2 & 0 \\ 0 & \delta x_2 \end{bmatrix}$$

$$+ \boldsymbol{R}_2(\bar{\boldsymbol{x}}, \boldsymbol{x})$$

where $\bar{m}_{(\cdot)}$ are $m_{(\cdot)}$ evaluated at $\bar{\boldsymbol{x}}$ and $\mathcal{D}_{x_{(\cdot)}}m_{(\cdot)} = \frac{\partial m_{(\cdot)}}{\partial x_{(\cdot)}}$ are evaluated at $\bar{\boldsymbol{x}}$.

A.4 Numerical Optimization: Convexity and Relevant Aspects

Let us consider a general constrained optimization problem of the form

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}, \end{aligned} \tag{A.15}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the decision vector (or optimization variable), $f : \mathbb{R}^n \to \mathbb{R}$ is the cost function, and $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are formed using scalar functions $h_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, 2, \ldots, q$ and $g_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, 2, \ldots, m$, respectively, i.e.,

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{x}) &= \begin{bmatrix} h_1(\boldsymbol{x}) & h_2(\boldsymbol{x}) & \cdots & h_q(\boldsymbol{x}) \end{bmatrix}^{\mathrm{T}}, \\ \boldsymbol{g}(\boldsymbol{x}) &= \begin{bmatrix} g_1(\boldsymbol{x}) & g_2(\boldsymbol{x}) & \cdots & g_m(\boldsymbol{x}) \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

Thus, in the optimization problem (A.15), we have $h_i(\boldsymbol{x}) = 0, i = 1, 2, \ldots, q$ and $g_i(\boldsymbol{x}) \geq 0, i = 1, 2, \ldots, m$, where the equations $h_i(\boldsymbol{x}) = 0$ and the inequalities $g_i(\boldsymbol{x}) \geq 0$ are termed equality constraints and inequality constraints, respectively. Also, the functions $h_i(\boldsymbol{x})$ and $g_i(\boldsymbol{x})$ are called equality constraint functions and inequality constraint functions, respectively. If there are no constraints involved (i.e., the scenario where $q = m = 0$), we call the optimization problem *unconstrained.* Note that the optimization problem (A.15) can be equivalently expressed as

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f'(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}, \end{aligned} \tag{A.16}$$

where $f' = -f$ is the objective function.

Convexity is an inherently important property for optimization problems, and a convex optimization problem is easier to solve compared to a non-convex one. In the context of convex optimization, we are interested in both convex functions and sets, which are described next:

- *Convex set [188]*: A given set $\mathbb{C}$ is termed convex if a straight line segment joining any two points in the set lies entirely within the set. In other words, a set $\mathbb{C}$ is convex if for any two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}$, it holds that

$$\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \mathbb{C},\ \forall\alpha \in [0,1]. \tag{A.17}$$

  Simple examples of convex sets include a ball of radius r (i.e., the set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \leq r^2\}$) and any polytope or polyhedron (i.e., the set $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$).

- *Convex function [188]*: A function $f : \mathbb{C} \to \mathbb{R}$ is called convex if its domain $\mathbb{C}$ is convex and if for any two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}$ the following property holds:

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}),\ \forall\alpha \in [0,1]. \tag{A.18}$$

  For example, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}$ with $\boldsymbol{Q} \geq 0$ is a convex function. In a similar vein, a function f is called *concave* if $-f$ is convex.

The general constrained optimization problem (A.15) is a convex one if the following three requirements are satisfied [180, 188]:

1. the cost function is convex,
2. the equality constraint functions are affine, i.e., $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{b}_i$,
3. the inequality constraint functions are concave.

Therefore, convex optimization problems involve a convex cost function and a *convex feasible set.* A general convex optimization problem can be expressed as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathbb{X} \end{aligned} \tag{A.19}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbb{X}$ are convex. Now, we state the following important property with regards to convex optimization problems.

Theorem A.4.1 ([180,188]). *Let $\boldsymbol{x}^\star$ be a locally optimal point of a convex optimization problem* (A.19)*. Then, it is indeed the global optimal point.*

Interior-point methods are powerful tools for solving a diverse range of optimization problems [188] and we briefly describe the working principles of a basic interior-point algorithm next.

A.4.1 A Basic Interior-Point Algorithm [188]

Consider the general constrained optimization problem (A.15). To relax the inequality constraints, a vector of slack variables $\boldsymbol{s} \in \mathbb{R}^m$ is introduced such that we have the approximated problem as follows:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{s}\in\mathbb{R}^m} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \\ & \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} = \boldsymbol{0}, \\ & \boldsymbol{s} \geq \boldsymbol{0}, \end{aligned} \tag{A.20}$$

where the last inequality means $s_i \geq 0, i = 1, 2, \ldots, m$. This optimization problem (A.20) can be further modified using *a logarithmic barrier function* as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{s}\in\mathbb{R}^m} \quad & f(\boldsymbol{x}) - \kappa \sum_{i=1}^{m} \log s_i \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \\ & \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} = \boldsymbol{0}, \end{aligned} \tag{A.21}$$

where κ is a positive parameter (called the 'barrier parameter' [188]) and log stands for the natural logarithm. Note that the inequality $\boldsymbol{s} \geq \boldsymbol{0}$ is no longer required for the optimization problem (A.21) due to the minimization of the barrier function [188]. Let us define a Lagrangian function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^m \to \mathbb{R}$ as

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{l}_h, \boldsymbol{l}_g) = f(\boldsymbol{x}) - \kappa \sum_{i=1}^{m} \log s_i - \boldsymbol{l}_h^{\mathrm{T}} \boldsymbol{h}(\boldsymbol{x}) - \boldsymbol{l}_g^{\mathrm{T}} \left(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s}\right) \tag{A.22}$$

where $\boldsymbol{l}_h \in \mathbb{R}^q$ and $\boldsymbol{l}_g \in \mathbb{R}^m$ are the Lagrange multiplies associated with the constraints $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$ and $\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} = \boldsymbol{0}$, respectively. Then, the Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (A.21) are given by [188]

$$\begin{aligned} \nabla f(\boldsymbol{x}) - \left(\boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x})\right)^{\mathrm{T}} \boldsymbol{l}_h - \left(\boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x})\right)^{\mathrm{T}} \boldsymbol{l}_g &= \boldsymbol{0}, \\ -\kappa \boldsymbol{S}^{-1}\boldsymbol{1} + \boldsymbol{l}_g &= \boldsymbol{0}, \\ \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0}, \\ \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} &= \boldsymbol{0}, \end{aligned} \tag{A.23}$$

where $\boldsymbol{S} = \operatorname{diag}(s_1, s_2, \ldots, s_m)$ and the Jacobian matrices are given by

$$\boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x}) = \begin{bmatrix} \left(\nabla h_1(\boldsymbol{x})\right)^{\mathrm{T}} \\ \left(\nabla h_2(\boldsymbol{x})\right)^{\mathrm{T}} \\ \vdots \\ \left(\nabla h_q(\boldsymbol{x})\right)^{\mathrm{T}} \end{bmatrix}, \quad \boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x}) = \begin{bmatrix} \left(\nabla g_1(\boldsymbol{x})\right)^{\mathrm{T}} \\ \left(\nabla g_2(\boldsymbol{x})\right)^{\mathrm{T}} \\ \vdots \\ \left(\nabla g_m(\boldsymbol{x})\right)^{\mathrm{T}} \end{bmatrix}.$$

The system of equations (A.23) can be equivalently expressed as

$$\begin{aligned} \nabla f(\boldsymbol{x}) - \left(\boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x})\right)^{\mathrm{T}} \boldsymbol{l}_h - \left(\boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x})\right)^{\mathrm{T}} \boldsymbol{l}_g &= \boldsymbol{0}, \\ -\kappa \boldsymbol{1} + \boldsymbol{S}\boldsymbol{l}_g &= \boldsymbol{0}, \\ \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0}, \\ \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} &= \boldsymbol{0}, \end{aligned} \tag{A.24}$$

where $-\kappa \boldsymbol{S}^{-1}\boldsymbol{1} + \boldsymbol{l}_g = \boldsymbol{0}$ is converted into $-\kappa \boldsymbol{1} + \boldsymbol{S}\boldsymbol{l}_g = \boldsymbol{0}$, a procedure that does not affect the solution of the system of equations (A.23) [188]. Also, the system of equations (A.24) is called the 'perturbed KKT system' [188].

Newton's method is a simple and well-known root-finding technique for nonlinear equations of the form $\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}$. In this method, it is assumed that $\boldsymbol{F}(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx \boldsymbol{F}(\boldsymbol{x}) + \boldsymbol{J}_{\boldsymbol{F}}(\boldsymbol{x})\Delta\boldsymbol{x} = \mathbf{0}$, which leads to

$$\Delta\boldsymbol{x} = -\left(\boldsymbol{J}_{\boldsymbol{F}}(\boldsymbol{x})\right)^{-1}\boldsymbol{F}(\boldsymbol{x}).$$

Next, Newton's method is applied to the system of equations (A.24) to obtain search directions for the algorithm. Let us assume that the current iterate is $(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{l}_h, \boldsymbol{l}_g)$. Then, the current search direction is calculated using [188]

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L} & \mathbf{0} & -\left(\boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x})\right)^{\mathrm{T}} & -\left(\boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x})\right)^{\mathrm{T}} \\ \mathbf{0} & \boldsymbol{L}_g & \mathbf{0} & \boldsymbol{S} \\ \boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x}) & -\boldsymbol{I} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{s} \\ \Delta\boldsymbol{l}_h \\ \Delta\boldsymbol{l}_g \end{bmatrix} = -\begin{bmatrix} \nabla f(\boldsymbol{x}) - \left(\boldsymbol{J}_{\boldsymbol{h}}(\boldsymbol{x})\right)^{\mathrm{T}}\boldsymbol{l}_h - \left(\boldsymbol{J}_{\boldsymbol{g}}(\boldsymbol{x})\right)^{\mathrm{T}}\boldsymbol{l}_g \\ -\kappa\mathbf{1} + \boldsymbol{S}\boldsymbol{l}_g \\ \boldsymbol{h}(\boldsymbol{x}) \\ \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{s} \end{bmatrix} \tag{A.25}$$

where $\boldsymbol{L}_g = \operatorname{diag}\left(l_{g_1}, l_{g_2}, \ldots, l_{g_m}\right)$. Note that the system of equations (A.25) is termed the *primal-dual system* [188]. Once the solution $\begin{bmatrix}(\Delta\boldsymbol{x})^{\mathrm{T}} & (\Delta\boldsymbol{s})^{\mathrm{T}} & (\Delta\boldsymbol{l}_h)^{\mathrm{T}} & (\Delta\boldsymbol{l}_g)^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ is obtained, an updated iterate is computed as [188]

$$\boldsymbol{x}^+ = \boldsymbol{x} + \alpha_1\Delta\boldsymbol{x}, \quad \boldsymbol{s}^+ = \boldsymbol{s} + \alpha_1\Delta\boldsymbol{s},$$

$$\boldsymbol{l}_h^+ = \boldsymbol{l}_h + \alpha_2\Delta\boldsymbol{l}_h, \quad \boldsymbol{l}_g^+ = \boldsymbol{l}_g + \alpha_2\Delta\boldsymbol{l}_g,$$

where α_1, α_2 are the step sizes that can be determined using the method provided in [188]. The iterations are repeated until a stopping criteria has been satisfied (e.g., convergence within a tolerance, maximum number of iterations reached, and so on). Although different modifications are needed to handle various issues, this simple iterative process forms the basis of modern interior-point methods [188].

A.4.2 Useful Results

Following are two useful results in the context of set-membership filtering:

Lemma A.4.2. *(S-procedure [52, 189]) Let $F_0, F_1, \ldots, F_q$ be quadratic functions of the variable $\boldsymbol{\xi} \in \mathbb{R}^n$, given by*

$$F_i(\boldsymbol{\xi}) = \boldsymbol{\xi}^{\mathrm{T}} \boldsymbol{T}_i \boldsymbol{\xi}, \ i = 0, 1, 2, \ldots, q \tag{A.26}$$

where $\boldsymbol{T}_i = \boldsymbol{T}_i^{\mathrm{T}}$. Then, the following condition

$$F_0(\boldsymbol{\xi}) \leq 0, \ \forall \boldsymbol{\xi} \text{ such that } F_i(\boldsymbol{\xi}) \leq 0, \ i = 1, 2, \ldots, q \tag{A.27}$$

holds if there exist $\tau_1 \geq 0$, $\tau_2 \geq 0, \ldots, \tau_q \geq 0$ such that

$$\boldsymbol{T}_0 - \sum_{i=1}^{q} \tau_i \boldsymbol{T}_i \leq 0. \tag{A.28}$$

Lemma A.4.3. *(Schur complements [52,189]) Consider the given matrices $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3$ with $\boldsymbol{S}_1 = \boldsymbol{S}_1^{\mathrm{T}}$ and $\boldsymbol{S}_3 < 0$. Then, $\boldsymbol{S}_1 - \boldsymbol{S}_2^{\mathrm{T}} \boldsymbol{S}_3^{-1} \boldsymbol{S}_2 \leq 0$ if and only if*

$$\begin{bmatrix} \boldsymbol{S}_3 & \boldsymbol{S}_2 \\ \boldsymbol{S}_2^{\mathrm{T}} & \boldsymbol{S}_1 \end{bmatrix} \leq 0. \tag{A.29}$$

A.5 Input-to-State Stability

Input-to-state stability of a system loosely translates to the fact that every trajectory corresponding to a bounded control input remains bounded and that the trajectory eventually becomes small if the control inputs are small irrespective of the initial state [172]. In other words, input-to-state stability implies bounded input bounded state (BIBS) stability of a system [185]. For the remainder of this section, the symbol $|\cdot|$ denotes standard Euclidean norm for vectors. Further, for any function $\boldsymbol{\theta} : \mathbb{Z}_\star \to \mathbb{R}^n$, we have $||\boldsymbol{\theta}|| = \sup\{|\boldsymbol{\theta}_k| : k \in \mathbb{Z}_\star\}$. This is the standard l_∞ norm for a bounded $\boldsymbol{\theta}$.

Definition A.5.1 ([172,190]). *A function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class $\mathcal{K}$ function if it is continuous, strictly increasing and $\gamma(0) = 0$. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class $\mathcal{KL}$ function if, for each fixed $t \geq 0$, the function $\beta(\cdot, t)$ is a class $\mathcal{K}$ function and for each fixed $s \geq 0$, the function $\beta(s, \cdot)$ is decreasing and $\beta(s, t) \to 0$ as $t \to \infty$.*

Next, we introduce the notion of input-to-state stability in the following definition.

Definition A.5.2. *A discrete-time system of the form $\boldsymbol{x}_{k+1} = \boldsymbol{\phi}(\boldsymbol{x}_k, \boldsymbol{u}_{1_k}, \boldsymbol{u}_{2_k}),\ k \in \mathbb{Z}_\star$ with $\boldsymbol{u}_1 : \mathbb{Z}_\star \to \mathbb{R}^{m_1}$, $\boldsymbol{u}_2 : \mathbb{Z}_\star \to \mathbb{R}^{m_2}$, $\boldsymbol{\phi}(\mathbf{0}_n, \mathbf{0}_{m_1}, \mathbf{0}_{m_2}) = \mathbf{0}_n$ is (globally) input-to-state stable (ISS) if there exist a class $\mathcal{KL}$ function β and two class $\mathcal{K}$ functions γ_1, γ_2 such that, for each pair of inputs $\boldsymbol{u}_1 \in l_\infty^{m_1}$, $\boldsymbol{u}_2 \in l_\infty^{m_2}$ and each $\boldsymbol{x}_0 \in \mathbb{R}^n$, it holds that*

$$|\boldsymbol{x}_k| \leq \beta(|\boldsymbol{x}_0|, k) + \gamma_1(||\boldsymbol{u}_1||) + \gamma_2(||\boldsymbol{u}_2||) \tag{A.30}$$

for each $k \in \mathbb{Z}_\star$.

Note that Definition A.5.2 is adopted from Definition 3.1 in [172] and has been suitably modified for systems with two inputs using Definition IV.3 in [190]. The ISS property (A.30) means that the origin of the unforced system $\boldsymbol{x}_{k+1} = \boldsymbol{\phi}(\boldsymbol{x}_k, \mathbf{0}_{m_1}, \mathbf{0}_{m_2}),\ k \in \mathbb{Z}_\star$ is globally asymptotically stable [172]. Also, an ISS system admits the so-called 'converging-input converging-state' property which means that every trajectory of the system converges to the origin if $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ vanish as $k \to \infty$ [172].

APPENDIX B

A Sketch of the Proof for Theorem 2.2.5

Our proof of the result in Theorem 2.2.5 is inspired by the sketch of the proof for Theorem 2 in Nesic et al. [8]. Therefore, we utilize the ideas given in Desoer and Shahruz [191] and the generalized singular perturbation framework given in Teel et al. [187]. First, we express the closed-loop system (2.8) as

$$\begin{aligned}
\dot{\boldsymbol{x}} &= \boldsymbol{f}\left(\boldsymbol{x}, \alpha(\boldsymbol{x}, \tilde{\theta} + \theta^\star + a\sin\omega t)\right), \\
\begin{bmatrix} \dot{\tilde{\theta}} \\ \dot{\xi} \\ \dot{\tilde{\eta}} \end{bmatrix} &= \omega \begin{bmatrix} \delta K' \xi \\ -\delta\omega_L' \xi + \delta\omega_L'(h(\boldsymbol{x}) - \tilde{\eta} - J(\theta^\star))\sin\omega t \\ -\delta\omega_H' \tilde{\eta} + \delta\omega_H'(h(\boldsymbol{x}) - J(\theta^\star)) \end{bmatrix}, \\
\dot{a} &= -\omega\epsilon\left(\delta\lambda_1' \; g_1(a) \; \exp(-\gamma_1|\hat{J}'(\tilde{\theta} + \theta^\star)|)\right).
\end{aligned} \tag{B.1}$$

It is clear that the system admits three time scales with ω and ϵ as the singular perturbation parameters. The requirement for δ to be sufficiently small will be established in the subsequent analysis (cf. Proposition B.1). Let us begin by putting the closed-loop system in (B.1) into the form of (23) in Teel et al. [187]. In doing so, we get the following system.

$$\begin{aligned}
\dot{\boldsymbol{x}}_{1_{s_1}} &= \boldsymbol{F}_{s_1}(\boldsymbol{x}_{s_1}, x_{s_2}, \boldsymbol{x}_f, \omega), \quad \dot{x}_{2_{s_1}} = \omega, \\
\dot{x}_{s_2} &= F_{s_2}(\boldsymbol{x}_{s_1}, x_{s_2}, \omega, \epsilon), \\
\dot{\boldsymbol{x}}_f &= \boldsymbol{F}_f(\boldsymbol{x}_{s_1}, x_{s_2}, \boldsymbol{x}_f),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{F}_{s_1}(\cdot) &= \omega \begin{bmatrix} \delta K' \xi \\ -\delta\omega_L' \xi + \delta\omega_L'(h(\boldsymbol{x}) - \tilde{\eta} - J(\theta^\star))\sin x_{2_{s_1}} \\ -\delta\omega_H' \tilde{\eta} + \delta\omega_H'(h(\boldsymbol{x}) - J(\theta^\star)) \end{bmatrix}, \\
\boldsymbol{x}_{s_1} &= (\tilde{\theta}, \xi, \tilde{\eta}, x_{2_{s_1}}) = (\boldsymbol{x}_{1_{s_1}}, x_{2_{s_1}}), \; x_{s_2} = a, \; \boldsymbol{x}_f = \boldsymbol{x}, \\
F_{s_2}(\cdot) &= -\omega\epsilon\left(\delta\lambda_1' \; g_1(a) \; \exp(-\gamma_1|\hat{J}'(\tilde{\theta} + \theta^\star)|)\right), \\
\boldsymbol{F}_f(\cdot) &= \boldsymbol{f}\left(\boldsymbol{x}, \alpha(\boldsymbol{x}, \tilde{\theta} + \theta^\star + a\sin x_{2_{s_1}})\right).
\end{aligned}$$

Note that the state $x_{2_{s_1}}$ denotes time in the slower time scale ωt. We define the boundary layer system as

$$\dot{\boldsymbol{x}} = \boldsymbol{F}_f(\boldsymbol{x}_{s_1}, x_{s_2}, \boldsymbol{x}),\ \dot{\boldsymbol{x}}_{s_1} = 0, \dot{x}_{s_2} = 0. \tag{B.2}$$

Let $\boldsymbol{x}_{\mathrm{bl}}(t) = (\boldsymbol{x}(t), \boldsymbol{x}_{s_1}(t_0), x_{s_2}(t_0))$ denote the solution to the boundary layer system (B.2) starting at time $t = t_0$. Next, we define the "average or reduced" function-1 based on Remark 15 in Teel et al. [187]. This is achieved by "freezing" $\boldsymbol{x}$ at its equilibrium value $\boldsymbol{x} = \boldsymbol{l}(\tilde{\theta}+\theta^\star+a\sin x_{2_{s_1}})$, fixing $x_{s_2} = a$ at its initial value $a(t_0) = a_0$, and taking the following limit

$$\begin{aligned}
&\boldsymbol{F}_{\mathrm{av}_1}(\boldsymbol{x}_{s_1}, a_0)\\
&= \lim_{\omega\to 0} \omega^{-1}\boldsymbol{F}_s(\boldsymbol{x}_{s_1}, a_0, \boldsymbol{l}(\tilde{\theta} + \theta^\star + a_0 \sin x_{2_{s_1}}), \omega),\\
&= \begin{bmatrix} \delta K'\xi \\ -\delta\omega_L'\xi + \delta\omega_L'(\nu(\tilde{\theta} + a_0 \sin x_{2_s}) - \tilde{\eta})\sin x_{2_{s_1}} \\ -\delta\omega_H'\tilde{\eta} + \delta\omega_H'\nu(\tilde{\theta} + a_0\sin x_{2_{s_1}}) \\ 1 \end{bmatrix},
\end{aligned}$$

where $\boldsymbol{F}_s(\cdot) = (\boldsymbol{F}_{s_1}(\cdot), \omega)$, $\nu(\tilde{\theta} + a_0 \sin x_{2_{s_1}}) = J(\tilde{\theta} + \theta^\star + a_0 \sin x_{2_{s_1}}) - J(\theta^\star)$. With our Assumption 2.1.3, we have the following results.

$$\nu(0) = 0,\ \nu'(0) = J'(\theta^\star) = 0,\ \nu''(0) = J''(\theta^\star) < 0.$$

Thus, we define the "average or reduced" system-1 as

$$\dot{\boldsymbol{x}}_{s_1} = \omega \boldsymbol{F}_{\mathrm{av}_1}(\boldsymbol{x}_{s_1}, a_0), \quad \dot{x}_{s_2} = 0. \tag{B.3}$$

Let $\boldsymbol{x}_{r_1}(t) = (\boldsymbol{x}_{s_1}(t), x_{s_2}(t_0))$ denote the solution to this system starting at time $t = t_0$. We state the following result about this system.

Proposition B.1. *Consider the system* (B.3) *under the Assumption 2.1.3. There exist positive constants $\bar{a}_1$ and $\bar{\delta}_1$ such that for all $a_0 \in (0, \bar{a}_1)$ and $\delta \in (0, \bar{\delta}_1)$ the* *solutions* $\boldsymbol{x}_{1_{s_1}}(t) = \Big(\tilde{\theta}(t), \xi(t), \tilde{\eta}(t)\Big)$ *exponentially converge to a unique* $\left(\frac{2\pi}{\omega}\right)$*-periodic solution* $\boldsymbol{x}^p_{1_{s_1}}(t, a_0) = \Big(\tilde{\theta}^p(t), \xi^p(t), \tilde{\eta}^p(t)\Big)$ *satisfying*

$$\left\| \begin{bmatrix} \tilde{\theta}^p(t) + \frac{\nu'''(0)}{8\nu''(0)} a_0^2 \\ \xi^p(t) \\ \tilde{\eta}^p(t) - \frac{\nu''(0)}{4} a_0^2 \end{bmatrix} \right\| \leq O(\delta) + O(a_0^3), \tag{B.4}$$

for all $t \geq t_0 \geq 0$ *and all* $\boldsymbol{x}_{1_{s_1}}(t_0) \in \mathcal{B}_{\boldsymbol{x}_{1_{s_1}}}$ *where* $\mathcal{B}_{\boldsymbol{x}_{1_{s_1}}}$ *is a closed ball centered at* $\Big(\tilde{\theta}(t), \xi(t), \tilde{\eta}(t)\Big) = \Big(-\frac{\nu'''(0)}{8\nu''(0)} a_0^2 + O(a_0^3), 0, \frac{\nu''(0)}{4} a_0^2 + O(a_0^3)\Big)$ *and contains a neighborhood of the origin* $\Big(\tilde{\theta}(t), \xi(t), \tilde{\eta}(t)\Big) = (0, 0, 0)$.

Proof. The system (B.3) can be equivalently expressed in the slower time scale $\tau = \omega t$ as

$$\frac{d}{d\tau} \begin{bmatrix} \tilde{\theta} \\ \xi \\ \tilde{\eta} \end{bmatrix} = \begin{bmatrix} \delta K' \xi \\ -\delta \omega_L' \xi + \delta \omega_L' (\nu(\tilde{\theta} + a_0 \sin \tau) - \tilde{\eta}) \sin \tau \\ -\delta \omega_H' \tilde{\eta} + \delta \omega_H' \nu(\tilde{\theta} + a_0 \sin \tau) \end{bmatrix}.$$

The rest of the proof follows from Krstic and Wang [5, Section 4] and noting that $t = \frac{\tau}{\omega}$. □

Finally, let us define the "average or reduced" system-2, utilizing the Remark 15 in Teel et al. [187], as

$$\dot{a} = -\omega \epsilon \delta \lambda_1' \ g_1(a) \ \exp(-\gamma_1 |\hat{J}'(\tilde{\theta}^p(t))|), \tag{B.5}$$

where we have dropped $\theta^\star$ from the argument of $\hat{J}'(\cdot)$ as it is a constant. Let $a(t)$ denote the solution to this system starting at $a(t_0) = a_0$. Now, let us investigate the assumptions in Teel et al. [187]. It is easy to check that the Assumptions 1 and 2 in Teel et al. [187] are satisfied. Assumption 3 in Teel et al. [187] is satisfied since solutions $\boldsymbol{x}(t)$ of (B.2) locally exponentially converge to the equilibrium $\boldsymbol{l}(\tilde{\theta}(t_0) + \theta^\star + a_0 \sin x_{2_{s_1}}(t_0)) = \boldsymbol{l}(\theta(t_0))$, uniformly in $\theta(t_0)$ (due to our Assumptions 2.1.1 and

2.1.2). With that, we choose $\omega_{f,o}(\boldsymbol{x}_{\mathrm{bl}}(t)) = |\boldsymbol{x}(t) - \boldsymbol{l}(\tilde{\theta}(t_0) + \theta^\star + a_0 \sin x_{2_{s_1}}(t_0))|$ and $\beta_f(\omega_{f,o}(\boldsymbol{x}_{\mathrm{bl}}(t_0)), t) = k_1 \exp(-\alpha_1(t - t_0))|\boldsymbol{x}(t_0) - \boldsymbol{l}(\tilde{\theta}(t_0) + \theta^\star + a_0 \sin x_{2_{s_1}}(t_0))|$ with some $k_1, \alpha_1 > 0$. Also, we take $\mathcal{H}_f = \mathcal{B}_{\boldsymbol{x}} \times \mathcal{B}_{\boldsymbol{x}_{1_{s_1}}} \times [0, \infty) \times (0, \bar{a}_1)$ where $\mathcal{B}_{\boldsymbol{x}}$ is a closed ball centered at $\boldsymbol{l}(\tilde{\theta}(t_0) + \theta^\star + a_0 \sin x_{2_{s_1}}(t_0))$ and containing a neighborhood of the point $\boldsymbol{l}(\theta^\star)$, $\mathcal{B}_{\boldsymbol{x}_{1_{s_1}}}$ and $\bar{a}_1$ are as in Proposition B.1. Similarly, Assumption 4 in Teel et al. [187] is satisfied with the solutions $\boldsymbol{x}_{1_{s_1}}(t)$ of (B.3) exponentially converging to $\boldsymbol{x}^p_{1_{s_1}}(t, a_0)$. Hence, we choose $\omega_{s_1,o}(\boldsymbol{x}_{r_1}(t)) = |\boldsymbol{x}_{1_{s_1}}(t) - \boldsymbol{x}^p_{1_{s_1}}(t, a_0)|$ and $\beta_{s_1}(\omega_{s_1,o}(\boldsymbol{x}_{r_1}(t_0)), \omega(t - t_0)) = k_2 \exp(-\alpha_2 \omega \delta (t - t_0))|\boldsymbol{x}_{1_{s_1}}(t_0) - \boldsymbol{x}^p_{1_{s_1}}(t, a_0)|$ with some $k_2, \alpha_2 > 0$. Again, we take $\mathcal{H}_{s_1} = \mathcal{B}_{\boldsymbol{x}_{1_{s_1}}} \times [0, \infty) \times (0, \bar{a}_1)$ where all the notations are the same as for $\mathcal{H}_f$. We need one additional condition for the system (B.5). From (B.5) and under our Assumption 2.1.4, it is obvious that there exists a class $\mathcal{KL}$ function β_a with $\beta_a(s, 0) = s$ such that $|a(t)| \leq \beta_a(|a(t_0)|, \omega\delta\epsilon(t - t_0))$ for an appropriate choice of γ_1. Therefore, we choose $\omega_{s_2,o}(a(t)) = |a(t)|$, $\beta_{s_2}(\omega_{s_2,o}(a(t_0)), \omega\epsilon(t - t_0)) = \beta_a(\omega_{s_2,o}(a(t_0)), \omega\delta\epsilon(t - t_0))$, and $\mathcal{H}_{s_2} = (0, \bar{a}_1)$. Note that all the input measuring functions in Teel et al. [187] are identically zero for our system as we do not consider any disturbances.

Introducing modifications to Assumptions 7 and 8 in Teel et al. [187] for our system are tedious and have been omitted. However, we point out that the basic properties desired by imposing the Assumptions 7 and 8 in Teel et al. [187] are staisfied. For Assumption 7 in Teel et al. [187], we take $\mathcal{K}_f = \mathcal{H}_f$, $\mathcal{K}_{s_1} = \mathcal{H}_{s_1}$, and $\mathcal{K}_{s_2} = \mathcal{H}_{s_2}$. Clearly, our measuring functions are bounded on these sets. Recurrence of the sets $\mathcal{K}_f$, $\mathcal{K}_{s_1}$, and $\mathcal{K}_{s_2}$ is guaranteed due to the choice of these sets, the results in Krstic and Wang [5, Theorem 5.1], and the monotonic decrease of $a(t)$. For Assumption 8 in Teel et al. [187], we require the functions describing the problem to be locally Lipschitz, which is satisfied by the assumptions on the functions $\boldsymbol{f}, h, \boldsymbol{l}, g_1$. Since the variable $x_{2_{s_1}}$ serves the purpose of time, the Lipschitz continuity of the functions $\boldsymbol{F}_f$, $\boldsymbol{F}_{\mathrm{av}_1}$ and $\boldsymbol{F}_{s_1}$ in the variable $x_{2_{s_1}}$ should be uniform. This is satisfied since $\boldsymbol{F}_f$, $\boldsymbol{F}_{\mathrm{av}_1}$ and $\boldsymbol{F}_{s_1}$ are periodic in $x_{2_{s_1}}$ (see Remarks 26-30 in Teel et al. [187]). Also, it is easy to verify the uniform continuity requirements on the measuring functions by making proper modifications.

With regards to the solutions of (B.1), let us define the following change of coordinates $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \boldsymbol{l}(\tilde{\theta}(t) + \theta^\star + a(t)\sin\omega t)$, $\tilde{\boldsymbol{x}}_{1_{s_1}}(t) = \boldsymbol{x}_{1_{s_1}}(t) - \boldsymbol{x}^p_{1_{s_1}}(t, a(t))$ where $\boldsymbol{x}^p_{1_{s_1}}(t, a(t))$ is the periodic solution from Proposition B.1, characterized by $a(t)$. Then, by introducing suitable modifications to the main result in Teel et al. [187], we conclude that (2.10) and (2.11) hold with $\boldsymbol{x}_{1_{s_1}} \equiv \boldsymbol{z}$ and $\boldsymbol{x}^p_{1_{s_1}} \equiv \boldsymbol{z}^p_1$. Further, from the dynamics of $\dot{a}$ in the system (B.1) and under our Assumption 2.1.4, we conclude that there exists a class $\mathcal{KL}$ function β_{a_1} with $\beta_{a_1}(s, 0) = s$ such that $a(t)$ satisfies (2.12) for all $\gamma_1 \in (0, \bar{\gamma}_1)$ with some $\bar{\gamma}_1 > 0$. This completes the proof. □

APPENDIX C

Expressions for the Terms in Equation (3.16)

$$
\begin{aligned}
I_{L_I} =& \int_0^b \Big[St_y^2\left(-\frac{J_1(a-St_y)}{(a-St_y)}\right)+(1+St_y^2)J_0(a-St_y)\Big]dy \\
=&\frac{1}{2b^2B_1^3}\Bigg[16bB_1St_0^2J_0(A_1+bB_1)+A_1\pi(-b^2B_1^2+8St_0^2)J_1(A_1)H_0(A_1) \\
&+J_1(A_1+bB_1)\Big(8b^2B_1^2St_0^2+\pi(A_1+bB_1)(b^2B_1^2-8St_0^2)H_0(A_1+bB_1)\Big) \\
&+A_1(b^2B_1^2-8St_0^2)J_0(A_1)(-2+\pi H_1(A_1)) \\
&+(A_1+bB_1)(b^2B_1^2-8St_0^2)J_0(A_1+bB_1)(2-\pi H_1(A_1+bB_1))\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
I_{L_{II}} =& \int_0^b \Big[St_y^2\left(\frac{J_1(a+St_y)}{(a+St_y)}\right)-(1+St_y^2)J_0(a+St_y)\Big]dy \\
=&\frac{1}{2b^2B_2^3}\Bigg[2A_1(b^2B_2^2-8St_0^2)J_0(A_1)-2(A_1+bB_2)(b^2B_2^2-8St_0^2)J_0(A_1+bB_2) \\
&-8bB_2St_0^2(bB_2J_1(A_1+bB_2)+2J_0(A_1+bB_2))+\pi(b^2B_2^2-8St_0^2) \\
&\times\Bigg(A_1J_1(A_1)H_0(A_1)-(A_1+bB_2)J_1(A_1+bB_2)H_0(A_1+bB_2) \\
&-A_1J_0(A_1)H_1(A_1)+(A_1+bB_2)J_0(A_1+bB_2)H_1(A_1+bB_2)\Bigg)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
F_{a_I} =& \int_0^b (x_a-0.25)(\frac{\pi}{4}\rho c^3)(St_y-\theta_0)\omega_f^2J_1(St_y)dy \\
=&(x_a-0.25)(\frac{\pi}{4}\rho c^3)\omega_f^2\left(\frac{b}{2St_0}\right)\Big(-\theta_0+\pi St_0J_1(2St_0)H_0(2St_0) \\
&+J_0(2St_0)\big(\theta_0-\pi St_0H_1(2St_0)\big)\Big)
\end{aligned}
$$

$$
\begin{aligned}
I_{D_I} =& \int_0^b \big((1+St_y^2)J_0(St_y)-St_yJ_1(St_y)\big)\,dy \\
=&\frac{1}{2}b\Big[J_1(2St_0)(4St_0-\pi H_0(2St_0))+J_0(2St_0)(2+\pi H_1(2St_0))\Big]
\end{aligned}
$$

$$
\begin{aligned}
I_{D_{II}} =& \int_0^b \Big[St_y^2\left(\frac{J_1(2a-St_y)}{(2a-St_y)}\right)-(1+St_y^2)J_0(2a-St_y)\Big]dy \\
=&\frac{1}{2b^2B_3^3}\Bigg[2A_2(b^2B_3^2-8St_0^2)J_0(A_2)-2(A_2+bB_3)(b^2B_3^2-8St_0^2)J_0(A_2+bB_3)
\end{aligned}
$$

$$
\begin{aligned}
&- 8bB_3St_0^2(bB_3J_1(A_2 + bB_3) + 2J_0(A_2 + bB_3)) + \pi(b^2B_3^2 - 8St_0^2) \\
&\times \Bigg(A_2J_1(A_2)H_0(A_2) - (A_2 + bB_3)J_1(A_2 + bB_3)H_0(A_2 + bB_3) \\
&- A_2J_0(A_2)H_1(A_2) + (A_2 + bB_3)J_0(A_2 + bB_3)H_1(A_2 + bB_3)\Bigg)\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
I_{D_{III}} =& \int_0^b \left[St_y^2\left(\frac{J_1(2a + St_y)}{(2a + St_y)}\right) - (1 + St_y^2)J_0(2a + St_y)\right]dy \\
=& \frac{1}{2b^2B_4^3}\Bigg[2A_2(b^2B_4^2 - 8St_0^2)J_0(A_2) - 2(A_2 + bB_4)(b^2B_4^2 - 8St_0^2)J_0(A_2 + bB_4) \\
&- 8bB_4St_0^2(bB_4J_1(A_2 + bB_4) + 2J_0(A_2 + bB_4)) + \pi(b^2B_4^2 - 8St_0^2) \\
&\times \Bigg(A_2J_1(A_2)H_0(A_2) - (A_2 + bB_4)J_1(A_2 + bB_4)H_0(A_2 + bB_4) \\
&- A_2J_0(A_2)H_1(A_2) + (A_2 + bB_4)J_0(A_2 + bB_4)H_1(A_2 + bB_4)\Bigg)\Bigg]
\end{aligned}
$$

APPENDIX D

Proof of Theorem 5.2.1

Utilizing the corrected state estimate in (5.8), the estimation error at the correction step is

$$\begin{aligned}
&\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} \\
=&(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) - \boldsymbol{L}_k(\boldsymbol{y}_k - \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\hat{\boldsymbol{x}}_{k|k-1}) \\
=&\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} - \boldsymbol{L}_k\Big[(\boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{K}_1\boldsymbol{\Delta}_1 + \boldsymbol{R}_{\boldsymbol{H}_2})(\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1}) \\
&\quad - \boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{v}_k\Big] \\
=&\left(\boldsymbol{E}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{H}(\hat{\boldsymbol{x}}_{k|k-1})\boldsymbol{E}_{k|k-1}\right)\boldsymbol{z}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{v}_k - \boldsymbol{L}_k\boldsymbol{K}_1\boldsymbol{\Delta}_1\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \\
&\quad - \boldsymbol{L}_k\boldsymbol{R}_{\boldsymbol{H}_2}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{K}_1\boldsymbol{\Delta}_1\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{L}_k\boldsymbol{R}_{\boldsymbol{H}_2}\hat{\boldsymbol{x}}_{k|k-1}
\end{aligned} \tag{D.1}$$

Denote the unknowns in (D.1) as

$$\begin{aligned}
\boldsymbol{\Delta}_3 &= \boldsymbol{\Delta}_1\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \in \mathbb{R}^{n^2} \\
\boldsymbol{\Delta}_4 &= \boldsymbol{R}_{\boldsymbol{H}_2}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \in \mathbb{R}^{p} \\
\boldsymbol{\Delta}_5 &= \boldsymbol{\Delta}_1\hat{\boldsymbol{x}}_{k|k-1} \in \mathbb{R}^{n^2} \\
\boldsymbol{\Delta}_6 &= \boldsymbol{R}_{\boldsymbol{H}_2}\hat{\boldsymbol{x}}_{k|k-1} \in \mathbb{R}^{p}.
\end{aligned} \tag{D.2}$$

Next, define a vector of all the unknowns in (D.1) as

$$\boldsymbol{\zeta} = \begin{bmatrix} 1 & \boldsymbol{z}_{k|k-1}^{\mathrm{T}} & \boldsymbol{v}_k^{\mathrm{T}} & \boldsymbol{\Delta}_3^{\mathrm{T}} & \boldsymbol{\Delta}_4^{\mathrm{T}} & \boldsymbol{\Delta}_5^{\mathrm{T}} & \boldsymbol{\Delta}_6^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}. \tag{D.3}$$

Therefore, the estimation error in (D.1) can be expressed in terms of $\boldsymbol{\zeta}$ as

$$\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{\Pi}_{k|k-1}\boldsymbol{\zeta} \tag{D.4}$$

where $\boldsymbol{\Pi}_{k|k-1}$ is as shown in (5.22). Now, $\boldsymbol{x}_k \in \mathcal{E}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k})$ can be expressed as

$$\boldsymbol{\zeta}^{\mathrm{T}}\Big[\boldsymbol{\Pi}_{k|k-1}^{\mathrm{T}}\boldsymbol{P}_{k|k}^{-1}\boldsymbol{\Pi}_{k|k-1} - \mathrm{diag}(1, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p)\Big]\boldsymbol{\zeta} \leq 0. \tag{D.5}$$

Using the definition of $\boldsymbol{\Delta}_1$, it can be shown that $||\boldsymbol{\Delta}_1|| \leq \gamma_{k|k-1}$ (Similar to (5.15)). With that, the following inequalities hold:

$$\begin{cases}
\boldsymbol{\Delta}_3^{\mathrm{T}}\boldsymbol{\Delta}_3 = & \boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{\Delta}_1^{\mathrm{T}}\boldsymbol{\Delta}_1\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \leq \gamma_{k|k-1}^2\boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1}, \\
\boldsymbol{\Delta}_5^{\mathrm{T}}\boldsymbol{\Delta}_5 = & \hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\boldsymbol{\Delta}_1^{\mathrm{T}}\boldsymbol{\Delta}_1\hat{\boldsymbol{x}}_{k|k-1} \leq \gamma_{k|k-1}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1}.
\end{cases}$$

Similarly, utilizing the upper bound on the norm of remainder $\boldsymbol{R}_{\boldsymbol{H}_2}$, the following inequalities are derived:

$$\begin{cases} \boldsymbol{\Delta}_4^{\mathrm{T}}\boldsymbol{\Delta}_4 = & \boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{H}_2}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{H}_2}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \leq r_{H_k}^2\boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1}, \\ \boldsymbol{\Delta}_6^{\mathrm{T}}\boldsymbol{\Delta}_6 = & \hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{H}_2}^{\mathrm{T}}\boldsymbol{R}_{\boldsymbol{H}_2}\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}} \leq r_{H_k}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1}. \end{cases}$$

Therefore, all the unknowns in $\boldsymbol{\zeta}$ should satisfy the following inequalities

$$\begin{cases} \boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{z}_{k|k-1} - 1 \leq 0, \\ \boldsymbol{v}_k^T\boldsymbol{R}_k^{-1}\boldsymbol{v}_k - 1 \leq 0, \\ \boldsymbol{\Delta}_3^{\mathrm{T}}\boldsymbol{\Delta}_3 - \gamma_{k|k-1}^2\boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \leq 0, \\ \boldsymbol{\Delta}_4^{\mathrm{T}}\boldsymbol{\Delta}_4 - r_{H_k}^2\boldsymbol{z}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}\boldsymbol{z}_{k|k-1} \leq 0, \\ \boldsymbol{\Delta}_5^{\mathrm{T}}\boldsymbol{\Delta}_5 - \gamma_{k|k-1}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1} \leq 0, \\ \boldsymbol{\Delta}_6^{\mathrm{T}}\boldsymbol{\Delta}_6 - r_{H_k}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1} \leq 0. \end{cases}$$

The above inequalities are expressed in terms of $\boldsymbol{\zeta}$ as follows

$$\begin{cases} \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-1, \boldsymbol{I}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p)\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-1, \boldsymbol{O}_n, \boldsymbol{R}_k^{-1}, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p)\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(0, -\gamma_{k|k-1}^2\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}, \boldsymbol{O}_p, \boldsymbol{I}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p)\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(0, -r_{H_k}^2\boldsymbol{E}_{k|k-1}^{\mathrm{T}}\boldsymbol{E}_{k|k-1}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{I}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p)\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-\gamma_{k|k-1}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{I}_{n^2}, \boldsymbol{O}_p)\boldsymbol{\zeta} \leq 0, \\ \boldsymbol{\zeta}^{\mathrm{T}}\mathrm{diag}(-r_{H_k}^2\hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}}\hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{I}_p)\boldsymbol{\zeta} \leq 0. \end{cases} \tag{D.6}$$

Next, the S-procedure (Lemma A.4.2) is applied to the inequalities in (D.5) and (D.6). The inequality in (D.5) holds if there exist $\tau_1 \geq 0, \tau_2 \geq 0, \tau_3 \geq 0, \tau_4 \geq 0, \tau_5 \geq 0, \tau_6 \geq 0$ such that the following is true :

$$
\begin{aligned}
&\boldsymbol{\Pi}_{k|k-1}^{\mathrm{T}} \boldsymbol{P}_{k|k}^{-1} \boldsymbol{\Pi}_{k|k-1} - \operatorname{diag}(1, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_1 \operatorname{diag}(-1, \boldsymbol{I}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_2 \operatorname{diag}(-1, \boldsymbol{O}_n, \boldsymbol{R}_k^{-1}, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_3 \operatorname{diag}(0, -\gamma_{k|k-1}^2 \boldsymbol{E}_{k|k-1}^{\mathrm{T}} \boldsymbol{E}_{k|k-1}, \boldsymbol{O}_p, \boldsymbol{I}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_4 \operatorname{diag}(0, -r_{H_k}^2 \boldsymbol{E}_{k|k-1}^{\mathrm{T}} \boldsymbol{E}_{k|k-1}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{I}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_5 \operatorname{diag}(-\gamma_{k|k-1}^2 \hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}} \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{I}_{n^2}, \boldsymbol{O}_p) \\
&- \tau_6 \operatorname{diag}(-r_{H_k}^2 \hat{\boldsymbol{x}}_{k|k-1}^{\mathrm{T}} \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{O}_n, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{O}_p, \boldsymbol{O}_{n^2}, \boldsymbol{I}_p) \\
&\leq 0.
\end{aligned}
$$

The above inequality can be expressed in a compact form as

$$
\boldsymbol{\Pi}_{k|k-1}^{\mathrm{T}} \boldsymbol{P}_{k|k}^{-1} \boldsymbol{\Pi}_{k|k-1} - \boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6) \leq 0 \tag{D.7}
$$

where $\boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)$ is as given in (5.22). Utilizing the Schur complement (Lemma A.4.3), the inequality in (D.7) can be equivalently expressed as

$$
\begin{bmatrix}
-\boldsymbol{P}_{k|k} & \boldsymbol{\Pi}_{k|k-1} \\
\boldsymbol{\Pi}_{k|k-1}^{\mathrm{T}} & -\boldsymbol{\Theta}(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6)
\end{bmatrix} \leq 0. \tag{D.8}
$$

Solving the inequality in (D.8) with $\boldsymbol{P}_{k|k} > 0$ and $\tau_i \geq 0,\ i = 1, 2, 3, 4, 5, 6$ yields *a correction ellipsoid* that contains the true state of the system. To obtain the minimal set containing the true state, the sum of the squared lengths of semi-axes of the correction ellipsoid is minimized by minimizing the trace of $\boldsymbol{P}_{k|k}$. This completes the proof. □

References

[1] F. Blanchini and S. Miani, *Set-theoretic methods in control.* Springer, 2008.

[2] M. Benosman, “Model-based vs data-driven adaptive control: an overview,” *International Journal of Adaptive Control and Signal Processing*, vol. 32, no. 5, pp. 753–776, 2018. [Online]. Available: https://doi.org/10.1002/acs.2862

[3] M. Haring and T. A. Johansen, “Asymptotic stability of perturbation-based extremum-seeking control for nonlinear plants,” *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2302–2317, 2017. [Online]. Available: https://doi.org/10.1109/TAC.2016.2603607

[4] ——, “On the accuracy of gradient estimation in extremum-seeking control using small perturbations,” *Automatica*, vol. 95, pp. 23–32, 2018. [Online]. Available: https://doi.org/10.1016/j.automatica.2018.05.001

[5] M. Krstic and H.-H. Wang, “Stability of extremum seeking feedback for general nonlinear dynamic systems,” *Automatica*, vol. 36, no. 4, pp. 595–601, 2000. [Online]. Available: https://doi.org/10.1016/S0005-1098(99)00183-1

[6] Y. Tan, D. Nesic, and I. Mareels, “On non-local stability properties of extremum seeking control,” *Automatica*, vol. 42, no. 6, pp. 889–903, 2006. [Online]. Available: https://doi.org/10.1016/j.automatica.2006.01.014

[7] S. Drakunov, U. Ozguner, P. Dix, and B. Ashrafi, “ABS control using optimum search via sliding modes,” *IEEE Transactions on Control Systems Technology*, vol. 3, no. 1, pp. 79–85, 1995. [Online]. Available: https://doi.org/10.1109/87.370698

[8] D. Nesic, A. Mohammadi, and C. Manzie, “A framework for extremum seeking control of systems with parameter uncertainties,” *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 435–448, 2012. [Online]. Available: https://doi.org/10.1109/TAC.2012.2215270

[9] L. Wang, S. Chen, and K. Ma, “On stability and application of extremum seeking control without steady-state oscillation,” *Automatica*, vol. 68, pp. 18–26, 2016. [Online]. Available: https://doi.org/10.1016/j.automatica.2016.01.009

[10] P. Binetti, K. B. Ariyur, M. Krstic, and F. Bernelli, “Formation flight optimization using extremum seeking feedback,” *Journal of Guidance, Control, and Dynamics*, vol. 26, no. 1, pp. 132–142, 2003. [Online]. Available: https://doi.org/10.2514/2.5024

[11] D. F. Chichka, J. L. Speyer, C. Fanti, and C. G. Park, “Peak-seeking control for drag reduction in formation flight,” *Journal of guidance, control, and dynamics*, vol. 29, no. 5, pp. 1221–1230, 2006. [Online]. Available: https://doi.org/10.2514/1.15424

[12] M. Brodecki and K. Subbarao, “Autonomous formation flight control system using in-flight sweet-spot estimation,” *Journal of Guidance, Control, and Dynamics*, vol. 38, no. 6, pp. 1083–1096, 2014. [Online]. Available: https://doi.org/10.2514/1.G000220

[13] D. DeHaan and M. Guay, “Extremum-seeking control of state-constrained nonlinear systems,” *Automatica*, vol. 41, no. 9, pp. 1567–1574, 2005. [Online]. Available: https://doi.org/10.1016/j.automatica.2005.03.030

[14] V. Adetola and M. Guay, “Parameter convergence in adaptive extremum-seeking control,” *Automatica*, vol. 43, no. 1, pp. 105–110, 2007. [Online]. Available: https://doi.org/10.1016/j.automatica.2006.07.021

[15] M. Benosman and G. M. Atınç, “Extremum seeking-based adaptive control for electromagnetic actuators,” *International Journal of Control*, vol. 88, no. 3, pp. 517–530, 2015. [Online]. Available: https://doi.org/10.1080/00207179.2014.964779

[16] J. I. Poveda and N. Quijano, “Shahshahani gradient-like extremum seeking,” *Automatica*, vol. 58, pp. 51–59, 2015. [Online]. Available: https://doi.org/10.1016/j.automatica.2015.05.002

[17] M. Guay, I. Vandermeulen, S. Dougherty, and P. J. McLellan, “Distributed extremum-seeking control over networks of dynamically coupled unstable dynamic agents,” *Automatica*, vol. 93, pp. 498–509, 2018. [Online]. Available: https://doi.org/10.1016/j.automatica.2018.03.081

[18] N. Ghods and M. Krstic, “Speed regulation in steering-based source seeking,” *Automatica*, vol. 46, no. 2, pp. 452–459, 2010. [Online]. Available: https://doi.org/10.1016/j.automatica.2009.11.023

[19] D. Bhattacharjee, K. Subbarao, and K. Bhaganagar, “Extremum seeking and adaptive sampling approaches for plume source estimation using unmanned aerial vehicles,” in *AIAA Scitech 2019 Forum*, 2019, AIAA 2019-1565. [Online]. Available: https://doi.org/10.2514/6.2019-1565

[20] H.-B. Dürr, M. S. Stanković, C. Ebenbauer, and K. H. Johansson, “Lie bracket approximation of extremum seeking systems,” *Automatica*, vol. 49, no. 6, pp. 1538–1552, 2013. [Online]. Available: https://doi.org/10.1016/j.automatica.2013.02.016

[21] H.-B. Dürr, M. Krstić, A. Scheinker, and C. Ebenbauer, “Extremum seeking for dynamic maps using lie brackets and singular perturbations,” *Automatica*, vol. 83, pp. 91–99, 2017. [Online]. Available: https://doi.org/10.1016/j.automatica.2017.05.002

[22] V. Grushkovskaya, A. Zuyev, and C. Ebenbauer, "On a class of generating vector fields for the extremum seeking problem: Lie bracket approximation and stability properties," *Automatica*, vol. 94, pp. 151–160, 2018. [Online]. Available: https://doi.org/10.1016/j.automatica.2018.04.024

[23] R. Suttner, "Extremum seeking control with an adaptive dither signal," *Automatica*, vol. 101, pp. 214–222, 2019. [Online]. Available: https://doi.org/10.1016/j.automatica.2018.11.055

[24] K. B. Ariyur and M. Krstic, *Real time optimization by extremum seeking control.* Wiley Online Library, 2003.

[25] Y. Tan, W. Moase, C. Manzie, D. Nesic, and I. Mareels, "Extremum seeking from 1922 to 2010," in *Proceedings of the 29th Chinese Control Conference.* IEEE, 2010, pp. 14–26.

[26] Y. Tan, D. Nesic, I. M. Mareels, and A. Astolfi, "On global extremum seeking in the presence of local extrema," *Automatica*, vol. 45, no. 1, pp. 245–251, 2009. [Online]. Available: https://doi.org/10.1016/j.automatica.2008.06.010

[27] W. H. Moase, C. Manzie, and M. J. Brear, "Newton-like extremum-seeking for the control of thermoacoustic instability," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2094–2105, 2010. [Online]. Available: https://doi.org/10.1109/TAC.2010.2042981

[28] G. Taylor, R. Nudds, and A. Thomas, "Flying and swimming animals cruise at a strouhal number tuned for high power efficiency," *Nature*, vol. 425, no. 6959, p. 707, 2003. [Online]. Available: https://doi.org/10.1038/nature02000

[29] G. Triantafyllou, M. Triantafyllou, and M. Grosenbaugh, "Optimal thrust development in oscillating foils with application to fish propulsion," *Journal of Fluids and Structures*, vol. 7, no. 2, pp. 205–224, 1993. [Online]. Available: https://doi.org/10.1006/jfls.1993.1012

[30] J. Anderson, K. Streitlien, D. Barrett, and M. Triantafyllou, "Oscillating foils of high propulsive efficiency," *Journal of Fluid Mechanics*, vol. 360, pp. 41–72, 1998. [Online]. Available: https://doi.org/10.1017/S0022112097008392

[31] M. Triantafyllou, G. Triantafyllou, and R. Gopalkrishnan, "Wake mechanics for thrust generation in oscillating foils," *Physics of Fluids A: Fluid Dynamics*, vol. 3, no. 12, pp. 2835–2837, 1991. [Online]. Available: https://doi.org/10.1063/1.858173

[32] D. Floryan, T. Van Buren, and A. J. Smits, "Efficient cruising for swimming and flying animals is dictated by fluid drag," *Proceedings of the National Academy of Sciences*, vol. 115, no. 32, pp. 8116–8118, 2018. [Online]. Available: https://doi.org/10.1073/pnas.1805941115

[33] K. D. von Ellenrieder, K. Parker, and J. Soria, "Fluid mechanics of flapping wings," *Experimental Thermal and Fluid Science*, vol. 32, no. 8, pp. 1578–1589, 2008. [Online]. Available: https://doi.org/10.1016/j.expthermflusci.2008.05.003

[34] A. Paranjape, S.-J. Chung, and H. Hilton, "Optimizing the forces and propulsive efficiency in bird-scale flapping flight," in *AIAA Atmospheric Flight Mechanics Conference*, 2013, AIAA 2013-4916. [Online]. Available: https://doi.org/10.2514/6.2013-4916

[35] S. Heathcote, Z. Wang, and I. Gursul, "Effect of spanwise flexibility on flapping wing propulsion," *Journal of Fluids and Structures*, vol. 24, no. 2, pp. 183–199, 2008. [Online]. Available: https://doi.org/10.1016/j.jfluidstructs.2007.08.003

[36] D. Bhattacharjee and K. Subbarao, "Closed-form expressions for cycle-averaged aerodynamic quantities at an airfoil section of an avian flapping wing," in *AIAA Scitech 2019 Forum*, 2019, AIAA 2019-0565. [Online]. Available: https://doi.org/10.2514/6.2019-0565

[37] T. Theodorsen, “General theory of aerodynamic instability and the mechanism of flutter,” *NACA Technical Report*, vol. 496, 1935.

[38] I. Garrick, “Propulsion of a flapping and oscillating airfoil,” *NACA Technical Report*, vol. 567, 1937.

[39] J. Rayner, “A vortex theory of animal flight. part 1. the vortex wake of a hovering animal,” *Journal of Fluid Mechanics*, vol. 91, no. 4, pp. 697–730, 1979. [Online]. Available: https://doi.org/10.1017/S0022112079000410

[40] ——, “A vortex theory of animal flight. part 2. the forward flight of birds,” *Journal of Fluid Mechanics*, vol. 91, no. 4, pp. 731–763, 1979. [Online]. Available: https://doi.org/10.1017/S0022112079000422

[41] J. M. Rayner, “A new approach to animal flight mechanics,” *Journal of Experimental Biology*, vol. 80, no. 1, pp. 17–54, 1979. [Online]. Available: https://doi.org/10.1242/jeb.80.1.17

[42] J. DeLaurier, “An aerodynamic model for flapping-wing flight,” *The Aeronautical Journal*, vol. 97, no. 964, pp. 125–130, 1993. [Online]. Available: https://doi.org/10.1017/S0001924000026002

[43] T. Nakata, H. Liu, and R. Bomphrey, “A CFD-informed quasi-steady model of flapping-wing aerodynamics,” *Journal of fluid mechanics*, vol. 783, pp. 323–343, 2015. [Online]. Available: https://doi.org/10.1017/jfm.2015.537

[44] D. Chin and D. Lentink, “Flapping wing aerodynamics: from insects to vertebrates,” *Journal of Experimental Biology*, vol. 219, no. 7, pp. 920–932, 2016. [Online]. Available: https://doi.org/10.1242/jeb.042317

[45] B. D. Anderson and J. B. Moore, *Optimal filtering.* Prentice Hall, Inc., 1979.

[46] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960. [Online]. Available: https://doi.org/10.1115/1.3662552

[47] R. E. Kalman and R. S. Bucy, “New Results in Linear Filtering and Prediction Theory,” *Journal of Basic Engineering*, vol. 83, no. 1, pp. 95–108, 1961. [Online]. Available: https://doi.org/10.1115/1.3658902

[48] H. Witsenhausen, “Sets of possible states of linear systems given perturbed observations,” *IEEE Transactions on Automatic Control*, vol. 13, no. 5, pp. 556–558, 1968. [Online]. Available: https://doi.org/10.1109/TAC.1968.1098995

[49] F. Schweppe, “Recursive state estimation: unknown but bounded errors and system inputs,” *IEEE Transactions on Automatic Control*, vol. 13, no. 1, pp. 22–28, 1968. [Online]. Available: https://doi.org/10.1109/TAC.1968.1098790

[50] D. Bertsekas and I. Rhodes, “Recursive state estimation for a set-membership description of uncertainty,” *IEEE Transactions on Automatic Control*, vol. 16, no. 2, pp. 117–128, 1971. [Online]. Available: https://doi.org/10.1109/TAC.1971.1099674

[51] G. Calafiore, “Reliable localization using set-valued nonlinear filters,” *IEEE Transactions on systems, man, and cybernetics-part A: systems and humans*, vol. 35, no. 2, pp. 189–197, 2005. [Online]. Available: https://doi.org/10.1109/TSMCA.2005.843383

[52] F. Yang and Y. Li, “Set-membership filtering for discrete-time systems with nonlinear equality constraints,” *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2480–2486, 2009. [Online]. Available: https://doi.org/10.1109/TAC.2009.2029403

[53] D. Maksarov and J. Norton, “State bounding with ellipsoidal set description of the uncertainty,” *International Journal of Control*, vol. 65, no. 5, pp. 847–866, 1996. [Online]. Available: https://doi.org/10.1080/00207179608921725

[54] L. El Ghaoui and G. Calafiore, “Robust filtering for discrete-time systems with bounded noise and parametric uncertainty,” *IEEE Transactions on*

Automatic Control, vol. 46, no. 7, pp. 1084–1089, 2001. [Online]. Available: https://doi.org/10.1109/9.935060

[55] B. T. Polyak, S. A. Nazin, C. Durieu, and E. Walter, "Ellipsoidal parameter or state estimation under model uncertainty," *Automatica*, vol. 40, no. 7, pp. 1171–1179, 2004. [Online]. Available: https://doi.org/10.1016/j.automatica.2004.02.014

[56] Y. Becis-Aubry, M. Boutayeb, and M. Darouach, "State estimation in the presence of bounded disturbances," *Automatica*, vol. 44, no. 7, pp. 1867–1873, 2008. [Online]. Available: https://doi.org/10.1016/j.automatica.2007.10.033

[57] J. S. Shamma and K.-Y. Tu, "Approximate set-valued observers for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 42, no. 5, pp. 648–658, 1997. [Online]. Available: https://doi.org/10.1109/9.580870

[58] J. Wan, S. Sharma, and R. Sutton, "Guaranteed state estimation for nonlinear discrete-time systems via indirectly implemented polytopic set computation," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4317–4322, 2018. [Online]. Available: https://doi.org/10.1109/TAC.2018.2816262

[59] Y. Wang, Z. Wang, V. Puig, and G. Cembrano, "Zonotopic set-membership state estimation for discrete-time descriptor LPV systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 2092–2099, 2018. [Online]. Available: https://doi.org/10.1109/TAC.2018.2863659

[60] 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*, vol. 69, pp. 126–136, 2016. [Online]. Available: https://doi.org/10.1016/j.automatica.2016.02.036

[61] 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*, vol. 111, p. 108614, 2020. [Online]. Available: https://doi.org/10.1016/j.automatica.2019.108614

[62] B. S. Rego, J. K. Scott, D. M. Raimondo, and G. V. Raffo, "Set-valued state estimation of nonlinear discrete-time systems with nonlinear invariants based on constrained zonotopes," *Automatica*, vol. 129, p. 109638, 2021. [Online]. Available: https://doi.org/10.1016/j.automatica.2021.109638

[63] G. Belforte, B. Bona, and V. Cerone, "Parameter estimation algorithms for a set-membership description of uncertainty," *Automatica*, vol. 26, no. 5, pp. 887–898, 1990. [Online]. Available: https://doi.org/10.1016/0005-1098(90)90005-3

[64] T. Raïssi, D. Efimov, and A. Zolghadri, "Interval state estimation for a class of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 260–265, 2011. [Online]. Available: https://doi.org/10.1109/TAC.2011.2164820

[65] F. Yang and Y. Li, "Set-membership filtering for systems with sensor saturation," *Automatica*, vol. 45, no. 8, pp. 1896–1902, 2009. [Online]. Available: https://doi.org/10.1016/j.automatica.2009.04.011

[66] G. Wei, S. Liu, Y. Song, and Y. Liu, "Probability-guaranteed set-membership filtering for systems with incomplete measurements," *Automatica*, vol. 60, pp. 12–16, 2015. [Online]. Available: https://doi.org/10.1016/j.automatica.2015.06.037

[67] Z. Wang, X. Shen, Y. Zhu, and J. Pan, "A tighter set-membership filter for some nonlinear dynamic systems," *IEEE Access*, vol. 6, pp. 25 351–25 362, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2830350

[68] Z. Wang, X. Shen, and Y. Zhu, "Ellipsoidal fusion estimation for multisensor dynamic systems with bounded noises," *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4725–4732, 2019. [Online]. Available: https://doi.org/10.1109/TAC.2019.2902722

[69] E. Scholte and M. E. Campbell, "A nonlinear set-membership filter for on-line applications," *International Journal of Robust and Nonlinear Control*, vol. 13, no. 15, pp. 1337–1358, 2003. [Online]. Available: https://doi.org/10.1002/rnc.856

[70] B. Zhou, J. Han, and G. Liu, "A UD factorization-based nonlinear adaptive set-membership filter for ellipsoidal estimation," *International Journal of Robust and Nonlinear Control*, vol. 18, no. 16, pp. 1513–1531, 2008. [Online]. Available: https://doi.org/10.1002/rnc.1289

[71] C. P. Mracek and J. R. Cloutier, "Control designs for the nonlinear benchmark problem via the state-dependent Riccati equation method," *International Journal of robust and nonlinear control*, vol. 8, no. 4-5, pp. 401–433, 1998. [Online]. Available: https://doi.org/10.1002/(SICI)1099-1239(19980415/30)8:4/5%3C401::AID-RNC361%3E3.0.CO;2-U

[72] T. Cimen, "Survey of state-dependent Riccati equation in nonlinear optimal feedback control synthesis," *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 4, pp. 1025–1047, 2012. [Online]. Available: https://doi.org/10.2514/1.55821

[73] C. Jaganath, A. Ridley, and D. S. Bernstein, "A SDRE-based asymptotic observer for nonlinear discrete-time systems," in *Proceedings of the American Control Conference.* IEEE, 2005, pp. 3630–3635. [Online]. Available: https://doi.org/10.1109/ACC.2005.1470537

[74] I. Chang and J. Bentsman, "High-fidelity discrete-time state-dependent Riccati equation filters for stochastic nonlinear systems with gaussian/non-gaussian noises," in *Proceedings of the American Control Conference.* IEEE, 2018, pp. 1132–1137. [Online]. Available: https://doi.org/10.23919/ACC.2018.8431771

[75] W. Ren and R. W. Beard, *Distributed consensus in multi-vehicle cooperative control*. Springer, 2008.

[76] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE transactions on automatic control*, vol. 49, no. 9, pp. 1465–1476, 2004. [Online]. Available: https://doi.org/10.1109/TAC.2004.834433

[77] R. M. Murray, "Recent research in cooperative control of multivehicle systems," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 129, no. 5, pp. 571–583, 2007. [Online]. Available: https://doi.org/10.1115/1.2766721

[78] Z. Li, Z. Duan, G. Chen, and L. Huang, "Consensus of multiagent systems and synchronization of complex networks: A unified viewpoint," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 1, pp. 213–224, 2009. [Online]. Available: https://doi.org/10.1109/TCSI.2009.2023937

[79] J. Wu, V. Ugrinovskii, and F. Allgöwer, "Cooperative estimation for synchronization of heterogeneous multi-agent systems using relative information," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 4662–4667, 2014. [Online]. Available: https://doi.org/10.3182/20140824-6-ZA-1003.01938

[80] R. Bhusal and K. Subbarao, "Sensitivity analysis of cooperating multi-agent systems with uncertain connection weights," in *2019 American Control Conference (ACC)*, 2019, pp. 4024–4029. [Online]. Available: https://doi.org/10.23919/ACC.2019.8815336

[81] H. L. Trentelman, K. Takaba, and N. Monshizadeh, "Robust synchronization of uncertain linear multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1511–1523, 2013. [Online]. Available: https://doi.org/10.1109/TAC.2013.2239011

[82] X. Wang, J. Zhu, and Z. Cheng, “Synchronization reachable topology and synchronization of discrete-time linear multi-agent systems,” *IEEE Transactions on Automatic Control*, vol. 60, no. 7, pp. 1927–1932, 2015. [Online]. Available: https://doi.org/10.1109/TAC.2014.2362990

[83] J. Back and J.-S. Kim, “Output feedback practical coordinated tracking of uncertain heterogeneous multi-agent systems under switching network topology,” *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6399–6406, 2017. [Online]. Available: https://doi.org/10.1109/TAC.2017.2651166

[84] Z. Li, G. Wen, Z. Duan, and W. Ren, “Designing fully distributed consensus protocols for linear multi-agent systems with directed graphs,” *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 1152–1157, 2014. [Online]. Available: https://doi.org/10.1109/TAC.2014.2350391

[85] F. L. Lewis, B. Cui, T. Ma, Y. Song, and C. Zhao, “Heterogeneous multi-agent systems: reduced-order synchronization and geometry,” *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1391–1396, 2016. [Online]. Available: https://doi.org/10.1109/TAC.2015.2471716

[86] E. Arabi, T. Yucelen, and W. M. Haddad, “Mitigating the effects of sensor uncertainties in networked multi-agent systems,” *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 139, no. 4, 2017. [Online]. Available: https://doi.org/10.1115/1.4035092

[87] D. Silvestre, P. Rosa, R. Cunha, J. P. Hespanha, and C. Silvestre, “Gossip average consensus in a byzantine environment using stochastic set-valued observers,” in *52nd IEEE conference on decision and control.* IEEE, 2013, pp. 4373–4378. [Online]. Available: https://doi.org/10.1109/CDC.2013.6760562

[88] D. Silvestre, P. Rosa, J. P. Hespanha, and C. Silvestre, “Finite-time average consensus in a byzantine environment using set-valued observers,” in *2014*

American Control Conference. IEEE, 2014, pp. 3023–3028. [Online]. Available: https://doi.org/10.1109/ACC.2014.6859426

[89] M. E. Valcher and P. Misra, "On the consensus and bipartite consensus in high-order multi-agent dynamical systems with antagonistic interactions," *Systems & Control Letters*, vol. 66, pp. 94–103, 2014. [Online]. Available: https://doi.org/10.1016/j.sysconle.2014.01.006

[90] K. Hengster-Movric, K. You, F. L. Lewis, and L. Xie, "Synchronization of discrete-time multi-agent systems on graphs using Riccati design," *Automatica*, vol. 49, no. 2, pp. 414–423, 2013. [Online]. Available: https://doi.org/10.1016/j.automatica.2012.11.038

[91] H. Zhang, F. L. Lewis, and A. Das, "Optimal design for synchronization of cooperative systems: state feedback, observer and output feedback," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1948–1952, 2011. [Online]. Available: https://doi.org/10.1109/TAC.2011.2139510

[92] F. Xiao and L. Wang, "Asynchronous rendezvous analysis via set-valued consensus theory," *SIAM Journal on Control and Optimization*, vol. 50, no. 1, pp. 196–221, 2012. [Online]. Available: https://doi.org/10.1137/100801202

[93] U. Munz, A. Papachristodoulou, and F. Allgower, "Delay robustness in non-identical multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1597–1603, 2011. [Online]. Available: https://doi.org/10.1109/TAC.2011.2178336

[94] A. Garulli and A. Giannitrapani, "Analysis of consensus protocols with bounded measurement errors," *Systems & Control Letters*, vol. 60, no. 1, pp. 44–52, 2011. [Online]. Available: https://doi.org/10.1016/j.sysconle.2010.10.005

[95] T. Sadikhov, W. M. Haddad, T. Yucelen, and R. Goebel, "Approximate consensus of multiagent systems with inaccurate sensor measurements," *ASME*

Journal of Dynamic Systems, Measurement, and Control, vol. 139, no. 9, 2017. [Online]. Available: https://doi.org/10.1115/1.4036031

[96] X. Ge, Q.-L. Han, and F. Yang, "Event-based set-membership leader-following consensus of networked multi-agent systems subject to limited communication resources and unknown-but-bounded noise," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5045–5054, 2016. [Online]. Available: https://doi.org/10.1109/TIE.2016.2613929

[97] J. B. Rawlings and D. Q. Mayne, *Model predictive control: Theory and design.* Nob Hill Pub., 2009.

[98] W. H. Kwon and S. H. Han, *Receding horizon control: model predictive control for state models.* Springer Science & Business Media, 2006.

[99] V. Bachtiar, C. Manzie, and E. C. Kerrigan, "Nonlinear model-predictive integrated missile control and its multiobjective tuning," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 40, no. 11, pp. 2961–2970, 2017. [Online]. Available: https://doi.org/10.2514/1.G002279

[100] V. Bachtiar, T. Mühlpfordt, W. Moase, T. Faulwasser, R. Findeisen, and C. Manzie, "Nonlinear model predictive missile control with a stabilising terminal constraint," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 457–462, 2014. [Online]. Available: https://doi.org/10.3182/20140824-6-ZA-1003.02122

[101] Z. Li, Y. Xia, C.-Y. Su, J. Deng, J. Fu, and W. He, "Missile guidance law based on robust model predictive control using neural-network optimization," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 8, pp. 1803–1809, 2014. [Online]. Available: https://doi.org/10.1109/TNNLS.2014.2345734

[102] S. He and D. Lin, "Guidance laws based on model predictive control and target manoeuvre estimator," *Transactions of the Institute of Measurement*

and Control, vol. 38, no. 12, pp. 1509–1519, 2016. [Online]. Available: https://doi.org/10.1177/0142331215597970

[103] J. Wang and S. He, "Optimal integral sliding mode guidance law based on generalized model predictive control," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 230, no. 7, pp. 610–621, 2016. [Online]. Available: https://doi.org/10.1177/0959651816640618

[104] S. Kang, J. Wang, G. Li, J. Shan, and I. R. Petersen, "Optimal cooperative guidance law for salvo attack: an mpc-based consensus perspective," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 5, pp. 2397–2410, 2018. [Online]. Available: https://doi.org/10.1109/TAES.2018.2816880

[105] S. Ghosh, D. Ghose, and S. Raha, "Capturability of augmented pure proportional navigation guidance against time-varying target maneuvers," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 37, no. 5, pp. 1446–1461, 2014. [Online]. Available: https://doi.org/10.2514/1.G000561

[106] M. Guelman, "Proportional navigation with a maneuvering target," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-8, no. 3, pp. 364–371, 1972. [Online]. Available: https://doi.org/10.1109/TAES.1972.309520

[107] ——, "Missile acceleration in proportional navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-9, no. 3, pp. 462–463, 1973. [Online]. Available: https://doi.org/10.1109/TAES.1973.309733

[108] S. Ghawghawe and D. Ghose, "Pure proportional navigation against time-varying target manoeuvres," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1336–1347, 1996. [Online]. Available: https://doi.org/10.1109/7.543854

[109] A. E. Bryson and Y.-C. Ho, *Applied optimal control: optimization, estimation and control.* Taylor & Francis, 1975.

[110] I.-S. Jeon and J.-I. Lee, "Optimality of proportional navigation based on nonlinear formulation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 4, pp. 2051–2055, 2010. [Online]. Available: https://doi.org/10.1109/TAES.2010.5595614

[111] W.-H. Chen, D. J. Ballance, and P. J. Gawthrop, "Optimal control of nonlinear systems: a predictive control approach," *Automatica*, vol. 39, no. 4, pp. 633–641, 2003. [Online]. Available: https://doi.org/10.1016/S0005-1098(02)00272-8

[112] D. W. Clarke and C. Mohtadi, "Properties of generalized predictive control," *Automatica*, vol. 25, no. 6, pp. 859–875, 1989. [Online]. Available: https://doi.org/10.1016/0005-1098(89)90053-8

[113] L.-G. Lin and M. Xin, "Missile guidance law based on new analysis and design of sdre scheme," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 42, no. 4, pp. 853–868, 2019. [Online]. Available: https://doi.org/10.2514/1.G003544

[114] I.-S. Jeon, M. Karpenko, and J.-I. Lee, "Connections between proportional navigation and terminal velocity maximization guidance," *Journal of Guidance, Control, and Dynamics*, vol. 43, no. 2, pp. 383–388, 2020. [Online]. Available: https://doi.org/10.2514/1.G004672

[115] R. Chai, A. Savvaris, and S. Chai, "Integrated missile guidance and control using optimization-based predictive control," *Nonlinear Dynamics*, vol. 96, no. 2, pp. 997–1015, 2019. [Online]. Available: https://doi.org/10.1007/s11071-019-04835-8

[116] X. Chen and J. Wang, “Optimal control based guidance law to control both impact time and impact angle,” *Aerospace Science and Technology*, vol. 84, pp. 454–463, 2019. [Online]. Available: https://doi.org/10.1016/j.ast.2018.10.036

[117] V. Shalumov, “Optimal cooperative guidance laws in a multiagent target–missile–defender engagement,” *AIAA Journal of Guidance, Control, and Dynamics*, pp. 1–14, 2019. [Online]. Available: https://doi.org/10.2514/1.G004054

[118] H. B. Oza and R. Padhi, “Impact-angle-constrained suboptimal model predictive static programming guidance of air-to-ground missiles,” *AIAA Journal of Guidance, Control, and Dynamics*, vol. 35, no. 1, pp. 153–164, 2012. [Online]. Available: https://doi.org/10.2514/1.53647

[119] A. Ratnoo and D. Ghose, “Impact angle constrained interception of stationary targets,” *AIAA Journal of Guidance, Control, and Dynamics*, vol. 31, no. 6, pp. 1817–1822, 2008. [Online]. Available: https://doi.org/10.2514/1.37864

[120] ——, “Impact angle constrained guidance against nonstationary nonmaneuvering targets,” *AIAA Journal of Guidance, Control, and Dynamics*, vol. 33, no. 1, pp. 269–275, 2010. [Online]. Available: https://doi.org/10.2514/1.45026

[121] V. Shaferman and T. Shima, “Linear quadratic guidance laws for imposing a terminal intercept angle,” *AIAA Journal of Guidance, Control, and Dynamics*, vol. 31, no. 5, pp. 1400–1412, 2008. [Online]. Available: https://doi.org/10.2514/1.32836

[122] ——, “Cooperative optimal guidance laws for imposing a relative intercept angle,” *AIAA Journal of Guidance, Control, and Dynamics*, vol. 38, no. 8, pp. 1395–1408, 2015. [Online]. Available: https://doi.org/10.2514/1.G000568

[123] C.-K. Ryoo, H. Cho, and M.-J. Tahk, “Optimal guidance laws with terminal impact angle constraint,” *AIAA Journal of Guidance, Control,*

and Dynamics, vol. 28, no. 4, pp. 724–732, 2005. [Online]. Available: https://doi.org/10.2514/1.8392

[124] A. Chakravarthy and D. Ghose, "Obstacle avoidance in a dynamic environment: A collision cone approach," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 28, no. 5, pp. 562–574, 1998. [Online]. Available: https://doi.org/10.1109/3468.709600

[125] ——, "Collision cones for quadric surfaces," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1159–1166, 2011. [Online]. Available: https://doi.org/10.1109/TRO.2011.2159413

[126] ——, "Generalization of the collision cone approach for motion safety in 3-d environments," *Autonomous Robots*, vol. 32, no. 3, pp. 243–266, 2012. [Online]. Available: https://doi.org/10.1007/s10514-011-9270-z

[127] W. Zuo, K. Dhal, A. Keow, A. Chakravarthy, and Z. Chen, "Model-based control of a robotic fish to enable 3d maneuvering through a moving orifice," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4719–4726, 2020. [Online]. Available: https://doi.org/10.1109/LRA.2020.3003862

[128] D. Bhattacharjee and K. Subbarao, "Extremum seeking control with attenuated steady-state oscillations," *Automatica*, vol. 125, p. 109432, 2021. [Online]. Available: https://doi.org/10.1016/j.automatica.2020.109432

[129] ——, "A flight mechanics-based justification of the unique range of strouhal numbers for avian cruising flight," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, p. 0954410020976597, 2020. [Online]. Available: https://doi.org/10.1177%2F0954410020976597

[130] ——, "Set-membership filter for discrete-time nonlinear systems using state dependent coefficient parameterization," *IEEE Transactions on Automatic Control*, 2021. [Online]. Available: https://doi.org/10.1109/TAC.2021.3082504

[131] ——, “Set-membership filtering-based leader–follower synchronization of discrete-time linear multi-agent systems,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 143, no. 6, p. 064502, 2021. [Online]. Available: https://doi.org/10.1115/1.4049553

[132] D. Bhattacharjee, A. Chakravarthy, and K. Subbarao, “Nonlinear model predictive control and collision-cone-based missile guidance algorithm,” *Journal of Guidance, Control, and Dynamics*, pp. 1–17, 2021. [Online]. Available: https://doi.org/10.2514/1.G005879

[133] D. Bhattacharjee, K. Subbarao, and A. Chakravarthy, “Set-membership filtering-based pure proportional navigation,” in *AIAA Scitech 2021 Forum*, 2021, AIAA 2021-1567. [Online]. Available: https://doi.org/10.2514/6.2021-1567

[134] D. Bhattacharjee, K. Subbarao, and K. Bhaganagar, “Reachable set estimation for discrete-time nonlinear systems using ellipsoidal set-membership frameworks,” in *AIAA Scitech 2021 Forum*, 2021, AIAA 2021-1459. [Online]. Available: https://doi.org/10.2514/6.2021-1459

[135] D. Bhattacharjee, A. Chakravarthy, and K. Subbarao, “Nonlinear model predictive control based missile guidance for target interception,” in *AIAA Scitech 2020 Forum*, 2020, AIAA 2020-0865. [Online]. Available: https://doi.org/10.2514/6.2020-0865

[136] S. J. Moura and Y. A. Chang, “Lyapunov-based switched extremum seeking for photovoltaic power maximization,” *Control Engineering Practice*, vol. 21, no. 7, pp. 971–980, 2013. [Online]. Available: https://doi.org/10.1016/j.conengprac.2013.02.009

[137] K. T. Atta and M. Guay, “Comment on “On stability and application of extremum seeking control without steady-state oscillation” [Automatica 68

(2016) 18–26],” *Automatica*, vol. 103, pp. 580–581, 2019. [Online]. Available: https://doi.org/10.1016/j.automatica.2018.01.016

[138] M. Krstić, “Performance improvement and limitations in extremum seeking control,” *Systems & Control Letters*, vol. 39, no. 5, pp. 313–326, 2000. [Online]. Available: https://doi.org/10.1016/S0167-6911(99)00111-5

[139] C. Orlowski and A. Girard, “Modeling and simulation of nonlinear dynamics of flapping wing micro air vehicles,” *AIAA journal*, vol. 49, no. 5, pp. 969–981, 2011. [Online]. Available: https://doi.org/10.2514/1.J050649

[140] M. Goman and A. Khrabrov, “State-space representation of aerodynamic characteristics of an aircraft at high angles of attack,” *Journal of Aircraft*, vol. 31, no. 5, pp. 1109–1115, 1994. [Online]. Available: https://doi.org/10.2514/3.46618

[141] I. H. Tuncer and M. F. Platzer, “Computational study of flapping airfoil aerodynamics,” *Journal of Aircraft*, vol. 37, no. 3, pp. 514–520, 2000. [Online]. Available: https://doi.org/10.2514/2.2628

[142] P. Withers, “An aerodynamic analysis of bird wings as fixed aerofoils,” *Journal of Experimental Biology*, vol. 90, no. 1, pp. 143–162, 1981. [Online]. Available: https://doi.org/10.1242/jeb.90.1.143

[143] A. Paranjape, S.-J. Chung, H. Hilton, and A. Chakravarthy, “Dynamics and performance of tailless micro aerial vehicle with flexible articulated wings,” *AIAA journal*, vol. 50, no. 5, pp. 1177–1188, 2012. [Online]. Available: https://doi.org/10.2514/1.J051447

[144] K. Hall, S. Pigott, and S. Hall, “Power requirements for large-amplitude flapping flight,” *Journal of Aircraft*, vol. 35, no. 3, pp. 352–361, 1998. [Online]. Available: https://doi.org/10.2514/2.2324

[145] C. Ellington, "Insects versus birds: the great divide (invited)," in *44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada*, 2006, AIAA 2006-35. [Online]. Available: https://doi.org/10.2514/6.2006-35

[146] S. Hoerner, "Fluid-dynamic drag," *Brick Town, NJ*, 1965.

[147] D. Viieru, R. Albertani, W. Shyy, and P. Ifju, "Effect of tip vortex on wing aerodynamics of micro air vehicles," *Journal of Aircraft*, vol. 42, no. 6, pp. 1530–1536, 2005. [Online]. Available: https://doi.org/10.2514/1.12805

[148] A. Paranjape, M. Dorothy, S.-J. Chung, and K.-D. Lee, "A flight mechanics-centric review of bird-scale flapping flight," *International Journal of Aeronautical and Space Sciences*, vol. 13, no. 3, pp. 267–281, 2012. [Online]. Available: https://doi.org/10.5139/IJASS.2012.13.3.267

[149] I. Tuncer and M. Kaya, "Optimization of flapping airfoils for maximum thrust and propulsive efficiency," *AIAA journal*, vol. 43, no. 11, pp. 2329–2336, 2005. [Online]. Available: https://doi.org/10.2514/1.816

[150] K. Isogai, Y. Shinmoto, and Y. Watanabe, "Effects of dynamic stall on propulsive efficiency and thrust of flapping airfoil," *AIAA journal*, vol. 37, no. 10, pp. 1145–1151, 1999. [Online]. Available: https://doi.org/10.2514/2.589

[151] D. Doman, M. Oppenheimer, and D. Sigthorsson, "Dynamics and control of a minimally actuated biomimetic vehicle: Part I-aerodynamic model," in *AIAA Guidance, Navigation, and Control Conference*, 2009, AIAA 2009-6160. [Online]. Available: https://doi.org/10.2514/6.2009-6160

[152] C. Orlowski and A. Girard, "Longitudinal flight dynamics of flapping-wing micro air vehicles," *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 4, pp. 1115–1131, 2012. [Online]. Available: https://doi.org/10.2514/1.55923

[153] D. Bhattacharjee, A. Paranjape, and R. Pant, "Optimization of the spanwise twist of a flapping wing for bird-sized aircraft using a quasi-

steady aerodynamic model," *International Journal of Aeronautical and Space Sciences*, vol. 20, pp. 571–583, 2019. [Online]. Available: https://doi.org/10.1007/s42405-019-00154-9

[154] D. J. Cleaver, Z. Wang, I. Gursul, and M. Visbal, "Lift enhancement by means of small-amplitude airfoil oscillations at low reynolds numbers," *AIAA journal*, vol. 49, no. 9, pp. 2018–2033, 2011. [Online]. Available: https://doi.org/10.2514/1.J051014

[155] D. Betteridge and R. Archer, "A study of the mechanics of flapping wings," *The Aeronautical Quarterly*, vol. 25, no. 2, pp. 129–142, 1974. [Online]. Available: https://doi.org/10.1017/S0001925900006892

[156] P. Phlips, R. East, and N. Pratt, "An unsteady lifting line theory of flapping wings with application to the forward flight of birds," *Journal of fluid mechanics*, vol. 112, pp. 97–125, 1981. [Online]. Available: https://doi.org/10.1017/S0022112081000311

[157] H. Oehme and U. Kitzler, "On the geometry of the avian wing (studies on the biophysics and physiology of avian flight II)," NASA-TT-F-16901, Tech. Rep., 1975.

[158] B. Tobalske and K. Dial, "Flight kinematics of black-billed magpies and pigeons over a wide range of speeds," *Journal of Experimental Biology*, vol. 199, no. 2, pp. 263–280, 1996. [Online]. Available: https://doi.org/10.1242/jeb.199.2.263

[159] B. Parslew and W. Crowther, "Simulating avian wingbeat kinematics," *Journal of Biomechanics*, vol. 43, no. 16, pp. 3191–3198, 2010. [Online]. Available: https://doi.org/10.1016/j.jbiomech.2010.07.024

[160] T. Liu, K. Kuykendoll, R. Rhew, and S. Jones, "Avian wing geometry and kinematics," *AIAA journal*, vol. 44, no. 5, pp. 954–963, 2006. [Online]. Available: https://doi.org/10.2514/1.16224

[161] T. Bachmann, “Anatomical, morphometrical and biomechanical studies of barn owls and pigeons wings,” *RWTH Aachen University, Germany (PhD thesis)*, 2010.

[162] C. Pennycuick, “Speeds and wingbeat frequencies of migrating birds compared with calculated benchmarks,” *Journal of Experimental Biology*, vol. 204, no. 19, pp. 3283–3294, 2001. [Online]. Available: https://doi.org/10.1242/jeb.204.19.3283

[163] M. Vidyasagar, “Nonlinear systems analysis.” SIAM, 2002, ch. 2, pp. 51–52.

[164] Y. Song and J. W. Grizzle, “The Extended Kalman filter as a local asymptotic observer for nonlinear discrete-time systems,” *Journal of Mathematical Systems Estimation and Control*, vol. 5, no. 1, pp. 59–78, 1995.

[165] W. J. Vetter, “Matrix calculus operations and taylor expansions,” *SIAM Review*, vol. 15, no. 2, pp. 352–369, 1973. [Online]. Available: https://doi.org/10.1137/1015034

[166] D. P. Bertsekas, “Convexification procedures and decomposition methods for nonconvex optimization problems,” *Journal of Optimization Theory and Applications*, vol. 29, no. 2, pp. 169–197, 1979. [Online]. Available: https://doi.org/10.1007/BF00937167

[167] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized algorithms for analysis and control of uncertain systems: with applications.* Springer Science & Business Media, 2005.

[168] S. H. Brooks, “A discussion of random methods for seeking maxima,” *Operations research*, vol. 6, no. 2, pp. 244–251, 1958.

[169] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996. [Online]. Available: https://doi.org/10.1137/1038003

[170] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, vol. 3. Taipei, Taiwan, 2004. [Online]. Available: https://doi.org/10.1109/CACSD.2004.1393890

[171] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[172] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001. [Online]. Available: https://doi.org/10.1016/S0005-1098(01)00028-0

[173] E. D. Sontag, "A remark on the converging-input converging-state property," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 313–314, 2003. [Online]. Available: https://doi.org/10.1109/TAC.2002.808490

[174] D. V. Balandin, R. S. Biryukov, and M. M. Kogan, "Ellipsoidal reachable sets of linear time-varying continuous and discrete systems in control and estimation problems," *Automatica*, vol. 116, p. 108926, 2020. [Online]. Available: https://doi.org/10.1016/j.automatica.2020.108926

[175] A. Bemporad, "A quadratic programming algorithm based on nonnegative least squares with applications to embedded model predictive control," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 1111–1116, 2015. [Online]. Available: https://doi.org/10.1109/TAC.2015.2459211

[176] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," *Acta numerica*, vol. 4, pp. 1–51, 1995. [Online]. Available: https://doi.org/10.1017/S0962492900002518

[177] Y. B. Shtessel, I. A. Shkolnikov, and A. Levant, "Guidance and control of missile interceptor using second-order sliding modes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 110–124, 2009. [Online]. Available: https://doi.org/10.1109/TAES.2009.4805267

[178] M. N. Zeilinger, M. Morari, and C. N. Jones, "Soft constrained model predictive control with robust stability guarantees," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1190–1202, 2014. [Online]. Available: https://doi.org/10.1109/TAC.2014.2304371

[179] E. Kerrigan and J. Maciejowski, "Soft constraints and exact penalty functions in model predictive control," in *Proc. UKACC Int. Conf. (Control)*, Cambridge, U.K., Sep. 2000.

[180] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[181] G. Pannocchia, M. Gabiccini, and A. Artoni, "Offset-free mpc explained: novelties, subtleties, and applications," *IFAC-PapersOnLine*, vol. 48, no. 23, pp. 342–351, 2015. [Online]. Available: https://doi.org/10.1016/j.ifacol.2015.11.304

[182] M. Idan, T. Shima, and O. M. Golan, "Integrated sliding mode autopilot-guidance for dual-control missiles," *AIAA Journal of Guidance, Control, and Dynamics*, vol. 30, no. 4, pp. 1081–1089, 2007. [Online]. Available: https://doi.org/10.2514/1.24953

[183] A. Koren, M. Idan, and O. M. Golan, "Integrated sliding mode guidance and control for missile with on-off actuators," *AIAA Journal of guidance, control, and dynamics*, vol. 31, no. 1, pp. 204–214, 2008. [Online]. Available: https://doi.org/10.2514/1.31328

[184] P. Kokotović, H. K. Khalil, and J. O'reilly, *Singular perturbation methods in control: analysis and design.* SIAM, 1999.

[185] H. K. Khalil, *Nonlinear Systems (Second Edition).* Upper Saddle River, NJ, USA: Prentice hall, 1996.

[186] P. D. Christofides and A. R. Teel, “Singular perturbations and input-to-state stability,” *IEEE Transactions on Automatic Control*, vol. 41, no. 11, pp. 1645–1650, 1996. [Online]. Available: https://doi.org/10.1109/9.544001

[187] A. R. Teel, L. Moreau, and D. Nesic, “A unified framework for input-to-state stability in systems with two time scales,” *IEEE Transactions on Automatic Control*, vol. 48, no. 9, pp. 1526–1544, 2003. [Online]. Available: https://doi.org/10.1109/TAC.2003.816966

[188] J. Nocedal and S. Wright, *Numerical Optimization.* Springer Science & Business Media, 2006.

[189] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory.* Siam, 1994, vol. 15.

[190] M. Lazar, W. P. M. H. Heemels, and A. R. Teel, “Further Input-to-State Stability subtleties for discrete-time systems,” *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1609–1613, June 2013. [Online]. Available: https://doi.org/10.1109/TAC.2012.2231611

[191] C. A. Desoer and S. Shahruz, “Stability of nonlinear systems with three time scales,” *Circuits, Systems and Signal Processing*, vol. 5, no. 4, pp. 449–464, 1986. [Online]. Available: https://doi.org/10.1007/BF01599620

Biographical Statement

Diganta Bhattacharjee was born in Kalna (a small subdivision town in the Purba Bardhaman district of West Bengal, India) in 1991. He received a Bachelor of Engineering from the Indian Institute of Engineering Science and Technology Shibpur in 2014 and a Master of Technology from the Indian Institute of Technology Bombay in 2016, both in Aerospace Engineering. He joined the University of Texas at Arlington (UTA) in Fall 2017 as a Ph.D. student majoring in Aerospace Engineering. During the doctoral studies, he has worked in the Aerospace Systems Laboratory at UTA. His research interests include (1) nonlinear system analysis, (2) nonlinear, optimal and adaptive control designs, (3) linear and nonlinear estimation and filtering, (4) distributed control of networked multi-agent systems, and (5) aerodynamics and flight mechanics of avian-scale (or bird-scale) flapping.

Upon completion of his Ph.D., Diganta will be joining the Department of Aerospace Engineering and Mechanics at the University of Minnesota, Twin Cities as a Postdoctoral Associate.