ANALYSES AND SCALABLE ALGORITHMS FOR BYZANTINE-RESILIENT DISTRIBUTED OPTIMIZATION

by

Kananart Kuwaranancharoen

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Electrical and Computer Engineering

West Lafayette, Indiana

August 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Shreyas Sundaram, Chair

School of Electrical and Computer Engineering

Dr. Jianghai Hu

School of Electrical and Computer Engineering

Dr. Christopher G. Brinton

School of Electrical and Computer Engineering

Dr. Gesualdo Scutari

School of Industrial Engineering

Approved by:

Dr. Milind Kulkarni

To all the intelligent species across the universe

ACKNOWLEDGMENTS

I extend my heartfelt gratitude to my supervisor, Dr. Shreyas Sundaram, for his unwavering guidance, support, and patience throughout my PhD journey. His insightful feedback and valuable suggestions have been instrumental in shaping my research and personal growth. I am incredibly fortunate to have had him as my supervisor, and I am also grateful for his generous offer of a graduate research assistantship, which provided crucial financial support throughout my PhD and helped alleviate the burden of funding my research.

My thesis committee members, Dr. Jianghai Hu, Dr. Christopher G. Brinton, and Dr. Gesualdo Scutari, deserve my sincere appreciation for generously dedicating their time to attend my PhD preliminary and final examinations. Their invaluable feedback and constructive criticism have been instrumental in shaping my research into what it is today.

I would like to express my sincere gratitude to Dr. Michael Kishinevsky for extending to me the opportunity to join Intel Labs as a research intern. This experience has been invaluable in allowing me to broaden my knowledge and gain new skills. Additionally, I would like to thank Dr. Raid Ayoub and Dr. Pietro Mercati for their outstanding collaboration on a project both during my internship and after. Their constant support and guidance have been instrumental in my growth as a researcher.

Special thanks to Dr. Watcharapong Khovidhungij, Dr. David Banjerdpongchai, and Dr. Manop Wongsaisuwan from Chulalongkorn University, who provided invaluable support in pursuing my PhD by writing recommendation letters. I am also grateful to Dr. Andrew Ng for igniting my passion for machine learning (ML) and inspiring me to pursue a career in this field. His boundless enthusiasm for ML and thirst for knowledge motivated me to expand my horizons and acquire a deeper understanding of the subject. Since then, I have been fortunate to gain a wealth of knowledge and skills in this area.

I am grateful to my colleagues, Aritra Mitra, Lintao Ye, Amritha Prasad, Lei Xin, Tong Yao, Nathaniel Woodford, and Mustafa Abdallah, for their support and camaraderie throughout my research journey. Our frequent research discussions have been instrumental in shaping my ideas and refining my approach. Lei Xin deserves special recognition for his contributions to Chapter 4 of this dissertation.

I would like to express my sincere gratitude to Cho-Hsin Tsai for the insightful academic discussions, knowledge sharing, and enjoyable extracurricular activities we engaged in during our time here. Cho-Hsin's unwavering support and assistance have been invaluable throughout my doctoral journey. Additionally, I would like to thank Marco Hadisurya for our productive collaboration, which I believe will feature in your dissertation. Your contributions to my academic and personal growth are much appreciated.

I am deeply grateful to the several students from the Purdue University Thai Student Association (PUTSA) for their tireless efforts in organizing various activities throughout the years and for providing me with the opportunity to be part of the committee. Jirayu Sirivorawong deserves special recognition for our philosophical discussions, which had a profound impact on my worldview and way of thinking. Thank you all for the unforgettable experiences that I will always cherish.

I would like to express my sincere appreciation to Punyaporn Tippayawat, Daokham Naitun, Thanaporn Techavorabot, and Patsavee Yinadsawaphan for their unwavering care, vital mental support, and counseling throughout my years of study. Their constant encouragement, advice, and guidance were essential in helping me navigate the challenges of the doctoral journey. I am incredibly grateful to have had such amazing friends who have shown me unyielding kindness and support.

I am deeply grateful to my family for their love, encouragement, and support throughout my academic journey. Their patience, sacrifices, and unwavering belief in my abilities have been the pillars of my success.

Finally, I am deeply grateful for the financial support provided by the National Science Foundation, which made all the research projects in this dissertation possible. I would also like to express my appreciation to the Purdue University community for fostering a stimulating academic environment that allowed me to interact with fellow researchers, scholars, and experts from various fields, and enriched my academic experience. Additionally, I would like to acknowledge Chulalongkorn University for providing me with a strong foundation during my undergraduate studies, which helped shape my academic journey.

I cannot express my appreciation enough for the contributions of all these individuals, and for that, I am forever grateful.

PREFACE

"*Mathematics is the language with which God wrote the universe.*"

– Galileo di Vincenzo Bonaiuti de' Galilei

This dissertation is focused on the problem of Byzantine-resilient distributed optimization, which arises in the context of multi-agent systems where agents must cooperate to optimize a global objective function while facing the challenge of malicious or faulty agents. The importance of this problem stems from its applicability in a wide range of domains, including sensor networks, control systems, and machine learning, among others. In this preface, I will provide some historical background on optimization, distributed systems, and security, and then highlight the contribution of this work in the context of existing literature.

"*True optimization is the revolutionary contribution of modern research to decision processes.*" – George Dantzig

The history of mathematical optimization can be traced back to ancient times, with the Greeks using geometric methods to solve optimization problems such as finding the maximum area of a rectangle given a fixed perimeter. In modern times, the development of optimization algorithms has been driven by applications in engineering, economics, and computer science, among other fields. With the advent of computers, it became possible to solve large-scale optimization problems using numerical methods, leading to the emergence of optimization as a core area of research in mathematics and computer science.

"*Computer science has as much to do with computers as astronomy has to do with telescopes.*" – Edsger Wybe Dijkstra

Distributed systems have a long history, dating back to the 1960s with the development of computer networks. Over the years, distributed systems have become increasingly important, with the rise of the internet and the proliferation of mobile devices. Distributed optimization, which involves solving an optimization problem across multiple agents in a network, has been an active area of research since the early 2000s. The key challenge in distributed optimization is to design algorithms that can effectively harness the computational power of multiple agents while ensuring convergence to the optimal solution.

"*The only truly secure system is one that is powered off, cast in a block of concrete and sealed in a lead-lined room with armed guards.*" – Gene Spafford

Security is a critical concern in any distributed system, and peer-to-peer networks in which we consider are no exception. In the context of distributed optimization, security is particularly important because malicious or faulty agents can undermine the integrity of the optimization process and compromise the quality of the solution. Byzantine-resilient distributed optimization is a variant of the problem that is designed to address this challenge by ensuring that the optimization process remains robust in the face of Byzantine faults, where agents can behave arbitrarily.

"*If I have seen further, it is by standing on the shoulders of giants.*"
– Sir Isaac Newton

Despite the significance of Byzantine-resilient distributed optimization, there has been relatively limited effort in developing scalable algorithms with provable guarantees. Most of the existing work in the literature focuses on the case where the number of decision variables is small or even only a single variable. In this dissertation, we address the challenge of large-scale resilient distributed optimization, where the number of decision variables and the problem size are both substantial. We provide a comprehensive analysis of the problem and propose a set of scalable algorithms that can converge to a neighborhood of the optimal solution, even in the presence of Byzantine faults.

"*All truths are easy to understand once they are discovered; the point is to discover them.*"
– Galileo di Vincenzo Bonaiuti de' Galilei

This dissertation is organized into four parts. The first part (Chapter 1) introduces each of the works separately. The second part (Chapters 2 and 3) establishes the foundation for resilient distributed convex optimization by analyzing the properties of the sum of convex functions. The third part (Chapters 4 and 5) presents a set of scalable algorithms with theoretical guarantees for different cases. Finally, Chapter 6 concludes our work and provides directions for future research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The advent of advanced communication technologies has given rise to large-scale networks comprised of numerous interconnected agents, which need to cooperate to accomplish various tasks, such as distributed message routing, formation control, robust statistical inference, and spectrum access coordination. These tasks can be formulated as distributed optimization problems, which require agents to agree on a parameter minimizing the average of their local cost functions by communicating only with their neighbors. However, distributed optimization algorithms are typically susceptible to malicious (or "Byzantine") agents that do not follow the algorithm. This thesis offers analysis and algorithms for such scenarios. As the malicious agent's function can be modeled as an unknown function with some fundamental properties, we begin in the first two parts by analyzing the region containing the potential minimizers of a sum of functions. Specifically, we explicitly characterize the boundary of this region for the sum of two unknown functions with certain properties. In the third part, we develop resilient algorithms that allow correctly functioning agents to converge to a region containing the true minimizer under the assumption of convex functions of each regular agent. Finally, we present a general algorithmic framework that includes most state-of-the-art resilient algorithms. Under the strongly convex assumption, we derive a geometric rate of convergence of all regular agents to a ball around the optimal solution (whose size we characterize) for some algorithms within the framework.

1. INTRODUCTION

1.1 Motivation

Optimization has been an important tool in various fields, including machine learning [1], signal processing [2], control theory, [3]–[5], and robotics [6]–[8]. Given an objective function to be optimized, there are several standard algorithms that can be applied to find the optimal variables [9]–[12]. However, recent advances in communication technologies have given rise to the study of distributed optimization problems which can be traced back to the works related to parallel and distributed computation [13], [14]. Their applications include machine learning [9], [15]–[17], control of large-scale systems [18]–[20], and cooperative robotic systems [21], [22]. In these settings, each node in a network has data which is commonly assumed to be private and represented as a convex function. The objective of the network is to collaboratively determine the minimizer of the sum of these functions. In order to tackle the task, several algorithms have been proposed, including [23]–[28], under some common assumptions on the functions such as strongly convexity or bounded gradients. However, these existing works typically make the assumption that all agents are trustworthy and cooperative (i.e., they follow the prescribed protocol) [29]; indeed, such protocols fail if even a single agent behaves in a malicious or incorrect manner [30].

As security becomes a more important consideration in large scale systems, a handful of recent papers have considered fault tolerant algorithms for the case where agent misbehavior follows specific patterns [31], [32]. Nevertheless, a more general (and serious) form of misbehavior is captured by the *Byzantine* adversary model from computer science, where misbehaving agents can send arbitrary (and conflicting) values to their neighbors at each iteration of the algorithm. Under such Byzantine behavior, it has been shown that it is impossible to guarantee computation of the true optimal point [30], [33]. Thus, it is crucial to analyze properties of the solution and develop algorithms that are resilient to agents that do not follow the prescribed algorithm.

1.2 Related Works

In the situation when the network contains Byzantine nodes which arbitrarily deviate from the prescribed algorithm, it is shown that all regular nodes will fail to achieve the true minimizer regardless of the implemented algorithm [30], [34]. To overcome the issue, one possible direction is to consider robust optimization, where the objective function contains some parametric uncertainty, and the goal is to choose the optimization variable to be robust to the possible realizations of the uncertainty [35]–[37]. The problem that we consider in Chapter 2 and Chapter 3 also has a similar flavor, in that we assume that the optimization objective is not fully known. However, rather than seeking to find a single solution that is simultaneously robust to all possible realizations of the uncertain parameter (or learning that parameter [37]), we instead seek to characterize the region where the minimizer could lie for *each* possible realization of the uncertainty. This approach has the potential to yield valuable insights into the nature of the possible solutions to the given uncertain optimization problems. By identifying the region in which the true minimizer can lie (*potential solution region*), we can better evaluate the effectiveness of resilient distributed optimization algorithms. Moreover, this knowledge would offer central servers a way to combine machine learning models in the context of federated learning setups [38], [39].

Another potential direction is to construct algorithms that allow the non-faulty nodes to converge to a certain region (possibly containing the solution) [33], [40]. However, one major limitation of the works in this direction [30], [34], [41] is that they make the assumption of scalar-valued objective functions, and the extension of the above ideas to general multi-dimensional functions remains largely open. In fact, one major challenge for minimizing multi-dimensional functions is that the region containing the minimizer of the sum of functions is itself difficult to characterize. Specifically, in contrast to the case of scalar functions, where the global minimizer always lies within the smallest interval containing all local minimizers [30], the region containing the minimizer of the sum of multi-dimensional functions may not necessarily be in the convex hull of the minimizers as shown in the analytical characterization and numerical experiments from Chapter 2 and Chapter 3.

There exists a branch of literature focusing on secure distributed machine learning in a client-server architecture [42], [43], where the server appropriately filters the information received from the clients. The papers [44], [45] consider a vector version of the resilient machine learning problem in a distributed (peer-to-peer) setting. These papers show that the states of regular nodes will converge to the statistical minimizer with high probability (as the amount of data of each node goes to infinity), but the analysis is restricted to i.i.d training data across the network. However, when each agent has a finite amount of data, these algorithms are still vulnerable to sophisticated attacks as shown in [46]. The recent work [47] considers a Byzantine distributed optimization problem for multi-dimensional functions, but relies on redundancy among the local functions, and also requires the underlying communication network to be complete. These previous works assume either specific functions or a restricted network topology. Hence, we consider developing resilient algorithms applicable for functions with several variables with convergence guarantees under less restrictions on both network topology and function properties in Chapter 4.

In addition to the issue of scalability, much of the existing research demonstrates only asymptotic convergence results for proposed algorithms, or fails to provide any convergence analysis at all. In particular, while some algorithms, such as those proposed in [48], [49], have shown promising numerical results, they lack rigorous theoretical analysis. Others, including [30], [33], [45], [50]–[52] which rely on a simple distributed gradient descent algorithm equipped with extreme value-based filtering, provide only asymptotic convergence results. To address these gaps, we propose an algorithmic framework that includes several state-of-the-art resilient distributed optimization algorithms in the literature as well as our algorithms in Chapter 4, and provide finite-time convergence analysis (including convergence rate) for this set of algorithms, under a mild assumption on the local functions. Our analysis not only provides finite-time convergence analysis for existing algorithms, but also proposes a simple and easily-checkable sufficient condition for ensuring geometric convergence. This work is presented in detail in Chapter 5.

1.3 Main Contributions

1.3.1 The Minimizer of the Sum of Two Strongly Convex Functions

In Chapter 2, we first argue that identification of the region containing the minimizer of multi-dimensional functions under mild assumptions on two unknown functions is challenging as the potential solution region is neither contained in the smallest hyperrectangle contains their local minimizers nor the convex hull of their local minimizers. Subsequently, we thoroughly study the potential solution region of the sum of two unknown multi-dimensional functions. Our main contributions are summarized as follows.

(i) We characterize a region containing all valid minimizers for the sum of two arbitrary strongly convex functions in which we call an *outer approximation* (Section 2.4). In fact, the outer approximation comes from a geometrical relationship derived by using the first-order necessary conditions for minimizers.

(ii) We characterize a region where every point is a valid minimizer for the sum of two arbitrary strongly convex functions in which we call an *inner approximation* (Section 2.5). This inner approximation can be obtained from analyzing quadratic functions with positive curvature (Section 2.5.1). Similar to the outer approximation case, we provide the characterization for all distances between the minimizers of two strongly convex function.

(iii) Lastly, we identify the region containing the minimizer of the sum of two arbitrary strongly convex functions by noticing that the inner approximation essentially almost coincides with the outer approximation. More precisely, we provide equations that characterize the boundary of such region (Section 2.6).

1.3.2 On the Set of Possible Minimizers of a Sum of Known and Unknown Functions

Since in many applications, it may be the case that the objective function is only partially known, we instead seek to characterize the region where the minimizer could lie for *each* possible realization of the uncertainty. This approach has the potential to yield insights

regarding the nature of the possible solutions to the given uncertain optimization problem. In contrast to Chapter 2, we shall consider the case of optimizing a sum of known and unknown functions where only limited information about the unknown function is available in Chapter 3. In this case, we are given some general characteristics of the unknown function, namely a region containing the minimizer, and the strong convexity parameter of the function. Our goal is to determine necessary conditions for a point to be a minimizer of the sum. In particular, we will determine a region where the potential minimizer of the sum can lie. Thus, if a point from within this region is chosen as an estimate of the true minimizer of the sum, the size of the region can be used to quantify how far the estimate can be from the true minimizer. Our main contributions are summarized as follows.

(i) We provide a necessary condition for a point to be a potential minimizer of the sum in the case that the uncertainty region is a compact set (Section 3.4). Essentially, the analysis is based on the first-order necessary conditions for minimizers.

(ii) We provide a necessary condition for a point to be a potential minimizer of the sum in the case that the uncertainty region is a ball (Section 3.5). In addition, it is computationally cheaper to verify such condition than the previous case.

(iii) In Section 3.6, we present an algorithm (Algorithm 1) to provide an over-approximation of the potential solution set in the case that the uncertainty region is a ball.

1.3.3 Scalable Distributed Optimization of Multi-Dimensional Functions Despite Byzantine Adversaries

While in Chapter 2 and Chapter 3, we consider the problems of potential locations of the minimizer of the sum of functions, in Chapter 4, we consider Byzantine-resilient distributed optimization algorithms. Most of the previous algorithms either are applicable to single-dimensional functions [30], [34], [41] or require some restricted conditions on the functions [44], [45] or communication network [47]. In this work, we focus on scalable Byzantine-resilient distributed optimization algorithms (i.e., applicable to high-dimensional functions) with mild assumptions on the network topology. Our main contributions are summarized as follows.

(i) We propose two multi-dimensional resilient distributed algorithms (Algorithm 2 and Algorithm 3). Algorithm 2 employs two types of filters: a distance-based filter and min-max filter while Algorithm 3 utilizes only a distance-based filter.

(ii) For the first algorithm (Algorithm 2), we show that the states of regular agents can asymptotically reach consensus. Furthermore, we provide convergence guarantees for both algorithms irrespective of the actions of Byzantine agents (Section 4.4). In particular, both algorithms have the asymptotic convergence to the same region and we explicitly characterize to size of this region. Even though Algorithm 3 does not have consensus guarantee, it requires a more relaxed condition on the network topology, and thus, it is more scalable.

(iii) We demonstrate the performance of both algorithms by providing the results from a synthetic quadratic functions task and a machine learning task (banknote authentication task) (Section 4.6) and empirically show that the states of the regular agents can get reasonably closed to the optimal solution even in the presence of sophisticated Byzantine agents.

1.3.4 On the Geometric Convergence of Byzantine-Resilient Distributed Optimization Algorithms

In Chapter 5, we consider Byzantine-resilient (peer-to-peer) distributed deterministic optimization problems as in Chapter 4. However, in this work, we consider different assumptions on the local functions. Specifically, we examine smooth, strongly convex local functions in Chapter 5 in contrast to general convex local functions with bounded gradients in Chapter 4. Furthermore, we focus on analyzing convergence properties of existing algorithms (including the algorithms proposed in Chapter 4) instead of proposing a new resilient algorithm. Our contributions are as follows:

(i) We introduce an algorithmic framework called RedGraf, a generalization of BRIDGE in [45], which includes some state-of-the-art Byzantine-resilient distributed optimization algorithms as special cases.

(ii) We propose a novel contraction property which we show provides a general method for proving geometric convergence of algorithms in RedGraf. To the best of our knowledge, our work is the first to provide a geometric rate of convergence of all regular agents' states to a ball containing the true minimizer for a class of resilient algorithms under the strong convexity assumption. In addition, we explicitly characterize the convergence rate and the size of the convergence region.

(iii) We introduce a general mixing dynamics property which is used to derive approximate consensus results for algorithms in RedGraf in which both the convergence rate and the final consensus diameter are explicitly characterized.

(iv) Using our framework, we analyze the contraction and mixing dynamics properties of some state-of-the-art algorithms, leading to convergence and consensus results for each algorithm. Our work is the first to show the finite-time convergence result for some existing resilient algorithms.

(v) We demonstrate and compare the performance of the resilient algorithms through numerical simulations to corroborate the theoretical results for convergence and approximate consensus.

2. THE MINIMIZER OF THE SUM OF TWO STRONGLY CONVEX FUNCTIONS

2.1 Introduction

The problem of optimizing a sum of functions arises in a variety of applications, including machine learning [9], [15], [16], control of large-scale systems [18], [19], and cooperative robotic systems [21], [22]. In these settings, each node in a network is assumed to possess a convex function. There are many proposed algorithms to find the minimizer of the sum of these functions [23]–[27], under some common assumptions such as the functions being strongly convex and the gradients being bounded.

In some cases, the exact functions themselves may not be fully known, and only certain characteristics (such as minimizer and convexity parameters) may be known. In this case, it is of interest to understand the potential set of minimizers of the sum of functions, despite this limited knowledge of the individual functions. For example, in resilient distributed optimization settings [30], [40], the network contains malicious nodes that do not follow the distributed optimization algorithm and one cannot guarantee that all nodes calculate the true minimizer. Instead, one must settle for algorithms that allow the non-malicious nodes to converge to a certain region [33], [53]. In such situations, knowing the region where the minimizer can lie would allow us to evaluate the efficacy of such resilient distributed optimization algorithms. Another example involves the scenario where two similar machine learning models trained on similar data are combined to achieve a common goal. However, due to privacy concerns [54], the datasets used to train these models may not available directly, which means we know the minimizer of each function for inference but not the function itself. In this case, it is of interest to obtain a potential new minimizer, i.e., a

combined machine learning model, which could offer enhanced performance for the original inference task. More specifically, consider a federated learning setup [38], [39] with two clients, each with access to local data. Each client, denoted by i, obtains a local optimal parameter $\boldsymbol{x}_i^*$ that minimizes its own loss function f_i. Now, if each client only reports $\boldsymbol{x}_i^*$ to a server, can the server determine the set of potential optimal parameters for the combined loss function $f_1 + f_2$, based solely on the information it has (i.e., $\boldsymbol{x}_1^*$, $\boldsymbol{x}_2^*$, and knowledge of the convexity properties of f_1 and f_2)?

When the local functions f_i at each node v_i are univariate (i.e., $f_i : \mathbb{R} \to \mathbb{R}$), and strongly convex, it is easy to argue that the minimizer of the sum must lie in the interval bracketed by the smallest and largest minimizers of the functions [30]. This is because the gradients of all the functions will have the same sign outside that region, and thus cannot sum to zero. However, a similar characterization of the region containing the minimizer of multivariate functions is lacking in the literature, and is significantly more challenging to obtain. For example, the conjecture that the minimizer of a sum of convex functions is in the convex hull of their local minimizers can be easily disproved via simple examples; consider $f_1(x, y) = x^2 - xy + \frac{1}{2}y^2$ and $f_2(x, y) = x^2 + xy + \frac{1}{2}y^2 - 4x - 2y$ with minimizers $(0, 0)$ and $(2, 0)$ respectively, whose sum has minimizer $(1, 1)$. In our recent work [55], we studied this problem and provided an *outer approximation* on the region containing the minimizer of two strongly convex functions in a specific case; this region is determined by the minimizers of the individual functions, their strong convexity parameters, and the specified bound on the norms of the gradients of the functions at the location of the minimizer.

In this chapter, **our goal is to characterize an outer approximation (i.e., a region containing all valid minimizers) as well as an inner approximation (i.e., a region where every point is a valid minimizer) for the sum of two unknown strongly convex functions.** More specifically, we provide an outer approximation that is more general than the one given in [55]. As we will see, the inner approximation essentially almost coincides with the outer approximation. More precisely, the boundary of both outer and inner approximations are the same under the assumption that the gradients of the two original functions are bounded by a finite number at the potential minimizer of the sum. Thus, our analysis in this chapter complements and completes the analysis in [55] by

fully characterizing the region containing the minimizer of the sum of two strongly convex functions. While the analysis is complicated even for this scenario involving two functions, our analysis provides insights that could be leveraged in future work to tackle the sum of multiple functions.

This chapter is organized as follows. The notations used throughout this chapter and preliminaries are provided in Section 2.2. The problem formulation is in Section 2.3. Our main results regarding the outer approximation are in Section 2.4, our analysis of the inner approximation and potential solution region is in Section 2.5, and the conclusions follow in Section 2.7.

2.2 Preliminaries

2.2.1 Sets

Let $\mathbb{R}$ denotes the set of real numbers. We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space. For a subset $\mathcal{E}$ of a topological space, we denote the complement, closure and interior of a set $\mathcal{E}$ by $\mathcal{E}^c$, $\overline{\mathcal{E}}$ and $\mathcal{E}^\circ$, respectively. The boundary of $\mathcal{E}$ is defined as $\partial\mathcal{E} = \overline{\mathcal{E}} \setminus \mathcal{E}^\circ$. We also use $\mathbf{dom}(f)$ to denote the domain of function f. In addition, we use $\sqcup$ to denote the disjoint union operation. We will use this simple lemma later in this chapter.

Lemma 2.2.1. *Let $\mathcal{G}$ and $\mathcal{H}$ be subsets of a topological space X such that $\mathcal{G} \subseteq \mathcal{H}$. Let $\mathfrak{P}$ be a partition of $\mathcal{H}$. Then,*

$$\mathcal{G}^\circ = \bigsqcup_{\mathcal{Z}\in\mathfrak{P}} \Big((\mathcal{G} \cap \mathcal{Z}) \setminus (\partial\mathcal{G} \cap \mathcal{Z})\Big).$$

Proof. For $\mathcal{Z} \in \mathfrak{P}$, since $\mathcal{G} \cap \mathcal{Z} \cap \mathcal{Z}^c = \emptyset$, we have

$$\mathcal{G}^\circ \cap \mathcal{Z} = (\mathcal{G} \cap (\partial\mathcal{G})^c \cap \mathcal{Z}) \cup (\mathcal{G} \cap \mathcal{Z} \cap \mathcal{Z}^c) = (\mathcal{G} \cap \mathcal{Z}) \cap (\partial\mathcal{G} \cap \mathcal{Z})^c = (\mathcal{G} \cap \mathcal{Z}) \setminus (\partial\mathcal{G} \cap \mathcal{Z}).$$

Using the above equation, we can write

$$\mathcal{G}^\circ = \bigsqcup_{\mathcal{Z}\in\mathfrak{P}} (\mathcal{G}^\circ \cap \mathcal{Z}) = \bigsqcup_{\mathcal{Z}\in\mathfrak{P}} \Big((\mathcal{G} \cap \mathcal{Z}) \setminus (\partial\mathcal{G} \cap \mathcal{Z})\Big).$$

□

2.2.2 Linear Algebra

For simplicity, we often use $(x_1, \ldots, x_n)$ and $[x_1 \ x_2 \ \cdots \ x_n]^\intercal$ to represent the column vector $\boldsymbol{x}$. We use $\mathbf{0}$ to denote the all-zero vector with appropriate dimension and $\boldsymbol{e}_i$ to denote the i-th basis vector (the vector of all zeros except for a one in the i-th position). We denote by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ the Euclidean inner product of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\langle \boldsymbol{u}, \boldsymbol{v} \rangle := \boldsymbol{u}^\intercal \boldsymbol{v}$, by $\|\boldsymbol{u}\|$ the Euclidean norm of $\boldsymbol{u}$, i.e., $\|\boldsymbol{u}\| := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle} = (\sum_i u_i^2)^{1/2}$. We define the functions $\angle : (\mathbb{R}^n \setminus \{\mathbf{0}\}) \times (\mathbb{R}^n \setminus \{\mathbf{0}\}) \to [0, \boldsymbol{\pi}]$ and $\measuredangle : (\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times (\mathbb{R}^2 \setminus \{\mathbf{0}\}) \to \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$ as

$$\angle(\boldsymbol{u}, \boldsymbol{v}) := \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} \right) \quad \text{and} \quad \measuredangle(\boldsymbol{u}, \boldsymbol{v}) := \arcsin\left(\frac{u_2 v_1 - u_1 v_2}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} \right), \tag{2.1}$$

respectively. Note that $\angle(\boldsymbol{u}, \boldsymbol{v}) = \angle(\boldsymbol{v}, \boldsymbol{u})$ but $\measuredangle(\boldsymbol{u}, \boldsymbol{v}) = -\measuredangle(\boldsymbol{v}, \boldsymbol{u})$. We use

$$\mathcal{B}(\boldsymbol{x}_0, r_0) := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_0\| < r_0\} \tag{2.2}$$

and $\overline{\mathcal{B}}(\boldsymbol{x}_0, r_0)$ to denote the open and closed balls, respectively, in $\mathbb{R}^n$ centered at $\boldsymbol{x}_0 \in \mathbb{R}^n$ and with radius $r_0 \in \mathbb{R}_{>0}$. We use $\boldsymbol{I}$ to denote the identity matrix of appropriate dimension. For square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we use $\lambda(\boldsymbol{A})$, $\lambda_{\min}(\boldsymbol{A})$ and $\mathrm{Tr}(\boldsymbol{A})$ to denote an eigenvalue, the minimum eigenvalue and the trace of matrix $\boldsymbol{A}$, respectively. For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we use $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$ to denote the column space and null space of matrix $\boldsymbol{A}$, respectively.

2.2.3 Convex Sets and Functions

A set $\mathcal{C} \subseteq \mathbb{R}^n$ is said to be convex if, for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathcal{C}$ and all t in the interval $(0, 1)$, the point $(1-t)\boldsymbol{x} + t\boldsymbol{y}$ also belongs to $\mathcal{C}$. A differentiable function f is called strongly convex with parameter $\sigma \in \mathbb{R}_{>0}$ (or σ-strongly convex) if

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq \sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2 \tag{2.3}$$

holds for all points $\boldsymbol{x}, \boldsymbol{y}$ in its domain. We use $\mathcal{S}(\boldsymbol{x}^*, \sigma)$ to denote the set of all differentiable and σ-strongly convex functions that have their minimizer at $\boldsymbol{x}^* \in \mathbb{R}^n$. Define S^n to be the set of symmetric matrices in $\mathbb{R}^{n\times n}$, and Q^n to be the set of all quadratic functions that map $\mathbb{R}^n$ to $\mathbb{R}$. A quadratic function $f \in \mathsf{Q}^n$ parameterized by $\boldsymbol{Q} \in \mathsf{S}^n$, $\boldsymbol{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$ is given by

$$f(x; \boldsymbol{Q}, \boldsymbol{b}, c) = \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^\intercal \boldsymbol{x} + c.$$

For $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}_{>0}$, define

$$\mathcal{Q}^{(n)}(\boldsymbol{x}^*, \sigma) := \Big\{ f(x; \boldsymbol{Q}, \boldsymbol{b}, c) \in \mathsf{Q}^n : \lambda_{\min}(\boldsymbol{Q}) = \sigma, \;\; \boldsymbol{Q}\boldsymbol{x}^* = -\boldsymbol{b} \Big\}. \tag{2.4}$$

We will omit the superscript (n) of $\mathcal{Q}^{(n)}$ when it is clear from contexts. Note that every function in $\mathcal{Q}(\boldsymbol{x}^*, \sigma)$ is σ-strongly convex quadratic and has the minimizer at $\boldsymbol{x}^*$, and

$$\mathcal{Q}(\boldsymbol{x}^*, \sigma) \subset \bigcup_{\tilde{\sigma} \geq \sigma} \mathcal{Q}(\boldsymbol{x}^*, \tilde{\sigma}) \subset \mathcal{S}(\boldsymbol{x}^*, \sigma). \tag{2.5}$$

The following lemma shows that the strong convexity of functions is invariant under some particular affine transformations. This property will help us to simplify the analysis throughout this chapter.

Lemma 2.2.2. *Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be an orthogonal matrix and $\boldsymbol{b} \in \mathbb{R}^n$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function and define $h(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$. Then, f is σ-strongly convex if and only if h is σ-strongly convex.*

Proof. By the definition of strongly convex functions in (2.3), we have that

$$\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \rangle \geq \sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2 \quad \text{for all} \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

Since $\boldsymbol{A}$ is invertible, we can replace $\boldsymbol{x}$ and $\boldsymbol{y}$ by $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$, respectively, and the above inequality is equivalent to

$$\Big\langle \nabla f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) - \nabla f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}),\ \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{y}) \Big\rangle \geq \sigma \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{y})\|^2 \ \text{ for all } \ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n. \tag{2.6}$$

Since $\nabla h(\boldsymbol{x}) = \boldsymbol{A}^\intercal \nabla f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$, we can rewrite the LHS of (2.6) as $\Big\langle \boldsymbol{A}^\intercal \Big(\nabla f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) - \nabla f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})\Big),\ \boldsymbol{x} - \boldsymbol{y}\Big\rangle = \langle \nabla h(\boldsymbol{x}) - \nabla h(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y}\rangle$. On the other hand, since $\boldsymbol{A}$ is an orthogonal matrix, the RHS of (2.6) becomes $\sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2$. □

2.3 Problem Formulation

Consider two (unknown) functions f_1 and f_2. In order to investigate the minimizer of the sum of two unknown functions $f_1 + f_2$, we will impose the following assumptions on the structure of both functions.

1. Given $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, the functions $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ are differentiable and strongly convex with parameters σ_1 and σ_2, respectively.

2. Given $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$, the minimizers of f_1 and f_2 are at $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, respectively.

3. Suppose $\boldsymbol{x}^* \in \mathbb{R}^n$ is the minimizer of $f_1 + f_2$. There is a finite (given) number $L \in \mathbb{R}_{>0}$ such that the norm of gradient of f_1 and f_2 evaluated at $\boldsymbol{x}^*$ is less than L.

Assumption 1 and 2 will be captured using the notations introduced earlier: $f_1 \in \mathcal{S}(\boldsymbol{x}_1^*, \sigma_1)$ and $f_2 \in \mathcal{S}(\boldsymbol{x}_2^*, \sigma_2)$. For Assumption 3, since $\boldsymbol{x}^*$ is the minimizer of $f_1 + f_2$, we have that $\nabla f_1(\boldsymbol{x}^*) = -\nabla f_2(\boldsymbol{x}^*)$. In addition, we can rewrite the bounded gradient at $\boldsymbol{x}^*$ condition as $\|\nabla f_1(\boldsymbol{x}^*)\| = \|\nabla f_2(\boldsymbol{x}^*)\| \le L$. Essentially, our goal is to estimate the region $\mathcal{M}$ containing all possible values $\boldsymbol{x}^*$ satisfying the above conditions. More specifically, given $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, and $L \in \mathbb{R}_{>0}$, we wish to estimate the potential solution region

$$\begin{aligned}\mathcal{M}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L) := \Big\{\boldsymbol{x} \in \mathbb{R}^n : \exists f_1 \in \mathcal{S}(\boldsymbol{x}_1^*, \sigma_1), \quad \exists f_2 \in \mathcal{S}(\boldsymbol{x}_2^*, \sigma_2),\\ \nabla f_1(\boldsymbol{x}) = -\nabla f_2(\boldsymbol{x}), \quad \|\nabla f_1(\boldsymbol{x})\| = \|\nabla f_2(\boldsymbol{x})\| \le L\Big\}. \quad (2.7)\end{aligned}$$

For simplicity of notation, we will omit the argument of the set $\mathcal{M}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L)$ and write it as $\mathcal{M}$.

2.3.1 Discussion of Assumptions

Functions that satisfy both differentiable and strongly convex conditions (Assumption 1) are common in many applications. In machine learning applications, for example, linear regression and logistic regression models with L_2-regularization are commonly used when only a small amount of training data is available [1].

Assumption 2 can be generalized by assuming that for $i \in \{1, 2\}$, the minimizer $\boldsymbol{x}_i^*$ of the function f_i is not available but instead $\boldsymbol{x}_i^*$ is located in a known compact set $\mathcal{A}_i \subset \mathbb{R}^n$ as in [56]. However, the analysis will be more involved, so we defer these assumptions to our future works.

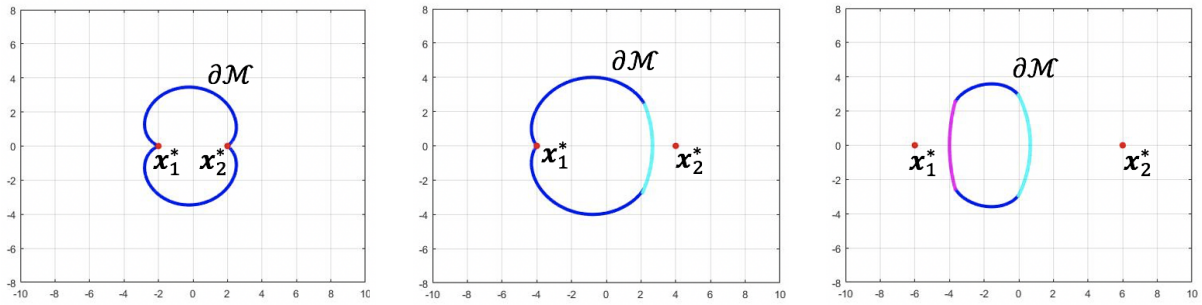
Assumption 3 is a technical assumption. Given $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$ such that $\boldsymbol{x}_1^* \neq \boldsymbol{x}_2^*$, let

$$\mathcal{L} = \Big\{ \boldsymbol{x} \in \mathbb{R}^n : \text{there exists } k \in \mathbb{R} \setminus (-1, 1) \text{ such that } \boldsymbol{x} - \boldsymbol{x}_1^* = k(\boldsymbol{x}_2^* - \boldsymbol{x}_1^*) \Big\}.$$

Without Assumption 3, i.e., the norm of the gradient of each function at the minimizer of the sum can be arbitrarily large, one can use the result from Proposition 2.5.1 to show that $\mathcal{M} = \mathbb{R}^n \setminus \mathcal{L}$. We can see that for $n \in \mathbb{N} \setminus \{1\}$, the set $\mathcal{L}$ has measure zero and hence, $\mathcal{M}$ covers almost the entire space. In other words, almost all points can be minimizers. One can think of imposing the bound on the gradients as one of the ways to implicitly limit the functions that we can choose from $\mathcal{S}(\boldsymbol{x}_1^*, \sigma_1)$ and $\mathcal{S}(\boldsymbol{x}_2^*, \sigma_2)$. However, there might be other ways to restrict the class of functions that we can select, for example, considering the functions that have Lipschitz continuous gradients in addition to Assumptions 1 and 2. For now, we restrict ourselves to the simpler assumption, Assumption 3, and leave such alternative assumptions for future work.

2.3.2 A Preview of the Solution

Recall the definition of the potential solution region $\mathcal{M}$ from (2.7). One way to characterize the set $\mathcal{M}$ is to provide an explicit formula for the boundary $\partial\mathcal{M}$ in terms of $\boldsymbol{x}_1^*$, $\boldsymbol{x}_2^*$, σ_1, σ_2 and L. In Fig. 2.1, we provide a preview of the boundary $\partial\mathcal{M}$ in $\mathbb{R}^2$ given fixed parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$, and $L = 10$, and a variable parameter $r \in \mathbb{R}_{>0}$. Suppose $\boldsymbol{x}_1^* = (-r, 0)$



(a) For $r = 2$, $\partial\mathcal{M}$ consists of 1 curve (blue curve) and $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.

(b) For $r = 4$, $\partial\mathcal{M}$ consists of 2 curves (blue and cyan curves) and $\{\boldsymbol{x}_1^*\}$.

(c) For $r = 6$, $\partial\mathcal{M}$ consists of 3 curves (blue, cyan and magenta curves).

Figure 2.1. The boundary $\partial\mathcal{M}$ in $\mathbb{R}^2$ is plotted given minimizers $\boldsymbol{x}_1^* = (-r, 0)$ and $\boldsymbol{x}_2^* = (r, 0)$ and fixed parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$, and $L = 10$. Different colors denote different equations that combine together to yield the boundary $\partial\mathcal{M}$.

and $\boldsymbol{x}_2^* = (r, 0)$. We illustrate $\partial\mathcal{M}$ for the case where $r = 2, 4$, and 6 in Fig. 2.1a, Fig. 2.1b and Fig. 2.1c, respectively. The different colors in the figures indicate different equations that combine together to yield the boundary (as we will explicitly characterize in the rest of this chapter).

2.3.3 Solution Approach

Since the analysis of the case $\boldsymbol{x}_1^* = \boldsymbol{x}_2^*$ is trivial (i.e., the potential solution region is $\mathcal{M} = \{\boldsymbol{x}_1^*\}$), without loss of generality, we assume that $\boldsymbol{x}_1^* = (-r, 0, \ldots, 0) \in \mathbb{R}^n$ and $\boldsymbol{x}_2^* = (r, 0, \ldots, 0) \in \mathbb{R}^n$ with $r = \frac{1}{2}\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\| > 0$.

To show this, given general $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$ with $\boldsymbol{x}_1^* \neq \boldsymbol{x}_2^*$, let the set of new bases $\mathcal{J} = \{\boldsymbol{e}_1', \boldsymbol{e}_2', \ldots, \boldsymbol{e}_n'\}$ be such that $\boldsymbol{e}_1' = \frac{\boldsymbol{x}_2^* - \boldsymbol{x}_1^*}{\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\|}$ and $\{\boldsymbol{e}_2', \boldsymbol{e}_3', \ldots, \boldsymbol{e}_n'\}$ is obtained by Gram-Schmidt orthogonalization. Let

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1' & \boldsymbol{e}_2' & \cdots & \boldsymbol{e}_n' \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \frac{1}{2}(\boldsymbol{x}_1^* + \boldsymbol{x}_2^*).$$

We let $\boldsymbol{x}_{\mathcal{J}} = \boldsymbol{E}^\intercal(\boldsymbol{x} - \boldsymbol{b})$ be the coordinate transformation. One can verify that if $\boldsymbol{x} = \boldsymbol{x}_1^*$ then $\boldsymbol{x}_{\mathcal{J}} = \left(-\frac{1}{2}\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\|, \ \boldsymbol{0}\right) = (-r, \boldsymbol{0})$ and if $\boldsymbol{x} = \boldsymbol{x}_2^*$ then $\boldsymbol{x}_{\mathcal{J}} = \left(\frac{1}{2}\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\|, \ \boldsymbol{0}\right) = (r, \boldsymbol{0})$. For $i \in \{1, 2\}$, let $\tilde{f}_i : \mathbb{R}^n \to \mathbb{R}$ be the function such that $\tilde{f}_i(\boldsymbol{x}_{\mathcal{J}}) = f_i(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$,

i.e., $\tilde{f}_i$'s value at the coordinate of point $\boldsymbol{x}$ on the new bases $\mathcal{J}$ is the same as f_i at point $\boldsymbol{x}$. We can write $\tilde{f}_i(\boldsymbol{x}) = f_i(\boldsymbol{E}\boldsymbol{x} + \boldsymbol{b})$ for $i \in \{1, 2\}$. Applying Lemma 2.2.2, we have that $\tilde{f}_i$ is σ_i-strongly convex for $i \in \{1, 2\}$. Once we attain the potential solution region $\mathcal{M}$ in terms of $\boldsymbol{x}_{\mathcal{J}}$, we can always use the transformation to obtain the region in terms of $\boldsymbol{x}$, i.e., the original coordinate system.

For convenience, we introduce the shorthand notation of sets that will be encountered throughout this chapter. Recall the definition of $\mathcal{B}$ from (2.2). For $i \in \{1, 2\}$, define

$$\mathcal{B}_i := \mathcal{B}\Big(\boldsymbol{x}_i^*, \frac{L}{\sigma_i}\Big). \tag{2.8}$$

Now, we introduce the functions that will be used to define the outer and inner approximations of $\mathcal{M}$. For $i \in \{1, 2\}$, define the functions $\tilde{\phi}_i : \overline{\mathcal{B}}_i \to \left[0, \frac{\pi}{2}\right]$ to be such that

$$\tilde{\phi}_i(\boldsymbol{x}) := \arccos\left(\frac{\sigma_i}{L}\|\boldsymbol{x} - \boldsymbol{x}_i^*\|\right), \tag{2.9}$$

and the functions $\alpha_i : \mathbb{R}^n \setminus \{\boldsymbol{x}_i^*\} \to [0, \boldsymbol{\pi}]$ to be such that

$$\alpha_i(\boldsymbol{x}) := \angle(\boldsymbol{x} - \boldsymbol{x}_i^*,\ \boldsymbol{x}_2^* - \boldsymbol{x}_1^*), \tag{2.10}$$

i.e., the angle between vectors $\boldsymbol{x} - \boldsymbol{x}_i^*$ and $\boldsymbol{x}_2^* - \boldsymbol{x}_1^*$. Note that $\alpha_2(\boldsymbol{x}) \geq \alpha_1(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ due to the assumption that $\boldsymbol{x}_1^* = (-r, \boldsymbol{0})$ and $\boldsymbol{x}_2^* = (r, \boldsymbol{0})$. We define $\psi : \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \to [0, \boldsymbol{\pi}]$ to be the function such that

$$\psi(\boldsymbol{x}) := \boldsymbol{\pi} - \Big(\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\Big). \tag{2.11}$$

The interpretation of the angles $\tilde{\phi}_i(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ will be clarified later (in Fig. 2.2). In addition, given $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, and $L \in \mathbb{R}_{>0}$, we define

$$\mathcal{X} := \begin{cases} \Big\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_i^*\| = \frac{L}{\sigma_i} \ \text{ for all } i \in \{1, 2\}\Big\} & \text{if } \|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\| = L\Big(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\Big), \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.12}$$

Due to the assumption that $\boldsymbol{x}_1^* = (-r, \mathbf{0})$ and $\boldsymbol{x}_2^* = (r, \mathbf{0})$, for $\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\| = L\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$, we have $\mathcal{X} = \left\{\left(-r + \frac{L}{\sigma_1},\ \mathbf{0}\right)\right\}$.

With these definitions in place, given $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^n$, $\sigma_1, \sigma_2 \in \mathbb{R}_{>0}$, and $L \in \mathbb{R}_{>0}$, we define the outer and inner approximations of $\mathcal{M}$ as

$$\mathcal{M}^{\uparrow}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L) := \left\{\boldsymbol{x} \in \mathbb{R}^n :\ \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})\right\} \tag{2.13}$$

and

$$\mathcal{M}_{\downarrow}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L) := \left\{\boldsymbol{x} \in \mathbb{R}^n :\ \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})\right\} \cup \mathcal{X}, \tag{2.14}$$

respectively. As before, we will omit the argument of the sets $\mathcal{M}^{\uparrow}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L)$ and $\mathcal{M}_{\downarrow}(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \sigma_1, \sigma_2, L)$, and write them as $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$, respectively.

Remark 1. Recall the definition of $\tilde{\phi}_i$ for $i \in \{1, 2\}$ and ψ from (2.9) and (2.11), respectively. Since $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$ are defined using $\tilde{\phi}_1$, $\tilde{\phi}_2$ and ψ, implicitly, they must be subsets of $\mathbf{dom}(\tilde{\phi}_1) \cap \mathbf{dom}(\tilde{\phi}_2) \cap \mathbf{dom}(\psi)$. In other words, the sets $\mathcal{M}^{\uparrow} \subseteq (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ and $\mathcal{M}_{\downarrow} \subseteq (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ where $\mathcal{B}_i$ for $i \in \{1, 2\}$ are defined in (2.8).

In order to characterize the potential solution region $\mathcal{M}$, we proceed as follows. First, in Proposition 2.4.1, we show that $\mathcal{M} \subseteq \mathcal{M}^{\uparrow}$ by considering a property of strongly convex functions. Then, we characterize the boundary and interior of the outer approximation ($\partial\mathcal{M}^{\uparrow}$ and $(\mathcal{M}^{\uparrow})^{\circ}$) for each value of r in Theorem 2.4.8. In Proposition 2.5.1, we consider quadratic functions and show that $\mathcal{M}_{\downarrow} \subseteq \mathcal{M}$ in Proposition 2.5.2. We use a similar approach as in Theorem 2.4.8 to characterize the boundary and interior of the inner approximation ($\partial\mathcal{M}_{\downarrow}$ and $(\mathcal{M}_{\downarrow})^{\circ}$) for each value of r which is presented in Theorem 2.5.2. Finally, by observing that $\partial\mathcal{M}^{\uparrow} = \partial\mathcal{M}_{\downarrow}$ and $(\mathcal{M}^{\uparrow})^{\circ} = (\mathcal{M}_{\downarrow})^{\circ}$ from Theorem 2.4.8 and Theorem 2.5.2, we conclude the chapter by showing that, in fact, the boundary of the potential solution region, outer approximation, and inner approximation are identical, i.e., $\partial\mathcal{M} = \partial\mathcal{M}^{\uparrow} = \partial\mathcal{M}_{\downarrow}$, in Theorem 2.6.1.

2.4 Outer Approximation

In this section, we derive necessary conditions for a point to be in the potential solution region $\mathcal{M}$ and show that $\mathcal{M} \subseteq \mathcal{M}^{\uparrow}$ in Proposition 2.4.1. Then, we explicitly characterize an important part of $\partial\mathcal{M}^{\uparrow}$ (and also $\partial\mathcal{M}_{\downarrow}$) in Proposition 2.4.2. In Theorem 2.4.8, which is the main result of this section, we identify $\partial\mathcal{M}^{\uparrow}$ and $(\mathcal{M}^{\uparrow})^{\circ}$, and also provide a property of $\mathcal{M}^{\uparrow}$. Other lemmas in this section are presented as tools that will be utilized in the proof of Theorem 2.4.8 (and also Theorem 2.5.2).

We will be using the following functions throughout our analysis. For $i \in \{1,2\}$, define $\boldsymbol{u}_i : \mathbb{R}^n \setminus \{\boldsymbol{x}_i^*\} \to \mathbb{R}^n$ to be the function such that

$$\boldsymbol{u}_i(\boldsymbol{x}) := \frac{\boldsymbol{x} - \boldsymbol{x}_i^*}{\|\boldsymbol{x} - \boldsymbol{x}_i^*\|}, \tag{2.15}$$

i.e., the unit vector in the direction of $\boldsymbol{x} - \boldsymbol{x}_i^*$. Recall the definition of $\angle(\cdot,\cdot)$ from (2.1). For $i \in \{1,2\}$, we define $\phi_i : \mathbb{R}^n \setminus \{\boldsymbol{x}_i^*\} \to \left[0, \frac{\pi}{2}\right]$ to be the function such that

$$\phi_i(\boldsymbol{x}) := \angle\Big(\nabla f_i(\boldsymbol{x}),\ \boldsymbol{u}_i(\boldsymbol{x})\Big), \tag{2.16}$$

and $\underline{L}_i : \mathbb{R}^n \to \mathbb{R}$ to be the function such that

$$\underline{L}_i(\boldsymbol{x}) := \sigma_i \|\boldsymbol{x} - \boldsymbol{x}_i^*\|. \tag{2.17}$$

Note that for $i \in \{1,2\}$, the quantity $\underline{L}_i(\boldsymbol{x})$ is a lower bound on the norm of the gradient of f_i at $\boldsymbol{x} \in \mathbb{R}^n$ if $f_i \in \mathcal{S}(\boldsymbol{x}_i^*, \sigma_i)$.

In Fig. 2.2, we illustrate the definition of $\boldsymbol{u}_i$, ϕ_i, $\tilde{\phi}_i$, α_i for $i \in \{1,2\}$, and ψ. Moreover, we illustrate the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})$ which is used to describe the outer approximation $\mathcal{M}^{\uparrow}$ in (2.13).

In the following proposition, we show a crucial result that the set $\mathcal{M}^{\uparrow}$ covers the set that we want to characterize, $\mathcal{M}$. In other words, the points in the set $\mathcal{M}^{\uparrow}$ satisfy necessary conditions of a point to be a minimizer of the sum $f_1 + f_2$.

Proposition 2.4.1. *Suppose the sets $\mathcal{M}$ and $\mathcal{M}^{\uparrow}$ are defined as in (2.7) and (2.13), respectively. Then, $\mathcal{M} \subseteq \mathcal{M}^{\uparrow}$.*

Proof. Recall the definition of the sets $\mathcal{B}_i$ for $i \in \{1,2\}$, the angles $\tilde{\phi}_i$ for $i \in \{1,2\}$, and the angle ψ from (2.8), (2.9), and (2.11), respectively. First, we want to show that the necessary conditions for a point $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ to be in $\mathcal{M}$ are

(i) $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, and

(ii) $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})$.

From the definition of strongly convex functions in (2.3), we have

$$\left\langle \nabla f_i(\boldsymbol{x}) - \nabla f_i(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \right\rangle \geq \sigma_i \|\boldsymbol{x} - \boldsymbol{y}\|^2$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and for $i \in \{1,2\}$. For $i \in \{1,2\}$, recall the definition of $\boldsymbol{u}_i(\boldsymbol{x})$ and $\phi_i(\boldsymbol{x})$ from (2.15) and (2.16), respectively. Since $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ are the minimizers of f_1 and f_2, respectively, for $\boldsymbol{x} \notin \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we get

$$\begin{aligned} & \left\langle \nabla f_i(\boldsymbol{x}) - \nabla f_i(\boldsymbol{x}_i^*),\ \boldsymbol{x} - \boldsymbol{x}_i^* \right\rangle \geq \sigma_i \|\boldsymbol{x} - \boldsymbol{x}_i^*\|^2, \\ \Leftrightarrow \quad & \|\nabla f_i(\boldsymbol{x})\| \cos(\phi_i(\boldsymbol{x})) = \langle \nabla f_i(\boldsymbol{x}),\ \boldsymbol{u}_i(\boldsymbol{x}) \rangle \geq \sigma_i \|\boldsymbol{x} - \boldsymbol{x}_i^*\| > 0. \end{aligned} \tag{2.18}$$

Suppose $\boldsymbol{x}$ is a candidate minimizer. Then, we have that $\|\nabla f_i(\boldsymbol{x})\| \leq L$ for $i \in \{1,2\}$ by our assumption. Recall the definition of $\underline{L}_i$ for $i \in \{1,2\}$ from (2.17). Inequality (2.18) becomes

$$\cos(\phi_i(\boldsymbol{x})) \geq \frac{\sigma_i}{L} \|\boldsymbol{x} - \boldsymbol{x}_i^*\| = \frac{\underline{L}_i(\boldsymbol{x})}{L}. \tag{2.19}$$

If $\underline{L}_1(\boldsymbol{x}) > L$ or $\underline{L}_2(\boldsymbol{x}) > L$, we have that $\boldsymbol{x}$ cannot be the minimizer of the function $f_1 + f_2$ since there is no $\phi_i(\boldsymbol{x})$ that can satisfy inequality (2.19). Thus, a necessary condition for $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ to be a minimizer of $f_1 + f_2$ is that $\underline{L}_i(\boldsymbol{x}) \leq L$ for $i \in \{1,2\}$ or equivalently, $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, yielding part (i) of the claim. We now prove part (ii).

From the definition of $\psi(\boldsymbol{x})$ in (2.11) and that $\angle(-\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x}))$, $\alpha_1(\boldsymbol{x})$ and $\boldsymbol{\pi} - \alpha_2(\boldsymbol{x})$ are the angles of the triangle formed by the points $\boldsymbol{x}$, $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$, we can write that for all $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$,

$$\psi(\boldsymbol{x}) = (\boldsymbol{\pi} - \alpha_2(\boldsymbol{x})) + \alpha_1(\boldsymbol{x}) = \boldsymbol{\pi} - \angle(-\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x})) = \angle(\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x})). \tag{2.20}$$

Suppose that $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$. Recall the definition of $\tilde{\phi}_i$ for $i \in \{1, 2\}$ from (2.9). From inequality (2.19), we have $\phi_i(\boldsymbol{x}) \leq \tilde{\phi}_i(\boldsymbol{x})$ for $i \in \{1, 2\}$. If $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) < \psi(\boldsymbol{x})$, then using (2.16) and (2.20), we have

$$\angle(\nabla f_1(\boldsymbol{x}), \boldsymbol{u}_1(\boldsymbol{x})) + \angle(-\nabla f_2(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x})) < \angle(\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x})).$$

However, using [57, Corollary 12], we can write $\angle(\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x})) \leq \angle(\nabla f_1(\boldsymbol{x}), \boldsymbol{u}_1(\boldsymbol{x})) + \angle(\nabla f_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x}))$. Therefore, if $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) < \psi(\boldsymbol{x})$, we have that $\nabla f_1(\boldsymbol{x}) \neq -\nabla f_2(\boldsymbol{x})$ which implies that $\boldsymbol{x}$ is not the minimizer of $f_1 + f_2$. This means that one of the necessary conditions is that $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})$ which completes the proof of the claim.

In the above analysis, we considered the case when $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. We are left with the case when $\boldsymbol{x} \in \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. From the definition of strongly convex functions, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

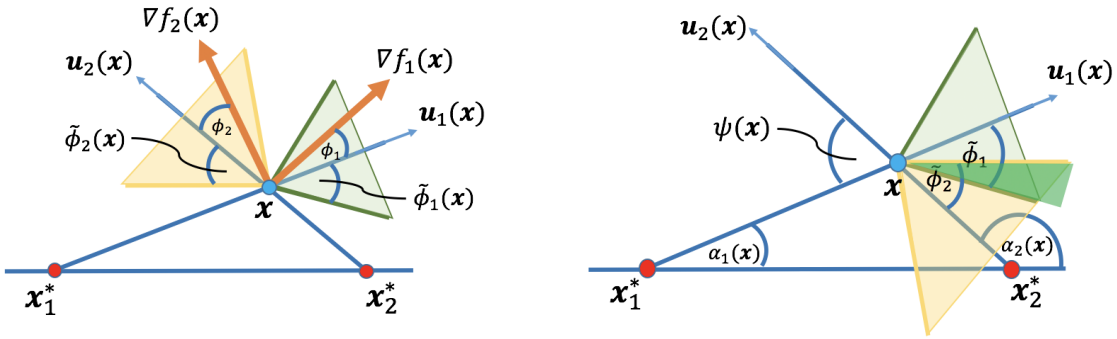
$$\left\langle \nabla f_2(\boldsymbol{x}) - \nabla f_2(\boldsymbol{y}),\ \boldsymbol{x} - \boldsymbol{y} \right\rangle \geq \sigma_2 \|\boldsymbol{x} - \boldsymbol{y}\|^2.$$

Since $\boldsymbol{x}_2^*$ is the minimizer of f_2 and $\boldsymbol{x}_1^* \neq \boldsymbol{x}_2^*$, we get

$$\left\langle \nabla f_2(\boldsymbol{x}_1^*),\ \boldsymbol{x}_1^* - \boldsymbol{x}_2^* \right\rangle = \left\langle \nabla f_2(\boldsymbol{x}_1^*) - \nabla f_2(\boldsymbol{x}_2^*),\ \boldsymbol{x}_1^* - \boldsymbol{x}_2^* \right\rangle \geq \sigma_2 \|\boldsymbol{x}_1^* - \boldsymbol{x}_2^*\|^2 > 0,$$

and thus, $\nabla f_2(\boldsymbol{x}_1^*) \neq \boldsymbol{0}$. This implies that $\nabla f_2(\boldsymbol{x}_1^*) + \nabla f_1(\boldsymbol{x}_1^*) \neq \boldsymbol{0}$ and $\boldsymbol{x}_1^*$ is not the minimizer of $f_1 + f_2$. By using similar approach, we can also conclude that $\boldsymbol{x}_2^*$ is not the minimizer of $f_1 + f_2$. □

Remark 2. The angle functions $\tilde{\phi}_i$ and α_i for $i \in \{1, 2\}$ defined in (2.16) and (2.10), respectively, can be expressed as functions of the distances $\|\boldsymbol{x}_1^* - \boldsymbol{x}_2^*\|$, $\|\boldsymbol{x} - \boldsymbol{x}_1^*\|$, and $\|\boldsymbol{x} - \boldsymbol{x}_2^*\|$.



(a) (b)

Figure 2.2. (a) The figure illustrates the definition of $\boldsymbol{u}_i$, ϕ_i, and $\tilde{\phi}_i$ for $i \in \{1, 2\}$. In particular, inequality (2.19) implies that $\phi_i(\boldsymbol{x}) \in [0, \tilde{\phi}_i(\boldsymbol{x})]$ for $i \in \{1, 2\}$, i.e., the gradient vectors $\nabla f_1(\boldsymbol{x})$ and $\nabla f_2(\boldsymbol{x})$ must lie in the corresponding shaded regions. (b) The figure illustrates the definition of α_i for $i \in \{1, 2\}$ and ψ. In addition, the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})$ in $\mathcal{M}^{\uparrow}$ means that there is an overlapping region (light green region in the figure) caused by one shaded region and the mirror of the other shaded region.

This means that the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \geq \psi(\boldsymbol{x})$ depends only on the distance among three points $\boldsymbol{x}$, $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$. Since $\boldsymbol{x}_1^* = (-r, \mathbf{0})$ and $\boldsymbol{x}_2^* = (r, \mathbf{0})$, we conclude that the shape of $\mathcal{M}^{\uparrow}$ (and $\mathcal{M}_{\downarrow}$) is symmetric around x_1-axis.

From this point, we will denote $\boldsymbol{x} = (x_1, \tilde{\mathbf{x}}) \in \mathbb{R}^n$ where $x_1 \in \mathbb{R}$ and $\tilde{\mathbf{x}} = (x_2, x_3, \ldots, x_n) \in \mathbb{R}^{n-1}$. Next, we will provide an algebraic expression for a certain portion of $\partial\mathcal{M}^{\uparrow}$ (and $\partial\mathcal{M}_{\downarrow}$) based on the geometric equation $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x})$, where $\tilde{\phi}_i$ for $i \in \{1, 2\}$ and ψ are defined in (2.9) and (2.11), respectively. For convenience, we define

$$\begin{aligned} d_1(\boldsymbol{x}) &:= \|\boldsymbol{x} - \boldsymbol{x}_1^*\| = \sqrt{(x_1 + r)^2 + \|\tilde{\mathbf{x}}\|^2} \quad \text{and} \\ d_2(\boldsymbol{x}) &:= \|\boldsymbol{x} - \boldsymbol{x}_2^*\| = \sqrt{(x_1 - r)^2 + \|\tilde{\mathbf{x}}\|^2}. \end{aligned} \tag{2.21}$$

Define the set of points

$$\mathcal{T} := \left\{ \boldsymbol{x} \in \mathbb{R}^n : \ \frac{\|\boldsymbol{x}\|^2 - r^2}{d_1^2(\boldsymbol{x}) \cdot d_2^2(\boldsymbol{x})} + \frac{\sigma_1 \sigma_2}{L^2} = \sqrt{\frac{1}{d_1^2(\boldsymbol{x})} - \frac{\sigma_1^2}{L^2}} \cdot \sqrt{\frac{1}{d_2^2(\boldsymbol{x})} - \frac{\sigma_2^2}{L^2}} \right\}. \tag{2.22}$$

Proposition 2.4.2. *The set $\mathcal{T}$ defined in (2.22) can equivalently be written as $\mathcal{T} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x}) \right\}$.*

Proof. Based on the definition of $\alpha_i(\boldsymbol{x})$ for $i \in \{1, 2\}$ in (2.10), for any point $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we have

$$\begin{aligned} & x_1 = d_1(\boldsymbol{x}) \cos(\alpha_1(\boldsymbol{x})) - r = d_2(\boldsymbol{x}) \cos(\alpha_2(\boldsymbol{x})) + r, \\ \Leftrightarrow \quad & \cos(\alpha_1(\boldsymbol{x})) = \frac{x_1 + r}{d_1(\boldsymbol{x})} \quad \text{and} \quad \cos(\alpha_2(\boldsymbol{x})) = \frac{x_1 - r}{d_2(\boldsymbol{x})}. \end{aligned} \tag{2.23}$$

Similarly,

$$\begin{aligned} & \|\tilde{\mathbf{x}}\| = d_1(\boldsymbol{x}) \sin(\alpha_1(\boldsymbol{x})) = d_2(\boldsymbol{x}) \sin(\alpha_2(\boldsymbol{x})), \\ \Leftrightarrow \quad & \sin(\alpha_1(\boldsymbol{x})) = \frac{\|\tilde{\mathbf{x}}\|}{d_1(\boldsymbol{x})} \quad \text{and} \quad \sin(\alpha_2(\boldsymbol{x})) = \frac{\|\tilde{\mathbf{x}}\|}{d_2(\boldsymbol{x})}. \end{aligned} \tag{2.24}$$

Since $\tilde{\phi}_i(\boldsymbol{x}) \in \left[0, \frac{\pi}{2}\right]$ for $i \in \{1,2\}$, we get $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \in [0, \boldsymbol{\pi}]$. Recall from (2.11) that $\psi(\boldsymbol{x}) \in [0, \boldsymbol{\pi}]$. Since the cosine function is one-to-one for this range of angles, equation $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x})$ is equivalent to

$$\cos\Big(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})\Big) = \cos\Big(\boldsymbol{\pi} - (\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x}))\Big) = -\cos\Big(\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\Big).$$

Expanding this equation and substituting (2.23), (2.24), and $\cos(\tilde{\phi}_i(\boldsymbol{x})) = \frac{\sigma_i}{L} d_i(\boldsymbol{x})$ for $i \in \{1,2\}$, we get

$$\frac{\sigma_1}{L} d_1(\boldsymbol{x}) \cdot \frac{\sigma_2}{L} d_2(\boldsymbol{x}) - \sqrt{1 - \left(\frac{\sigma_1}{L} d_1(\boldsymbol{x})\right)^2} \cdot \sqrt{1 - \left(\frac{\sigma_2}{L} d_2(\boldsymbol{x})\right)^2} = -\frac{x_1 - r}{d_2(\boldsymbol{x})} \cdot \frac{x_1 + r}{d_1(\boldsymbol{x})} - \frac{\|\tilde{\mathbf{x}}\|}{d_2(\boldsymbol{x})} \cdot \frac{\|\tilde{\mathbf{x}}\|}{d_1(\boldsymbol{x})}.$$

Dividing the above equation by $d_1(\boldsymbol{x}) \cdot d_2(\boldsymbol{x})$ and rearranging it yields the result. □

The subsequent lemmas (Lemma 2.4.1 - 2.4.7) are useful ingredients for proving the characterization of the outer approximation $\mathcal{M}^{\uparrow}$ (defined in (2.13)) given in Theorem 2.4.8, and their proofs are provided in Appendix A.1.

The following lemma provides a sufficient condition for the minimizers $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ to be on the boundary of the outer approximation $\mathcal{M}^{\uparrow}$ and the inner approximation $\mathcal{M}_{\downarrow}$.

Lemma 2.4.1. *Let $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$ be as defined in (2.13) and (2.14), respectively.*

(i) If $r \in \left(0, \frac{L}{2\sigma_2}\right]$ then $\boldsymbol{x}_1^* \in \partial\mathcal{M}^{\uparrow}$ and $\boldsymbol{x}_1^* \in \partial\mathcal{M}_{\downarrow}$.

(ii) If $r \in \left(0, \frac{L}{2\sigma_1}\right]$ then $\boldsymbol{x}_2^* \in \partial\mathcal{M}^{\uparrow}$ and $\boldsymbol{x}_2^* \in \partial\mathcal{M}_{\downarrow}$.

In the next lemma, we provide a property of points in a particular set which will be used to characterize the sets $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$ defined in (2.13) and (2.14), respectively. Roughly speaking, if $\boldsymbol{x} \in \mathcal{M}^{\uparrow}$ and $x_1 \in [-r, r]$, then each point that has the same first component and is closer to the x_1-axis is also in $\mathcal{M}^{\uparrow}$.

Lemma 2.4.2. *Consider two points $\boldsymbol{x} = (x_1, \tilde{\mathbf{x}})$ and $\boldsymbol{y} = (y_1, \tilde{\mathbf{y}})$. Suppose $-r \le x_1 = y_1 \le r$ and $\|\tilde{\mathbf{x}}\| > \|\tilde{\mathbf{y}}\|$. If $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \ge \psi(\boldsymbol{x})$ then either $\tilde{\phi}_1(\boldsymbol{y}) + \tilde{\phi}_2(\boldsymbol{y}) > \psi(\boldsymbol{y})$ or $\boldsymbol{y} \in \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.*

Recall the definition of $\mathcal{B}$ from (2.8). Since $\boldsymbol{x}_1^* = (-r, \mathbf{0})$ and $\boldsymbol{x}_2^* = (r, \mathbf{0})$ by our assumption, we can explicitly write $\partial\mathcal{B}_i$ for $i \in \{1, 2\}$ as follows:

$$\begin{aligned} \partial\mathcal{B}_1 &= \partial\mathcal{B}\left(\boldsymbol{x}_1^*, \frac{L}{\sigma_1}\right) = \left\{\boldsymbol{x} \in \mathbb{R}^n : (x_1 + r)^2 + \|\tilde{\mathbf{x}}\|^2 = \frac{L^2}{\sigma_1^2}\right\}, \\ \partial\mathcal{B}_2 &= \partial\mathcal{B}\left(\boldsymbol{x}_2^*, \frac{L}{\sigma_2}\right) = \left\{\boldsymbol{x} \in \mathbb{R}^n : (x_1 - r)^2 + \|\tilde{\mathbf{x}}\|^2 = \frac{L^2}{\sigma_2^2}\right\}. \end{aligned} \tag{2.25}$$

For convenience, we define

$$\gamma_i := \frac{L^2}{\sigma_i^2} \quad \text{for} \quad i \in \{1, 2\} \quad \text{and} \quad \beta := \frac{\sigma_2}{\sigma_1}. \tag{2.26}$$

By using the definitions above, we define

$$\lambda_1 := \left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} \quad \text{and} \quad \lambda_2 := -\left(\frac{1+\beta}{2+\beta}\right)\frac{\gamma_2}{2r} + \frac{\beta r}{2+\beta}. \tag{2.27}$$

In the following lemma, we will show that if $\boldsymbol{x} \in \{\partial\mathcal{B}_1, \partial\mathcal{B}_2\}$, the value of the first component x_1 is necessary and sufficient to determine whether $\boldsymbol{x}$ is in $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$, which are defined in (2.13) and (2.14), respectively. In other words, the angle condition $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \lesseqgtr \psi(\boldsymbol{x})$ can be simplified if we consider a point in $\partial\mathcal{B}_1$ or $\partial\mathcal{B}_2$.

Lemma 2.4.3. *Let λ_1 and λ_2 be as defined in (2.27). Consider $\boldsymbol{x} = (x_1, \tilde{\mathbf{x}}) \in (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.*

(i) If $\boldsymbol{x} \in \partial\mathcal{B}_1$ then $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \lesseqgtr \psi(\boldsymbol{x})$ if and only if $x_1 \lesseqgtr \lambda_1$.

(ii) If $\boldsymbol{x} \in \partial\mathcal{B}_2$ then $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \lesseqgtr \psi(\boldsymbol{x})$ if and only if $x_1 \gtreqless \lambda_2$.

In the following lemma, we will show that the points in the set of intersection between $\mathcal{T}$ and $\partial\mathcal{B}_1$ (resp. $\mathcal{T}$ and $\partial\mathcal{B}_2$) have the same first component, if the intersection is non-empty.

Moreover, the first component of these points is λ_1 (resp. λ_2) where λ_i for $i \in \{1,2\}$ are defined in (2.27). By using the definition of γ_1, γ_2 and β in (2.26), define

$$\begin{aligned} \nu_1 &:= \frac{r}{2(1+2\beta)}\sqrt{-\left(\frac{\gamma_1}{r^2}-4\right)\left((1+\beta)^2\frac{\gamma_1}{r^2}-4\beta^2\right)} \quad \text{and} \\ \nu_2 &:= \frac{r}{2(2+\beta)}\sqrt{-\left(\frac{\gamma_2}{r^2}-4\right)\left((1+\beta)^2\frac{\gamma_2}{r^2}-4\right)}. \end{aligned} \tag{2.28}$$

Lemma 2.4.4. *Consider the sets of points $\mathcal{T}$ and $\partial\mathcal{B}_i$ for $i \in \{1,2\}$ defined in (2.22) and (2.25), respectively. Let λ_i and ν_i for $i \in \{1,2\}$ be as defined in (2.27) and (2.28), respectively.*

(i) For $i \in \{1,2\}$, if $r \in \left(0, \frac{L}{2\sigma_i}\right]$, then $\mathcal{T} \cap \partial\mathcal{B}_i = \emptyset$.

(ii) For $i \in \{1,2\}$, if $r \in \left(\frac{L}{2\sigma_i}, \frac{L}{2}(\frac{1}{\sigma_1}+\frac{1}{\sigma_2})\right]$, then $\mathcal{T} \cap \partial\mathcal{B}_i = \left\{\boldsymbol{x} \in \mathbb{R}^n : x_1 = \lambda_i, \ \ \|\tilde{\mathbf{x}}\| = \nu_i\right\}$.

Recall that λ_1 and λ_2 are defined in (2.27). In the following lemma, for $i \in \{1,2\}$, we consider a relationship between $\frac{L}{\sigma_i r}$ and $\frac{\lambda_i}{r}$. In particular, for $r \in \left(0, \frac{L}{2}(\frac{1}{\sigma_1}+\frac{1}{\sigma_2})\right]$, recall from Lemma 2.4.4 that if $\mathcal{T} \cap \partial\mathcal{B}_1 \neq \emptyset$ (resp. $\mathcal{T} \cap \partial\mathcal{B}_2 \neq \emptyset$), then every point in the intersection has the first component equal to λ_1 (resp. λ_2). The next lemma compares λ_1 to the maximum value of the first component over all points of $\partial\mathcal{B}_1$ (which is $-r+\frac{L}{\sigma_1}$), and compares λ_2 to the minimum value of the first component over all points of $\partial\mathcal{B}_2$ (which is $r-\frac{L}{\sigma_1}$), respectively.

Lemma 2.4.5. *Let λ_i for $i \in \{1,2\}$ be as defined in (2.27).*

(i) If $r \in \left(0, \frac{L}{2\sigma_1}\right]$ then $\lambda_1 \geq \frac{L}{\sigma_1} - r$, with equality only if $r = \frac{L}{2\sigma_1}$.

(ii) $r \in \left(\frac{L}{2\sigma_1}, \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ if and only if $\lambda_1 < \frac{L}{\sigma_1} - r$.

(iii) If $r \in \left(0, \frac{L}{2\sigma_2}\right]$ then $\lambda_2 \leq r - \frac{L}{\sigma_2}$, with equality only if $r = \frac{L}{2\sigma_2}$.

(iv) $r \in \left(\frac{L}{2\sigma_2}, \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ if and only if $\lambda_2 > r - \frac{L}{\sigma_2}$.

In the next lemma, we will consider a relationship between $\mathcal{T}$ and $\partial\mathcal{M}^{\uparrow}$ (the boundary of the outer approximation), and $\mathcal{T}$ and $\partial\mathcal{M}_{\downarrow}$ (the boundary of the inner approximation). In particular, we will show that $\mathcal{T} \subseteq \partial\mathcal{M}^{\uparrow}$ and $\mathcal{T} \subseteq \partial\mathcal{M}_{\downarrow}$ for a particular range of r.

Lemma 2.4.6. *Let $\mathcal{M}^{\uparrow}$, $\mathcal{M}_{\downarrow}$ and $\mathcal{T}$ be defined as in (2.13), (2.14) and (2.22), respectively. If $r \in \left(0,\ \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$, then $\mathcal{T} \subseteq \partial\mathcal{M}^{\uparrow}$ and $\mathcal{T} \subseteq \partial\mathcal{M}_{\downarrow}$.*

For $i \in \{1,2\}$, define half-planes

$$\mathcal{H}_i^+ := \{\boldsymbol{x} \in \mathbb{R}^n : x_1 \geq \lambda_i\} \quad \text{and} \quad \mathcal{H}_i^- := \{\boldsymbol{x} \in \mathbb{R}^n : x_1 \leq \lambda_i\}, \tag{2.29}$$

where λ_i for $i \in \{1,2\}$ are defined in (2.27). In the lemma below, we examine properties of points $\boldsymbol{x} \in \partial\mathcal{B}_1 \cap \mathcal{H}_1^+$ (resp. $\boldsymbol{x} \in \partial\mathcal{B}_2 \cap \mathcal{H}_2^-$).

Lemma 2.4.7. *Let the sets $\mathcal{B}_i$ for $i \in \{1,2\}$, and $\mathcal{H}_1^+$ and $\mathcal{H}_2^-$ be defined as in (2.8) and (2.29), respectively.*

(i) If $r \in \left(\frac{L}{2\sigma_1}, \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ then $\left[\lambda_1,\ -r+\frac{L}{\sigma_1}\right] \subseteq (-r,r)$ and $\partial\mathcal{B}_1 \cap \mathcal{H}_1^+ \subseteq (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.

(ii) If $r \in \left(\frac{L}{2\sigma_2}, \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ then $\left[r-\frac{L}{\sigma_2},\ \lambda_2\right] \subseteq (-r,r)$ and $\partial\mathcal{B}_2 \cap \mathcal{H}_2^- \subseteq (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.

In the theorem below, we give the characterization of the boundary $\partial\mathcal{M}^{\uparrow}$ and interior $(\mathcal{M}^{\uparrow})^{\circ}$, and also a property of the set $\mathcal{M}^{\uparrow}$ for each range of r. Define the set

$$\widetilde{\mathcal{T}} := \{\boldsymbol{x} \in \mathbb{R}^n : \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})\}, \tag{2.30}$$

which will be used especially in Theorem 2.4.8 and Theorem 2.5.2.

Theorem 2.4.8. *Assume $\sigma_1 \geq \sigma_2$. Let the sets $\mathcal{M}^{\uparrow}$, $\mathcal{T}$, $\widetilde{\mathcal{T}}$, and $\mathcal{B}_i$ for $i \in \{1,2\}$ be defined as in (2.13), (2.22), (2.30), and (2.8), respectively. Also, let the sets $\mathcal{H}_i^+$ and $\mathcal{H}_i^-$ for $i \in \{1,2\}$ be defined as in (2.29).*

(i) If $r \in \left(0,\ \frac{L}{2\sigma_1}\right]$ then $\mathcal{M}^{\uparrow} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ is closed,

$$\partial\mathcal{M}^{\uparrow} = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \quad \text{and} \quad (\mathcal{M}^{\uparrow})^{\circ} = \widetilde{\mathcal{T}}.$$

(ii) If $r \in \left(\frac{L}{2\sigma_1},\ \frac{L}{2\sigma_2}\right]$ *then* $\mathcal{M}^{\uparrow} \sqcup \{\boldsymbol{x}_1^*\}$ *is closed,*

$$\begin{aligned}\partial\mathcal{M}^{\uparrow} &= \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\} \quad \textit{and}\\ (\mathcal{M}^{\uparrow})^{\circ} &= \left[\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\widetilde{\mathcal{T}} \cap \mathcal{H}_1^-\right].\end{aligned}$$

(iii) If $r \in \left(\frac{L}{2\sigma_2},\ \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ *then* $\mathcal{M}^{\uparrow}$ *is closed,*

$$\begin{aligned}\partial\mathcal{M}^{\uparrow} &= \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \mathcal{T} \quad \textit{and}\\ (\mathcal{M}^{\uparrow})^{\circ} &= \left[\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \left[\widetilde{\mathcal{T}} \cap (\mathcal{H}_1^- \cap \mathcal{H}_2^+)\right].\end{aligned}$$

(iv) If $r = \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)$ *then* $\mathcal{M}^{\uparrow} = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\}$.

(v) If $r \in \left(\frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right),\ \infty\right)$ *then* $\mathcal{M}^{\uparrow} = \emptyset$.

Proof. For convenience, we define function $\varphi : (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \to [-\boldsymbol{\pi}, \boldsymbol{\pi}]$ to be such that

$$\varphi(\boldsymbol{x}) := \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) - \psi(\boldsymbol{x}), \tag{2.31}$$

where $\tilde{\phi}_i$ for $i \in \{1,2\}$ and ψ are defined in (2.9) and (2.11), respectively.

Part (i): $r \in \left(0,\ \frac{L}{2\sigma_1}\right]$. First, we want to show that

$$\text{if} \quad \boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2) \quad \text{then} \quad \boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) < 0\} \cup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}. \tag{2.32}$$

Suppose $\boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2)$ and $\boldsymbol{x} \in \partial\mathcal{B}_1$. Since $\overline{\mathcal{B}}_1$ is closed and $x_1 \in \left[-r-\frac{L}{\sigma_1}, -r+\frac{L}{\sigma_1}\right]$, from Lemma 2.4.5 part (i), we get $x_1 \le \frac{L}{\sigma_1} - r \le \lambda_1$. If $x_1 < \lambda_1$, from Lemma 2.4.3 part (i), we obtain $\varphi(\boldsymbol{x}) < 0$. On the other hand, if $x_1 = \lambda_1$ (i.e., $\frac{L}{\sigma_1} - r = \lambda_1$), from Lemma 2.4.5 part (i), we get $r = \frac{L}{2\sigma_1}$. Substituting into $x_1 = \frac{L}{\sigma_1} - r$, we obtain that $x_1 = \frac{L}{2\sigma_1} = r$. Since $\boldsymbol{x} \in \partial\mathcal{B}(\boldsymbol{x}_1^*, 2r)$ and $x_1 = r$, we conclude that $\boldsymbol{x} = \boldsymbol{x}_2^* = (r, \mathbf{0})$.

From the assumption $\sigma_1 \ge \sigma_2$ and the inequality $r \le \frac{L}{2\sigma_1}$, we get $r \le \frac{L}{2\sigma_2}$. We can similarly show that if $\boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2)$ and $\boldsymbol{x} \in \partial\mathcal{B}_2$ then either $\varphi(\boldsymbol{x}) < 0$ or $\boldsymbol{x} = \boldsymbol{x}_1^* = (-r, \mathbf{0})$

by using Lemma 2.4.5 part (iii) and Lemma 2.4.3 part (ii). Since $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \subseteq \partial\mathcal{B}_1 \cup \partial\mathcal{B}_2$, we have proved our claim.

Since $\mathcal{M}^\uparrow \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$ from the definition of $\mathcal{M}^\uparrow$ in (2.13) and $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \subset (\mathcal{M}^\uparrow)^c$ from (2.32), we have $\mathcal{M}^\uparrow \subseteq \mathcal{B}_1 \cap \mathcal{B}_2$. Recall the definition of φ in (2.31). Let $\mathcal{R} = (\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. We then partition the set $\mathcal{R}$ into 3 parts as follows:

$$\mathcal{R}_1 = \Big\{\boldsymbol{z} \in \mathcal{R} : \varphi(\boldsymbol{z}) > 0\Big\}, \quad \mathcal{R}_2 = \Big\{\boldsymbol{z} \in \mathcal{R} : \varphi(\boldsymbol{z}) < 0\Big\},$$
$$\text{and} \quad \mathcal{R}_3 = \Big\{\boldsymbol{z} \in \mathcal{R} : \varphi(\boldsymbol{z}) = 0\Big\} = \mathcal{T},$$

where the last equality comes from Proposition 2.4.2. We will show that

$$\begin{cases} \mathcal{R}_1 \subset (\partial\mathcal{M}^\uparrow)^c, \\ \mathcal{R}_2 \subset (\partial\mathcal{M}^\uparrow)^c, \\ \mathcal{R}_3 \subseteq \partial\mathcal{M}^\uparrow, \\ \partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subset (\partial\mathcal{M}^\uparrow)^c, \\ (\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subset (\partial\mathcal{M}^\uparrow)^c. \end{cases} \tag{2.33}$$

Suppose $\boldsymbol{x} \in \mathcal{R}_1$. Since φ is continuous, there exists $\epsilon > 0$ such that for all $\boldsymbol{x}_0 \in \mathcal{B}(\boldsymbol{x}, \epsilon)$, we have $\boldsymbol{x}_0 \in \mathcal{R}_1$ and $\varphi(\boldsymbol{x}_0) > 0$. Since $\mathcal{R}_1 \subseteq \mathcal{M}^\uparrow$ and is open, we have $\mathcal{R}_1 \subseteq (\mathcal{M}^\uparrow)^\circ$. Similarly, we have $\mathcal{R}_2 \subseteq (\mathcal{R} \setminus \mathcal{M}^\uparrow)^\circ$. Suppose $\boldsymbol{x} \in \mathcal{R}_3 = \mathcal{T}$. Using Lemma 2.4.6, we have that $\mathcal{R}_3 \subseteq \partial\mathcal{M}^\uparrow$. Since $(\mathcal{M}^\uparrow)^\circ$, $(\mathcal{R} \setminus \mathcal{M}^\uparrow)^\circ$, and $\partial\mathcal{M}^\uparrow$ are disjoint, we conclude that $\mathcal{R}_1 \subset (\partial\mathcal{M}^\uparrow)^c$ and $\mathcal{R}_2 \subset (\partial\mathcal{M}^\uparrow)^c$.

Consider $\boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. From (2.32), we have $\boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) < 0\}$. Since $\mathbf{dom}(\varphi) = (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ and φ is continuous, there exists $\epsilon > 0$ such that for all $\boldsymbol{x}_0 \in \mathcal{B}(\boldsymbol{x}, \epsilon) \cap \mathbf{dom}(\varphi)$, we have $\varphi(\boldsymbol{x}_0) < 0$. Thus, $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subset ((\mathcal{M}^\uparrow)^c)^\circ$ which implies that $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subset (\partial\mathcal{M}^\uparrow)^c$. In addition, we have $(\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subseteq ((\mathcal{M}^\uparrow)^c)^\circ$ (since $(\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subseteq (\mathcal{M}^\uparrow)^c$ and is open) which implies $(\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subset (\partial\mathcal{M}^\uparrow)^c$. Therefore, we have proved the claim (2.33).

Since we can partition $\mathbb{R}^n$ into $\mathcal{R}$, $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, $(\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ and $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, using (2.33), we obtain that $\partial\mathcal{M}^\uparrow \subseteq \mathcal{R}_3 \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. However, we know that $\mathcal{R}_3 \subseteq \partial\mathcal{M}^\uparrow$ from the above analysis and $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subseteq \partial\mathcal{M}^\uparrow$ from Lemma 2.4.1. Thus, we have $\partial\mathcal{M}^\uparrow = \mathcal{R}_3 \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ by Proposition 2.4.2.

From Proposition 2.4.2, we have $\mathcal{T} = \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) = 0\}$. Using the definition of $\mathcal{M}^\uparrow$ in (2.13) and $\partial\mathcal{M}^\uparrow = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we can write $(\mathcal{M}^\uparrow)^\circ = \mathcal{M}^\uparrow \setminus \partial\mathcal{M}^\uparrow = \widetilde{\mathcal{T}}$ where $\widetilde{\mathcal{T}}$ is defined in (2.30). Since $\mathcal{T} \subseteq \mathcal{M}^\uparrow$, this implies that $\partial\mathcal{M}^\uparrow = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subseteq \mathcal{M}^\uparrow \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. Thus, the set $\mathcal{M}^\uparrow \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ is closed.

Part (ii): $r \in \left(\frac{L}{2\sigma_1}, \frac{L}{2\sigma_2}\right]$. We separately consider three disjoint regions: $(\mathcal{H}_1^-)^c$, $\mathcal{H}_1^+ \cap \mathcal{H}_1^-$, $(\mathcal{H}_1^+)^c$. For the first region, we want to show that

$$\mathcal{T} \cap (\mathcal{H}_1^-)^c = \emptyset \quad \text{and} \quad \partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c = \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c. \tag{2.34}$$

From Lemma 2.4.7 part (i), we have $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c \subseteq \partial\mathcal{B}_1 \cap \mathcal{H}_1^+ \subseteq \mathbf{dom}(\varphi)$. Consider $\boldsymbol{x} \in \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$ and note that $x_1 \in \left(\lambda_1, \ -r + \frac{L}{\sigma_1}\right]$. From Lemma 2.4.3 part (i), we have $\varphi(\boldsymbol{x}) > 0$. Since $\left[\lambda_1, \ -r + \frac{L}{\sigma_1}\right] \subseteq (-r, r)$ from Lemma 2.4.7 part (i), we can apply Lemma 2.4.2 to get that for all $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c$, we have $\varphi(\boldsymbol{x}) > 0$. This implies that

$$\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{T}^c \quad \text{and} \quad \overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c, \tag{2.35}$$

by Proposition 2.4.2 and the definition of $\mathcal{M}^\uparrow$ in (2.13), respectively. Since $\mathcal{T} \subseteq \mathbf{dom}(\varphi)$, we have $(\overline{\mathcal{B}}_1)^c \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{T}^c$. Using this inclusion and the first inclusion in (2.35), we can write

$$\emptyset = \left[\mathcal{T} \cap \left(\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c\right)\right] \cup \left[\mathcal{T} \cap \left((\overline{\mathcal{B}}_1)^c \cap (\mathcal{H}_1^-)^c\right)\right] = \mathcal{T} \cap (\mathcal{H}_1^-)^c, \tag{2.36}$$

which completes the first part of claim (2.34). Next, note that since $-r + \frac{L}{\sigma_1} < r$, we have $\boldsymbol{x}_2^* \in \mathcal{H}_1^+$. However, we have $\mathcal{M}^\uparrow \subseteq \overline{\mathcal{B}}_1$ from the definition of $\mathcal{M}^\uparrow$ in (2.13). Combine this argument with the second inclusion in (2.35) yields

$$\mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c = \overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c. \tag{2.37}$$

Since $(\mathcal{H}_1^-)^c$ is open, (2.37) implies that $(\mathcal{M}^\uparrow)^\circ \cap (\mathcal{H}_1^-)^c = \mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$. Then, subtracting this equation from (2.37), we obtain that $\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c = \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$, which completes the second part of claim (2.34).

Next, consider the second region $\mathcal{H}_1^+ \cap \mathcal{H}_1^- = \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1\}$. Recall the definition of ν_1 in (2.28). Consider the following three cases.

- Suppose $\boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| > \nu_1\}$. Then, $\boldsymbol{x} \in (\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ which implies that $\boldsymbol{x} \notin \mathcal{T}$. Since $(\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \subseteq (\mathcal{M}^\uparrow)^c$ and is open, we also have $\boldsymbol{x} \notin \partial\mathcal{M}^\uparrow$.

- Suppose $\boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| = \nu_1\}$. From Lemma 2.4.4 part (ii), we have $\boldsymbol{x} \in \mathcal{T}$. Using Lemma 2.4.6, we obtain that $\boldsymbol{x} \in \partial\mathcal{M}^\uparrow$.

- Suppose $\boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| < \nu_1\}$. Since $\{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| = \nu_1\} \subseteq \mathcal{T}$, from Lemma 2.4.2, we have $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})$ which implies that $\boldsymbol{x} \notin \mathcal{T}$. Since $\boldsymbol{x} \in (\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ and φ is continuous, there exists $\epsilon \in \mathbb{R}_{>0}$ such that for all $\boldsymbol{x}_0 \in \mathcal{B}(\boldsymbol{x}, \epsilon)$, we have $\boldsymbol{x}_0 \in \mathcal{M}^\uparrow$ by the definition of $\mathcal{M}^\uparrow$ in (2.13). This means that $\boldsymbol{x} \in (\mathcal{M}^\uparrow)^\circ$ and thus, $\boldsymbol{x} \notin \partial\mathcal{M}^\uparrow$.

Combining the analysis of these three cases, we have that

$$\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) = \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| = \nu_1\} = \mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-). \tag{2.38}$$

Next, consider the third region $(\mathcal{H}_1^+)^c = \mathbb{R}^n \setminus \mathcal{H}_1^+$. First, we want to show that

$$\text{if} \quad \boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c \quad \text{then} \quad \boldsymbol{x} \in \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) < 0\} \cup \{\boldsymbol{x}_1^*\}. \tag{2.39}$$

Suppose $\boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2)$ and $\boldsymbol{x} \in \partial\mathcal{B}_1 \cap (\mathcal{H}_1^+)^c$. Since $x_1 < \lambda_1$, from Lemma 2.4.3 part (i), we obtain $\varphi(\boldsymbol{x}) < 0$. By using the result from the proof of part (i), we have that if $\boldsymbol{x} \in \partial(\mathcal{B}_1 \cap \mathcal{B}_2)$ and $\boldsymbol{x} \in \partial\mathcal{B}_2$ then either $\varphi(\boldsymbol{x}) < 0$ or $\boldsymbol{x} = \boldsymbol{x}_1^*$. Combining the two results, we have proved the claim.

Since $\mathcal{M}^\uparrow \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$ from the definition of $\mathcal{M}^\uparrow$ in (2.13) and $\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c \subset (\mathcal{M}^\uparrow)^c$ from (2.39), we have $\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c \subseteq (\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c$. Let $\mathcal{R}' = \mathcal{R} \cap (\mathcal{H}_1^+)^c$. We then

partition the set $\mathcal{R}'$ into $\mathcal{R}'_1$, $\mathcal{R}'_2$, and $\mathcal{R}'_3$ where $\mathcal{R}'_i = \mathcal{R}_i \cap (\mathcal{H}_1^+)^c$ for $i \in \{1,2,3\}$. We can use a similar argument as in the proof of (2.33) to show that

$$\begin{cases} \mathcal{R}'_1 \subset (\partial\mathcal{M}^\uparrow)^c \cap (\mathcal{H}_1^+)^c, \\ \mathcal{R}'_2 \subset (\partial\mathcal{M}^\uparrow)^c \cap (\mathcal{H}_1^+)^c, \\ \mathcal{R}'_3 \subseteq \partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c, \\ (\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*\}) \cap (\mathcal{H}_1^+)^c \subset (\partial\mathcal{M}^\uparrow)^c \cap (\mathcal{H}_1^+)^c, \\ \Big((\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*\}\Big) \cap (\mathcal{H}_1^+)^c \subset (\partial\mathcal{M}^\uparrow)^c \cap (\mathcal{H}_1^+)^c. \end{cases} \tag{2.40}$$

Since we can partition $(\mathcal{H}_1^+)^c$ into $\mathcal{R}'$, $(\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*\}) \cap (\mathcal{H}_1^+)^c$, $\Big((\mathbf{dom}(\varphi))^c \setminus \{\boldsymbol{x}_1^*\}\Big) \cap (\mathcal{H}_1^+)^c$ and $\{\boldsymbol{x}_1^*\}$, using (2.40), we obtain that $\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c \subseteq \mathcal{R}'_3 \sqcup \{\boldsymbol{x}_1^*\}$. However, we know that $\mathcal{R}'_3 \subseteq \partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c$ from (2.40) and $\{\boldsymbol{x}_1^*\} \subseteq \partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c$ from Lemma 2.4.1. Thus, using $\mathcal{R}'_3 = \mathcal{R}_3 \cap (\mathcal{H}_1^+)^c$ and Proposition 2.4.2 we have

$$\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c = \mathcal{R}'_3 \sqcup \{\boldsymbol{x}_1^*\} = \Big[\mathcal{T} \cap (\mathcal{H}_1^+)^c\Big] \sqcup \{\boldsymbol{x}_1^*\}. \tag{2.41}$$

Since $\mathbb{R}^n = (\mathcal{H}_1^-)^c \sqcup (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) \sqcup (\mathcal{H}_1^+)^c$, using (2.34), (2.38) and (2.41), we obtain that

$$\partial\mathcal{M}^\uparrow = \Big[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\Big] \sqcup \Big[\mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\Big] \sqcup \Big[\mathcal{T} \cap (\mathcal{H}_1^+)^c\Big] \sqcup \{\boldsymbol{x}_1^*\}. \tag{2.42}$$

However, from (2.36), we can write $\mathcal{T} = \Big[\mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\Big] \sqcup \Big[\mathcal{T} \cap (\mathcal{H}_1^+)^c\Big]$ which means that we can rewrite (2.42) as

$$\partial\mathcal{M}^\uparrow = \Big[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\Big] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\}. \tag{2.43}$$

From (2.37) and (2.43), we can write $\Big[\mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c\Big] \setminus \Big[\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^-)^c\Big] = \Big[\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c\Big] \setminus \Big[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\Big] = \mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$. From the definition of $\mathcal{M}^\uparrow$ in (2.13) and equation (2.38), we can write $\Big[\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\Big] \setminus \Big[\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\Big] = \widetilde{\mathcal{T}} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)$. From the definition of $\mathcal{M}^\uparrow$ in (2.13) and equation (2.41), we can write $(\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c) \setminus (\partial\mathcal{M}^\uparrow \cap (\mathcal{H}_1^+)^c) = \widetilde{\mathcal{T}} \cap (\mathcal{H}_1^+)^c$. Applying the above three equations to Lemma 2.2.1, we obtain the result of $(\mathcal{M}^\uparrow)^\circ$. Consider

the characterization of $\partial\mathcal{M}^{\uparrow}$ in (2.43). Since $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{M}^{\uparrow}$ from (2.37), and $\mathcal{T} \subseteq \mathcal{M}^{\uparrow}$ from Proposition 2.4.2 and the definition of $\mathcal{M}^{\uparrow}$, we can write $\partial\mathcal{M}^{\uparrow} \subseteq \mathcal{M}^{\uparrow} \sqcup \{\boldsymbol{x}_1^*\}$ and thus, $\mathcal{M}^{\uparrow} \sqcup \{\boldsymbol{x}_1^*\}$ is closed.

Part (iii): $r \in \left(\frac{L}{2\sigma_2},\ \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$. We can use a similar argument as in the proof of part (ii) to show that

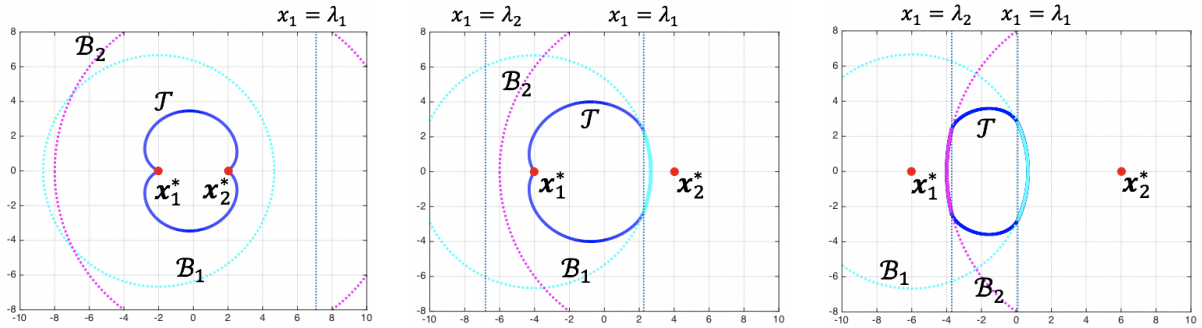
$$\begin{cases} \partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_1^-)^c = \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{M}^{\uparrow}, & \text{(similar to proving (2.34))} \\ \partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_2^+)^c = \partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c \subseteq \mathcal{M}^{\uparrow}, & \text{(similar to proving (2.34))} \\ \partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) = \mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) \subseteq \mathcal{M}^{\uparrow}, & \text{(similar to proving (2.38))} \\ \partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_2^+ \cap \mathcal{H}_2^-) = \mathcal{T} \cap (\mathcal{H}_2^+ \cap \mathcal{H}_2^-) \subseteq \mathcal{M}^{\uparrow}, & \text{(similar to proving (2.38))} \\ \partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c = \mathcal{T} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c \subseteq \mathcal{M}^{\uparrow}. & \text{(similar to proving (2.41))} \end{cases}$$

Similar to (2.36), in this case, we have that $\mathcal{T} \cap (\mathcal{H}_1^-)^c = \emptyset$ and $\mathcal{T} \cap (\mathcal{H}_2^+)^c = \emptyset$. This means that the last three equations regarding $\partial\mathcal{M}^{\uparrow}$ above can be combined into $\partial\mathcal{M}^{\uparrow} \cap (\mathcal{H}_1^- \cap \mathcal{H}_2^+) = \mathcal{T}$. Combining this equation with the first two equations regarding $\partial\mathcal{M}^{\uparrow}$ above, we obtain the characterization of $\partial\mathcal{M}^{\uparrow}$. For the characterization of $(\mathcal{M}^{\uparrow})^{\circ}$, we can use the same technique as shown in the analysis of part (ii) to obtain the result. From the five inclusions regarding $\partial\mathcal{M}^{\uparrow}$ above, we can write $\partial\mathcal{M}^{\uparrow} \subseteq \mathcal{M}^{\uparrow}$ and we conclude that $\mathcal{M}^{\uparrow}$ is closed.

Part (iv): $r = \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)$. In this case, we have $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2 = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\}$. Suppose $\boldsymbol{x} = \left(\frac{L}{2}\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right),\ \mathbf{0}\right)$. Since $\mathcal{M}^{\uparrow} \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, we only need to check point $\boldsymbol{x}$. At this point, we get $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x}) = 0$ (since $d_1(\boldsymbol{x}) = \frac{L}{\sigma_1}$, $d_2(\boldsymbol{x}) = \frac{L}{\sigma_2}$, $\alpha_1(\boldsymbol{x}) = 0$ and $\alpha_2(\boldsymbol{x}) = \boldsymbol{\pi}$). So, we conclude that $\mathcal{M}^{\uparrow} = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\}$.

Part (v): $r \in \left(\frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right),\ \infty\right)$. Since $r > \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)$, we have $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2 = \emptyset$. Since $\mathcal{M}^{\uparrow} \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, we conclude that $\mathcal{M}^{\uparrow} = \emptyset$. □

Examples of the boundary $\partial\mathcal{M}^{\uparrow}$ in $\mathbb{R}^2$ for the first three cases of Theorem 2.4.8 are shown in Fig. 2.3. We consider parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$ and $L = 10$. For $r = 2$, as we can see from Fig. 2.3a, we have $\partial\mathcal{M}^{\uparrow} = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ (i.e., solid blue line + two red dots) consistent with part (i) of Theorem 2.4.8. For $r = 4$, as we can see from Fig. 2.3b, we have $\partial\mathcal{M}^{\uparrow} = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\}$ (i.e., solid cyan line + solid blue line + left red



(a) For $r = 2$, the characterization of boundary $\partial\mathcal{M}^\uparrow$ corresponds to Theorem 2.4.8 part (i).

(b) For $r = 4$, the characterization of boundary $\partial\mathcal{M}^\uparrow$ corresponds to Theorem 2.4.8 part (ii).

(c) For $r = 6$, the characterization of boundary $\partial\mathcal{M}^\uparrow$ corresponds to Theorem 2.4.8 part (iii).

Figure 2.3. The boundary $\partial\mathcal{M}^\uparrow$ in $\mathbb{R}^2$ with different values of r for two original minimizers $x_1^* = (-r, 0)$ and $x_2^* = (r, 0)$ is plotted given fixed parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$, and $L = 10$. The sets $\mathcal{T}$, $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$ and $\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c$ are shown by solid blue, cyan and magenta lines, respectively. The vertical dotted lines represent the equations $x_1 = \lambda_1$ and $x_1 = \lambda_2$ and note that the value of λ_1 and λ_2 depends on r.

dot) consistent with part (ii) of Theorem 2.4.8. For $r = 6$, as we can see from Fig. 2.3c, we have $\partial\mathcal{M}^\uparrow = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \mathcal{T}$ (i.e., solid cyan line + solid magenta line + solid blue line) consistent with part (iii) of Theorem 2.4.8. Note that the solid blue line, solid cyan line and solid magenta line in the figures indicate that the corresponding sets of points are subsets of the outer approximation $\mathcal{M}^\uparrow$, i.e., $\mathcal{T} \subseteq \mathcal{M}^\uparrow$, $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{M}^\uparrow$ and $\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c \subseteq \mathcal{M}^\uparrow$, respectively.

2.5 Inner Approximation and Potential Solution Region

In the previous section, we showed that the set $\mathcal{M}^\uparrow$ defined in (2.13) is an outer approximation for the desired set $\mathcal{M}$ defined in (2.7), in that $\mathcal{M} \subseteq \mathcal{M}^\uparrow$. We now turn our attention to the set $\mathcal{M}_\downarrow$ defined in (2.14). We will show that $\mathcal{M}_\downarrow \subseteq \mathcal{M}$, and consequently, provide a tight characterization of $\mathcal{M}$.

Since we have $\bigcup_{\tilde{\sigma} \geq \sigma} \mathcal{Q}(\boldsymbol{x}^*, \tilde{\sigma}) \subset \mathcal{S}(\boldsymbol{x}^*, \sigma)$ for all $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}_{>0}$ from (2.5), we can provide a region contained in the potential solution region $\mathcal{M}$ by restricting our consideration to only some classes of quadratic functions. In Section 2.5.1, we analyze a sufficient and

necessary condition for constructing a quadratic function with a given minimizer, gradient and curvature. Then, using the result from Section 2.5.1, in Section 2.5.2, we prove a relationship between the potential solution region $\mathcal{M}$ and the inner approximation $\mathcal{M}_{\downarrow}$, and also provide a characterization of $\mathcal{M}_{\downarrow}$.

2.5.1 Quadratic Functions Analysis

In this subsection, first we consider an equivalent condition for the existence of a quadratic function with a given minimizer, gradient at a specific point and the smallest eigenvalue associated to the quadratic term in n-dimensional space (i.e., in $\mathbb{R}^n$) which is presented in Proposition 2.5.1. Then, we present Corollary 2.5.1 in which we provide a similar equivalent condition for the class $\bigcup_{\tilde{\sigma}\geq\sigma}\mathcal{Q}(\boldsymbol{x}^*,\tilde{\sigma})$ for a given $\boldsymbol{x}^*\in\mathbb{R}^n$ and $\sigma\in\mathbb{R}_{>0}$.

In the following proposition (whose proof is provided in Appendix A.2.1), we consider an equivalent condition for the existence of a quadratic function with more than one independent variable satisfying certain properties.

Proposition 2.5.1. *Let $\mathcal{Q}$ be defined as in (2.4). For $n\in\mathbb{N}\setminus\{1\}$, suppose we are given points $\boldsymbol{x}^*\in\mathbb{R}^n$ and $\boldsymbol{x}_0\in\mathbb{R}^n$ such that $\boldsymbol{x}_0\neq\boldsymbol{x}^*$, vector $\boldsymbol{g}\in\mathbb{R}^n$, and scalar $\sigma\in\mathbb{R}_{>0}$. Then, there exists a function $f\in\mathcal{Q}^{(n)}(\boldsymbol{x}^*,\sigma)$ with a gradient $\nabla f(\boldsymbol{x}_0)=\boldsymbol{g}$ if and only if*

(i) $\boldsymbol{x}_0\in\overline{\mathcal{B}}\Big(\boldsymbol{x}^*,\frac{\|\boldsymbol{g}\|}{\sigma}\Big)$ *and*

(ii) $\angle(\boldsymbol{g},\boldsymbol{x}_0-\boldsymbol{x}^*)\in\{0\}\cup\Big[0,\ \arccos\Big(\frac{\sigma}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0-\boldsymbol{x}^*\|\Big)\Big)$.

Note that if $\sigma\|\boldsymbol{x}_0-\boldsymbol{x}^*\|=\|\boldsymbol{g}\|$, then $\Big[0,\ \arccos(\frac{\sigma}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0-\boldsymbol{x}^*\|)\Big)=\emptyset$.

Recall from (2.5) that $\bigcup_{\hat{\sigma}\geq\sigma}\mathcal{Q}(\boldsymbol{x}^*,\hat{\sigma})\subset\mathcal{S}(\boldsymbol{x}^*,\sigma)$ for all $\boldsymbol{x}^*\in\mathbb{R}^n$ and $\sigma\in\mathbb{R}_{>0}$. One way to characterize the inner approximation $\mathcal{M}_{\downarrow}$ is to utilize sufficient conditions for the construction of a function $f\in\bigcup_{\hat{\sigma}\geq\sigma}\mathcal{Q}(\boldsymbol{x}^*,\hat{\sigma})$. More generally, the following corollary presents necessary and sufficient conditions for such construction.

Corollary 2.5.1. *Let $\mathcal{Q}$ be defined as in (2.4). For $n\in\mathbb{N}\setminus\{1\}$, suppose we are given points $\boldsymbol{x}^*\in\mathbb{R}^n$ and $\boldsymbol{x}_0\in\mathbb{R}^n$ such that $\boldsymbol{x}_0\neq\boldsymbol{x}^*$, a vector $\boldsymbol{g}\in\mathbb{R}^n$, scalar $L\in\mathbb{R}_{>0}$ such that $\|\boldsymbol{g}\|=L$, and a scalar $\sigma\in\mathbb{R}_{>0}$. Then, there exists a function $f\in\bigcup_{\hat{\sigma}\geq\sigma}\mathcal{Q}(\boldsymbol{x}^*,\hat{\sigma})$ with $\nabla f(\boldsymbol{x}_0)=\boldsymbol{g}$ if and only if*

(i) $\boldsymbol{x}_0 \in \overline{\mathcal{B}}\Big(\boldsymbol{x}^*, \frac{L}{\sigma}\Big)$ *and*

(ii) $\angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) \in \{0\} \cup \Big[0,\ \arccos\Big(\frac{\sigma}{L}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\Big)\Big)$.

Note that if $\sigma\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| = L$, *then* $\Big[0,\ \arccos(\frac{\sigma}{L}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|)\Big) = \emptyset$.

Proof. From Proposition 2.5.1, we can write that there exists a function $f \in \bigcup_{\hat{\sigma}\geq\sigma} \mathcal{Q}(\boldsymbol{x}^*, \hat{\sigma})$ with a gradient $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$ and $\|\nabla f(\boldsymbol{x}_0)\| = L$ if and only if

$$\boldsymbol{x}_0 \in \bigcup_{\hat{\sigma}\geq\sigma} \overline{\mathcal{B}}\Big(\boldsymbol{x}^*, \frac{\|\boldsymbol{g}\|}{\hat{\sigma}}\Big) = \overline{\mathcal{B}}\Big(\boldsymbol{x}^*, \frac{L}{\sigma}\Big),$$

and

$$\begin{aligned}\angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) &\in \{0\} \cup \bigcup_{\hat{\sigma}\geq\sigma} \Big[0,\ \arccos\Big(\frac{\hat{\sigma}}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\Big)\Big) \\ &= \{0\} \cup \Big[0,\ \arccos\Big(\frac{\sigma}{L}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\Big)\Big).\end{aligned}$$

□

2.5.2 Inner Approximation Characterization

In this subsection, we use results from Section 2.5.1 to derive a sufficient condition for a point to be in the potential solution region $\mathcal{M}$, defined in (2.7). In fact, the sufficient condition is encapsulated in the description of the inner approximation $\mathcal{M}_\downarrow$; therefore, $\mathcal{M}_\downarrow \subseteq \mathcal{M}$ which is presented in Proposition 2.5.2. Then, in Theorem 2.5.2, we characterize the boundary $\partial\mathcal{M}_\downarrow$ and interior $(\mathcal{M}_\downarrow)^\circ$, and provide a property of $\mathcal{M}_\downarrow$ similar to Theorem 2.4.8.

Recall the definition of $\underline{L}_i$ for $i \in \{1, 2\}$ from (2.17). Given $i \in \{1, 2\}$, $\boldsymbol{x}_i^* \in \mathbb{R}^n$, $\sigma_i \in \mathbb{R}_{>0}$, and $L \in \mathbb{R}_{>0}$, from Corollary 2.5.1, we define the set of gradient angles $\angle(\nabla f_i(\boldsymbol{x}),\ \boldsymbol{x} - \boldsymbol{x}_i^*)$ that we can choose to construct a quadratic function $f_i \in \bigcup_{\hat{\sigma}_i\geq\sigma_i} \mathcal{Q}(\boldsymbol{x}_i^*, \hat{\sigma}_i)$ with $\|\nabla f_i(\boldsymbol{x})\| \leq L$ as $\Phi_i : \overline{\mathcal{B}}\Big(\boldsymbol{x}_i^*, \frac{L}{\sigma_i}\Big) \to 2^{[0,\frac{\pi}{2})}$ with

$$\Phi_i(\boldsymbol{x}) := \begin{cases} \Big[0,\ \arccos\Big(\frac{\underline{L}_i(\boldsymbol{x})}{L}\Big)\Big) & \text{if} \quad \underline{L}_i(\boldsymbol{x}) < L, \\ \{0\} & \text{if} \quad \underline{L}_i(\boldsymbol{x}) = L. \end{cases} \tag{2.44}$$

Notice that the supremum of the set of angles $\Phi_i(\boldsymbol{x})$ is equal to $\tilde{\phi}_i(\boldsymbol{x})$ which is defined in (2.9). That is, for $i \in \{1,2\}$, for all $\boldsymbol{x} \in \overline{\mathcal{B}}\left(\boldsymbol{x}_i^*, \frac{L}{\sigma_i}\right)$, we have

$$\sup \Phi_i(\boldsymbol{x}) = \arccos\left(\frac{L_i(\boldsymbol{x})}{L}\right) = \tilde{\phi}_i(\boldsymbol{x}).$$

In two-dimensional space, for a given $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, for $i \in \{1,2\}$, the set of admissible angles $\Phi_i(\boldsymbol{x})$ and the quantity $\tilde{\phi}_i(\boldsymbol{x})$ are shown in Fig. 2.4a. In the next proposition, using Corollary 2.5.1, we show that the inner approximation $\mathcal{M}_\downarrow$ is contained in the potential solution region $\mathcal{M}$.

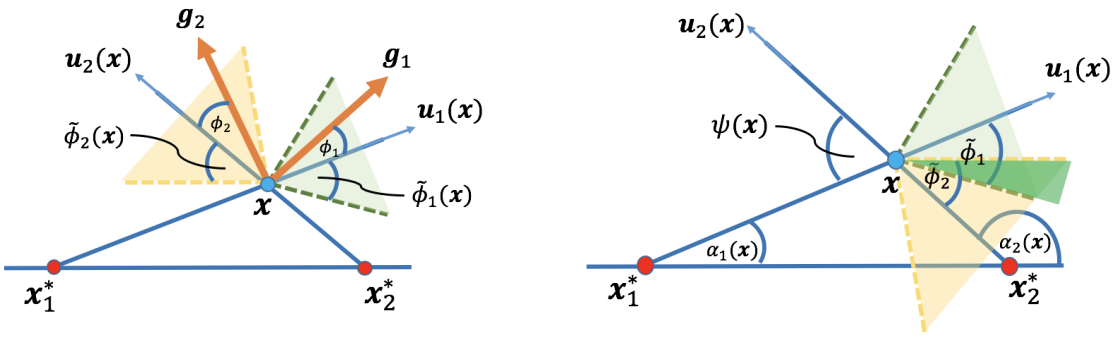
Proposition 2.5.2. *Suppose the sets $\mathcal{M}$ and $\mathcal{M}_\downarrow$ are defined as in (2.7) and (2.14), respectively. Then, $\mathcal{M} \supseteq \mathcal{M}_\downarrow$.*

Proof. Suppose $\boldsymbol{x} \in (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. Recall the definition of $\mathcal{X}$ from (2.12). We want to show that if $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})$ or $\boldsymbol{x} \in \mathcal{X}$, then $\boldsymbol{x} \in \mathcal{M}$. Recall the definition of $\boldsymbol{u}_i(\boldsymbol{x})$ for $i \in \{1,2\}$ from (2.15). Consider the following two cases.

- Suppose $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})$. Since $\psi(\boldsymbol{x}) = \angle(\boldsymbol{u}_1(\boldsymbol{x}), -\boldsymbol{u}_2(\boldsymbol{x}))$ from (2.20), there exists a vector $\boldsymbol{g} \in \mathbb{R}^n$ with $\|\boldsymbol{g}\| = L$ such that $\angle(\boldsymbol{g}, \boldsymbol{u}_1(\boldsymbol{x})) < \tilde{\phi}_1(\boldsymbol{x})$ and $\angle(\boldsymbol{g}, -\boldsymbol{u}_2(\boldsymbol{x})) < \tilde{\phi}_2(\boldsymbol{x})$. By the definition of Φ_i in (2.44), this means that $\angle(\boldsymbol{g}, \boldsymbol{u}_1(\boldsymbol{x})) \in \Phi_1(\boldsymbol{x})$ and $\angle(-\boldsymbol{g}, \boldsymbol{u}_2(\boldsymbol{x})) \in \Phi_2(\boldsymbol{x})$.

- Suppose $\boldsymbol{x} \in \mathcal{X}$. Then, the point $\boldsymbol{x} = \left(-r + \frac{L}{\sigma_1}, \boldsymbol{0}\right)$ and $-r + \frac{L}{\sigma_1} \in (-r, r)$ as discussed below (2.12). In this case, we choose $\boldsymbol{g} = L\boldsymbol{e}_1'$ where $\boldsymbol{e}_1' = \frac{\boldsymbol{x}_2^* - \boldsymbol{x}_1^*}{\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\|}$. This implies that $\angle(\boldsymbol{g}, \boldsymbol{u}_1(\boldsymbol{x})) = 0 \in \Phi_1(\boldsymbol{x})$ and $\angle(-\boldsymbol{g}, \boldsymbol{u}_2(\boldsymbol{x})) = 0 \in \Phi_2(\boldsymbol{x})$ by the definition of Φ_i in (2.44).

Using Corollary 2.5.1, for both cases, we have that for $i \in \{1,2\}$, there exist functions $f_i \in \bigcup_{\hat{\sigma}>\sigma} \mathcal{Q}(\boldsymbol{x}_i^*, \hat{\sigma})$ such that $\boldsymbol{g} = \nabla f_1(\boldsymbol{x}) = -\nabla f_2(\boldsymbol{x})$ and $\|\nabla f_1(\boldsymbol{x})\| = \|\nabla f_2(\boldsymbol{x})\| = L$. Using (2.5), we have that there exist $f_i \in \mathcal{S}(\boldsymbol{x}_i^*, \sigma)$ with $\|\nabla f_i(\boldsymbol{x})\| \leq L$ for $i \in \{1,2\}$ and $\boldsymbol{x}$ is the minimizer of $f_1 + f_2$. Therefore, $\boldsymbol{x} \in \mathcal{M}$. Since $\mathcal{M}_\downarrow \subseteq (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we conclude that $\mathcal{M} \supseteq \mathcal{M}_\downarrow$. □



(a) (b)

Figure 2.4. (a) For given $\boldsymbol{x}$, $\boldsymbol{x}^*$ and σ, the figure illustrates the regions where the vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ with $\|\boldsymbol{g}_1\| = \|\boldsymbol{g}_2\| = L$ must lie, so that we can construct quadratic functions $f_i \in \bigcup_{\hat{\sigma}\geq\sigma} \mathcal{Q}(\boldsymbol{x}^*, \hat{\sigma})$ with $\nabla f_i(\boldsymbol{x}) = \boldsymbol{g}_i$ for $i \in \{1, 2\}$. Recall the definition of ϕ_i and Φ_i from (2.16) and (2.44), respectively. In particular, for $i \in \{1, 2\}$, it is sufficient to have $\phi_i(\boldsymbol{x}) \in \Phi_i(\boldsymbol{x})$ from Corollary 2.5.1, i.e., pictorially, the vectors $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ must **strictly** lie in the corresponding shaded regions. (b) The figure illustrates the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})$ in the description of $\mathcal{M}_\downarrow$ which means that there is an overlapping region (light green region in the figure) caused by one shaded region and the mirror of the other shaded region.

In two-dimensional space, for a given $\boldsymbol{x} \in (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, the geometrical interpretation of the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) > \psi(\boldsymbol{x})$, which is used to describe the inner approximation $\mathcal{M}_\downarrow$, is represented in Fig. 2.4b.

Before characterizing the set $\mathcal{M}_\downarrow$, recall the definition of λ_i and ν_i in (2.27) and (2.28), respectively. For $i \in \{1, 2\}$, we define

$$\mathcal{C}_i := \Big\{\boldsymbol{x} \in \mathbb{R}^n : x_1 = \lambda_i, \ \ \|\tilde{\mathbf{x}}\| = \nu_i\Big\}. \tag{2.45}$$

Comparing the definition of $\mathcal{M}^\uparrow$ in (2.13) to that of $\mathcal{M}_\downarrow$ in (2.14), we see that the description of $\mathcal{M}_\downarrow$ involves a strict inequality while it is not for $\mathcal{M}^\uparrow$. Since the only difference is the inequality sign, we could expect to see similar results as in Theorem 2.4.8. Specifically, the following theorem provides a characterization of $\partial\mathcal{M}_\downarrow$ and $(\mathcal{M}_\downarrow)^\circ$ explicitly, and also a property of $\mathcal{M}_\downarrow$. Since most parts of the proof are similar to that of Theorem 2.4.8, we defer the proof to Appendix A.2.2.

Theorem 2.5.2. *Assume $\sigma_1 \geq \sigma_2$. Let the sets $\mathcal{M}_\downarrow$, $\mathcal{T}$, $\widetilde{\mathcal{T}}$, and $\mathcal{B}_i$ for $i \in \{1, 2\}$ be defined as in (2.13), (2.22), (2.30), and (2.8), respectively. Also, let the sets $\mathcal{H}_i^+$ and $\mathcal{H}_i^-$ for $i \in \{1, 2\}$ be defined as in (2.29), and the sets $\mathcal{C}_i$ for $i \in \{1, 2\}$ be defined as in (2.45).*

(i) If $r \in \left(0, \ \frac{L}{2\sigma_1}\right]$ then $\mathcal{M}_\downarrow$ is open and

$$\partial\mathcal{M}_\downarrow = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \quad \text{and} \quad (\mathcal{M}_\downarrow)^\circ = \widetilde{\mathcal{T}}.$$

(ii) If $r \in \left(\frac{L}{2\sigma_1}, \ \frac{L}{2\sigma_2}\right]$ then $(\mathcal{M}_\downarrow \cup \mathcal{C}_1) \cap \mathcal{H}_1^+$ is closed while $\mathcal{M}_\downarrow \cap (\mathcal{H}_1^+)^c$ is open, and

$$\begin{aligned} \partial\mathcal{M}_\downarrow &= \Big[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\Big] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\} \quad \text{and} \\ (\mathcal{M}_\downarrow)^\circ &= \Big[\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\Big] \sqcup \Big[\widetilde{\mathcal{T}} \cap \mathcal{H}_1^-\Big]. \end{aligned}$$

(iii) *If* $r \in \left(\frac{L}{2\sigma_2},\ \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)\right)$ *then* $(\mathcal{M}_\downarrow \cup \mathcal{C}_1) \cap \mathcal{H}_1^+$ *and* $(\mathcal{M}_\downarrow \cup \mathcal{C}_2) \cap \mathcal{H}_2^-$ *are closed while* $\mathcal{M}_\downarrow \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c$ *is open, and*

$$\partial\mathcal{M}_\downarrow = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \mathcal{T} \quad \textit{and}$$
$$(\mathcal{M}_\downarrow)^\circ = \left[\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \left[\widetilde{\mathcal{T}} \cap (\mathcal{H}_1^- \cap \mathcal{H}_2^+)\right].$$

(iv) *If* $r = \frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right)$ *then* $\mathcal{M}_\downarrow = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1}-\frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\}$.

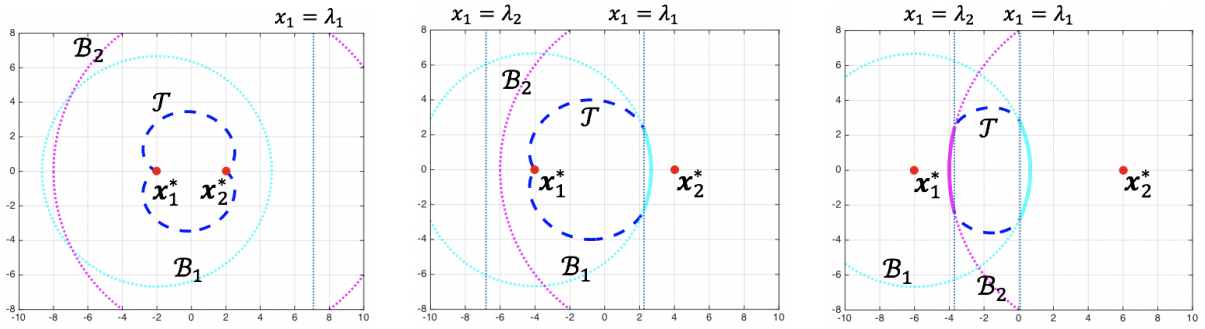
(v) *If* $r \in \left(\frac{L}{2}\left(\frac{1}{\sigma_1}+\frac{1}{\sigma_2}\right),\ \infty\right)$ *then* $\mathcal{M}_\downarrow = \emptyset$.

Examples of the boundary $\partial\mathcal{M}_\downarrow$ in $\mathbb{R}^2$ for the first three cases of Theorem 2.5.2 are shown in Fig. 2.5. Again, we consider parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$ and $L = 10$. For $r = 2$, as we can see from Fig. 2.5a, we have $\partial\mathcal{M}_\downarrow = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$ (i.e., dotted blue line + two red dots) consistent with part (i) of Theorem 2.5.2. For $r = 4$, as we can see from Fig. 2.5b, we have $\partial\mathcal{M}_\downarrow = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\}$ (i.e., solid cyan line + dotted blue line + left red dot) consistent with part (ii) of Theorem 2.5.2. For $r = 6$, as we can see from Fig. 2.5c, we have $\partial\mathcal{M}_\downarrow = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \mathcal{T}$ (i.e., solid cyan line + solid magenta line + dotted blue line) consistent with part (iii) of Theorem 2.5.2. Note that the dotted blue line in the figures indicates that the corresponding set of points is not a subset of the inner approximation $\mathcal{M}_\downarrow$, i.e., $\mathcal{T} \not\subseteq \mathcal{M}_\downarrow$ whereas the solid cyan line and solid magenta line indicate that $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c \subseteq \mathcal{M}_\downarrow$ and $\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c \subseteq \mathcal{M}_\downarrow$, respectively.

2.6 Potential Solution Region

In this section, using results from analyzing the outer approximation $\mathcal{M}^\uparrow$ in Section 2.4 and the inner approximation $\mathcal{M}_\downarrow$ in Section 2.5.2, we derive relationships among the potential solution region $\mathcal{M}$ (which is the set that we want to identify), outer approximation $\mathcal{M}^\uparrow$ and inner approximation $\mathcal{M}_\downarrow$.

Before stating the main theorem, we summarize the inclusions among the three sets. Specifically, based on Proposition 2.4.1 and Proposition 2.5.2, we get $\mathcal{M} \subseteq \mathcal{M}^\uparrow$ and $\mathcal{M}_\downarrow \subseteq \mathcal{M}$, respectively, and we can state the following proposition.



(a) For $r = 2$, the characterization of boundary $\partial\mathcal{M}_\downarrow$ corresponds to Theorem 2.5.2 part (i).

(b) For $r = 4$, the characterization of boundary $\partial\mathcal{M}_\downarrow$ corresponds to Theorem 2.5.2 part (ii).

(c) For $r = 6$, the characterization of boundary $\partial\mathcal{M}_\downarrow$ corresponds to Theorem 2.5.2 part (iii).

Figure 2.5. The boundary $\partial\mathcal{M}_\downarrow$ in $\mathbb{R}^2$ with different values of r for two original minimizers $x_1^* = (-r, 0)$ and $x_2^* = (r, 0)$ is plotted given fixed parameters $\sigma_1 = 1.5$, $\sigma_2 = 1$, and $L = 10$. The set $\mathcal{T}$ is represented by blue dashed lines since $\mathcal{T} \subseteq (\mathcal{M}_\downarrow)^c$, while the sets $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$ and $\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c$ are represented by cyan and magenta solid lines, respectively since they are both subsets of $\mathcal{M}_\downarrow$. The vertical dotted lines represent the equations $x_1 = \lambda_1$ and $x_1 = \lambda_2$ and note that the value of λ_1 and λ_2 depends on r.

Proposition 2.6.1. *Suppose the sets $\mathcal{M}$, $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$ are defined as in (2.7), (2.13) and (2.14), respectively. Then, $\mathcal{M}_{\downarrow} \subseteq \mathcal{M} \subseteq \mathcal{M}^{\uparrow}$.*

Since Theorem 2.4.8 and Theorem 2.5.2 are similar, we will see that, in fact, the boundary of the outer approximation $\partial\mathcal{M}^{\uparrow}$ and inner approximation $\partial\mathcal{M}_{\downarrow}$ are equal to the boundary of the potential solution region $\partial\mathcal{M}$ for all values of r. This means that we obtain the explicit characterization of $\partial\mathcal{M}$ from Theorem 2.4.8 or Theorem 2.5.2. We present this result in the following theorem.

Theorem 2.6.1. *Suppose $\mathcal{M}$, $\mathcal{M}^{\uparrow}$ and $\mathcal{M}_{\downarrow}$ are defined as in (2.7), (2.13) and (2.14), respectively. Then, $\partial\mathcal{M} = \partial\mathcal{M}^{\uparrow} = \partial\mathcal{M}_{\downarrow}$.*

Proof. Recall from Proposition 2.6.1 that $\mathcal{M}_{\downarrow} \subseteq \mathcal{M} \subseteq \mathcal{M}^{\uparrow}$. This entails that $(\mathcal{M}_{\downarrow})^{\circ} \subseteq (\mathcal{M})^{\circ} \subseteq (\mathcal{M}^{\uparrow})^{\circ}$ and $\overline{\mathcal{M}_{\downarrow}} \subseteq \overline{\mathcal{M}} \subseteq \overline{\mathcal{M}^{\uparrow}}$. On the other hand, we have $(\mathcal{M}_{\downarrow})^{\circ} = (\mathcal{M}^{\uparrow})^{\circ}$ and $\partial\mathcal{M}_{\downarrow} = \partial\mathcal{M}^{\uparrow}$ from Theorem 2.4.8 and Theorem 2.5.2. Combine the facts regarding the interiors to obtain that $(\mathcal{M}_{\downarrow})^{\circ} = (\mathcal{M})^{\circ} = (\mathcal{M}^{\uparrow})^{\circ}$. Then, we can write

$$\overline{\mathcal{M}_{\downarrow}} = (\mathcal{M}_{\downarrow})^{\circ} \sqcup \partial\mathcal{M}_{\downarrow} = (\mathcal{M}^{\uparrow})^{\circ} \sqcup \partial\mathcal{M}^{\uparrow} = \overline{\mathcal{M}^{\uparrow}}.$$

Combining the above equation with $\overline{\mathcal{M}_{\downarrow}} \subseteq \overline{\mathcal{M}} \subseteq \overline{\mathcal{M}^{\uparrow}}$, we can write $\overline{\mathcal{M}_{\downarrow}} = \overline{\mathcal{M}} = \overline{\mathcal{M}^{\uparrow}}$. Since $\partial\mathcal{E} = \overline{\mathcal{E}} \backslash \mathcal{E}^{\circ}$ for any subset $\mathcal{E}$ in a topological space, we conclude that $\partial\mathcal{M} = \partial\mathcal{M}^{\uparrow} = \partial\mathcal{M}_{\downarrow}$. $\square$

Recall the definition of $\mathcal{M}$, $\mathcal{T}$, $\{\partial\mathcal{B}_i \text{ for } i \in \{1,2\}\}$ and $\{\mathcal{H}_1^-, \mathcal{H}_2^+\}$ from (2.7), (2.22), (2.25), and (2.29), respectively. Assuming that $\sigma_1 \geq \sigma_2$, we summarize a characterization of the potential solution region $\mathcal{M}$ as follows:

$$\begin{cases} \partial\mathcal{M} = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} & \text{if} \quad r \in \left(0,\ \frac{L}{2\sigma_1}\right], \\ \partial\mathcal{M} = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \mathcal{T} \sqcup \{\boldsymbol{x}_1^*\} & \text{if} \quad r \in \left(\frac{L}{2\sigma_1},\ \frac{L}{2\sigma_2}\right], \\ \partial\mathcal{M} = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \left[\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c\right] \sqcup \mathcal{T} & \text{if} \quad r \in \left(\frac{L}{2\sigma_2},\ \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)\right), \\ \mathcal{M} = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\} & \text{if} \quad r \in \left(\frac{L}{2\sigma_2},\ \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)\right), \\ \mathcal{M} = \emptyset & \text{if} \quad r \in \left(\frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right),\ \infty\right), \end{cases}$$

where the first three equations are obtained by applying Theorem 2.4.8 and Theorem 2.5.2 to Theorem 2.6.1, and the last two equations are obtained by applying Theorem 2.4.8 and Theorem 2.5.2 to Proposition 2.6.1. Examples of the boundary $\partial\mathcal{M}$ in $\mathbb{R}^2$ with different values of r are shown in Fig. 2.1. In particular, the solid blue, cyan and magenta curves in the figure correspond to the sets $\mathcal{T}$, $\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c$ and $\partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c$, respectively.

2.7 Discussion and Conclusions

In this chapter, we studied the possible locations of the minimizer of the sum of two strongly convex functions. Based on the location of the two minimizers of the individual functions, strong convexity parameters and a bound on the gradients at the minimizer of the sum, we established a necessary condition and a sufficient condition for a given point to be a minimizer, and called the set of points that satisfies the conditions as the outer approximation $\mathcal{M}^{\uparrow}$ and inner approximation $\mathcal{M}_{\downarrow}$, respectively. We then explicitly characterized the boundary and interior of the outer and inner approximations. The characterization of these boundaries and interiors turned out to be identical. Subsequently, we showed that the boundary of the potential solution region $\partial\mathcal{M}$ is also identical to those boundaries. In particular, we showed that it is sufficient to consider quadratic functions to establish (almost) the entire set of potential minimizers. To visualize the boundary of the potential solution region $\partial\mathcal{M}$, we provided examples with different distances between the two original minimizers in Fig. 2.1.

Our work in this chapter focused on the case of two functions. Future work could include identifying the region that the minimizer of the sum can lie for the case of multiple strongly convex functions. One can also modify some assumptions, for example by considering strongly convex functions with Lipschitz continuous gradient condition.

3. ON THE SET OF POSSIBLE MINIMIZERS OF A SUM OF KNOWN AND UNKNOWN FUNCTIONS

3.1 Introduction

As discussed earlier in this thesis, optimization is an important tool in various fields, including machine learning [1], signal processing [2], control theory, [3]–[5], and robotics [6]–[8]. Given an objective function to be optimized, there are several standard algorithms that can be applied to find the optimal variables [9]–[12].

However, in many applications, it may be the case that the objective function is only partially known. For example, such scenarios are central to the field of robust optimization, where the objective function contains some parametric uncertainty, and the goal is to choose the optimization variable to be robust to the possible realizations of the uncertainty [35]–[37]. The problem that we consider in this chapter also has a similar flavor, in that we assume that the optimization objective is not fully known. However, rather than seeking to find a single solution that is simultaneously robust to all possible realizations of the uncertain parameter (or learning that parameter [37]), we instead seek to characterize the region where the minimizer could lie for *each* possible realization of the uncertainty. This approach has the potential to yield insights regarding the nature of the possible solutions to the given uncertain optimization problem.

In the previous chapter (Chapter 2), we determined a region containing the possible minimizers of a sum of two arbitrary strongly convex functions, given only the minimizers of the local functions, their strong convexity parameters, and a bound on their gradients. In contrast, in this chapter, we shall consider the case of optimizing a sum of known and unknown functions where only limited information about the unknown function is available.

In this case, we are given some general characteristics of the unknown function, namely a region containing the minimizer, and the strong convexity parameter of the function. Our goal is to determine necessary conditions for a point to be a minimizer of the sum. In particular, we will determine a region where the potential minimizer of the sum can lie. Thus, if a point from within this region is chosen as an estimate of the true minimizer of the sum, the size of the region can be used to quantify how far the estimate can be from the true minimizer. Below, we describe an example scenario to illustrate this problem.

An Example Scenario

In supervised machine learning problems, one uses labeled training data in order to construct a model that can be used to perform regression or classification tasks. The training data consists of pairs $\boldsymbol{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ which are the feature vector and label of the i-th example, respectively. For simplicity, assume that we have 2 training sets denoted by $\mathcal{D}_j = \{\boldsymbol{x}_i^{(j)}, y_i^{(j)}\}_{i=1}^{N_j}$ for $j \in \{1, 2\}$. We can write the aggregate loss function of the whole dataset $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ as

$$L(\boldsymbol{w}; \mathcal{D}) = \underbrace{\sum_{i=1}^{N_1} \ell(\boldsymbol{w};\ \boldsymbol{x}_i^{(1)}, y_i^{(1)})}_{L_1(\boldsymbol{w};\, \mathcal{D}_1)} + \underbrace{\sum_{i=1}^{N_2} \ell(\boldsymbol{w};\ \boldsymbol{x}_i^{(2)}, y_i^{(2)})}_{L_2(\boldsymbol{w};\, \mathcal{D}_2)},$$

where $\boldsymbol{w}$ is a model parameter that we need to optimize and $\ell(\boldsymbol{w};\ \boldsymbol{x}_i^{(j)}, y_i^{(j)})$ is a loss function for each sample. Assume that $L(\boldsymbol{w}; \mathcal{D})$ is a strongly convex function (which will be the case when we consider linear regression problems or functions incorporating l_2 regularization [58]). Suppose $\boldsymbol{w}^*$ and $\boldsymbol{w}_2^*$ are the minimizer of $L(\boldsymbol{w}; \mathcal{D})$ and $L_2(\boldsymbol{w}; \mathcal{D}_2)$, respectively.

Now suppose that the entity trying to find the optimal parameter $\boldsymbol{w}$ for $L(\boldsymbol{w}, \mathcal{D})$ can only access the data set $\mathcal{D}_1$, but not $\mathcal{D}_2$ (or alternatively, can only access a corrupted or *poisoned* version of $\mathcal{D}_2$ [59], [60]). In this case, the entity may only know certain properties of the function $L_2(\boldsymbol{w}; \mathcal{D}_2)$ (such as its general form, convexity parameters, etc.), and a region containing the minimizer of $L_2(\boldsymbol{w}; \mathcal{D}_2)$ (e.g., based on the statistical properties of the underlying data). Given this limited information about $L_2(\boldsymbol{w}; \mathcal{D}_2)$, and with $L_1(\boldsymbol{w}; \mathcal{D}_1)$ fully

known, the entity could seek to find a region that is guaranteed to contain the minimizer of the true function $L(\boldsymbol{w}; \mathcal{D})$. This is the problem tackled in this chapter.

3.2 Notation and Preliminaries

3.2.1 Sets

We denote the closure, interior, and boundary of a set $\mathcal{E}$ by $\bar{\mathcal{E}}$, $\mathcal{E}^\circ$, and $\partial\mathcal{E} = \bar{\mathcal{E}} \setminus \mathcal{E}^\circ$, respectively.

3.2.2 Linear Algebra

We denote by $\mathbb{R}^n$ the n-dimensional Euclidean space. For simplicity, we often use $\boldsymbol{x} = (x_1, \ldots, x_n)$ to represent the column vector $\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^T$. We use $\boldsymbol{e}_i$ to denote the i-th basis vector (the vector of all zeros except for a one in the i-th position). We denote by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ the Euclidean inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$ i.e., $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^T \boldsymbol{v}$, by $\|\cdot\|$ the Euclidean norm $\|\boldsymbol{x}\| := (\sum_i x_i^2)^{1/2}$ and by $\angle(\boldsymbol{u}, \boldsymbol{v})$ the angle between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Note that

$$\angle(\boldsymbol{u}, \boldsymbol{v}) = \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right).$$

We use $\mathcal{B}(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_0\| < r\}$ and $\bar{\mathcal{B}}(\boldsymbol{x}_0, r)$ to denote the open and closed ball, respectively, centered at $\boldsymbol{x}_0$ of radius r. Moreover, the function $\boldsymbol{u}(\boldsymbol{x}_1, \boldsymbol{x}_2) : (\mathbb{R}^n \times \mathbb{R}^n) \setminus \{(\boldsymbol{z}_1, \boldsymbol{z}_2) \in \mathbb{R}^n \times \mathbb{R}^n : \boldsymbol{z}_1 = \boldsymbol{z}_2\} \to \mathbb{R}^n$ denotes the unit vector in the direction of $\boldsymbol{x}_1 - \boldsymbol{x}_2$, i.e.,

$$\boldsymbol{u}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\boldsymbol{x}_1 - \boldsymbol{x}_2}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|} \quad \text{with} \quad \boldsymbol{x}_1 \neq \boldsymbol{x}_2. \tag{3.1}$$

3.2.3 Convex Sets and Convex Functions

A set $\mathcal{C}$ in $\mathbb{R}^n$ is said to be convex if, for all $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathcal{C}$ and all θ in the interval $(0, 1)$, the point $(1 - \theta)\boldsymbol{x}_1 + \theta\boldsymbol{x}_2 \in \mathcal{C}$.

We say a vector $\boldsymbol{g} \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x} \in \mathbf{dom}\, f$ if for all $\boldsymbol{z} \in \mathbf{dom}\, f$, $f(\boldsymbol{z}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{z} - \boldsymbol{x} \rangle$.

If f is convex and differentiable, then its gradient at $\boldsymbol{x}$ is a subgradient; however, a subgradient can exist even when f is not differentiable at $\boldsymbol{x}$. A function f is called subdifferentiable at $\boldsymbol{x}$ if there exists at least one subgradient at $\boldsymbol{x}$. The set of subgradients of f at the point $\boldsymbol{x}$ is called the subdifferential of f at $\boldsymbol{x}$, and is denoted $\partial f(\boldsymbol{x})$. The subdifferential $\partial f(\boldsymbol{x})$ is always a closed convex set, even if f is not convex. In addition, if f is continuous at $\boldsymbol{x}$, then the subdifferential $\partial f(\boldsymbol{x})$ is bounded.

A function f is called strongly convex with parameter $\sigma > 0$ (or σ-strongly convex) if for all points $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$, $\langle \boldsymbol{g}_{\boldsymbol{x}} - \boldsymbol{g}_{\boldsymbol{y}}, \boldsymbol{x} - \boldsymbol{y} \rangle \geq \sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2$ for all $\boldsymbol{g}_{\boldsymbol{x}} \in \partial f(\boldsymbol{x})$ and $\boldsymbol{g}_{\boldsymbol{y}} \in \partial f(\boldsymbol{y})$. We denote the set of all convex functions by $\mathcal{F}$, and the set of all σ-strongly convex functions with minimizer $\boldsymbol{x}_u^*$ in the set $\mathcal{A} \subseteq \mathbb{R}^n$ and $\mathbf{dom}(\cdot) = \mathbb{R}^n$ by $\mathcal{S}(\mathcal{A}, \sigma)$.

3.3 Problem Statement

We consider a function of the form

$$f(\boldsymbol{x}) = f^k(\boldsymbol{x}) + f^u(\boldsymbol{x}), \tag{3.2}$$

where f^k and f^u are convex functions. We assume that we know f^k exactly, but do not know f^u, other than some general properties described below.

We assume that $f^k \in \mathcal{F}$ and $f^u \in \mathcal{S}(\mathcal{A}, \sigma)$ where $\mathcal{A}$ is a compact set (i.e., we only know that f^u is σ-strongly convex and that its minimizer lies in some set $\mathcal{A}$). Our goal is to find the set of points $\boldsymbol{x} \in \mathbb{R}^n$ that could potentially be the minimizer of $f(\boldsymbol{x})$ in (3.2). To this end, we will seek to characterize the region

$$\mathcal{M}(f^k, \mathcal{A}, \sigma) := \Big\{ \boldsymbol{x} \in \mathbb{R}^n : \ \exists f^u \in \mathcal{S}(\mathcal{A}, \sigma),\ \mathbf{0} \in \partial f^k(\boldsymbol{x}) + \partial f^u(\boldsymbol{x}) \Big\}. \tag{3.3}$$

For simplicity of notation, we will omit the argument of the set $\mathcal{M}(f^k, \mathcal{A}, \sigma)$ and write it as $\mathcal{M}$. Note that $\mathcal{M}$ contains all points $\boldsymbol{x} \in \mathbb{R}^n$ that can potentially be a minimizer of f, given f^k, and the quantity σ and the set $\mathcal{A}$ pertaining to f^u.

Remark 3. Returning to the regression scenario involving data that is not directly available to the optimizing entity (described in the Introduction), the unknown function would be of the form $f^u(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{y}\|^2$ where A is a matrix containing (unknown) training data and $\boldsymbol{y}$ is the (unknown) vector of corresponding labels. When A has full rank, the loss function is strongly convex. In addition, if some general underlying statistical properties of the data are known to the optimizing entity, it could estimate a lower bound on the strong convexity parameter σ, and a region containing the possible minimizer of $f^u(\boldsymbol{x})$. Thus, using this information, the central entity seeks to find the set of possible minimizers of the sum of this unknown function and its own loss function (corresponding to data that it has access to directly).

3.4 Analysis for General Uncertainty Region

In this section, we provide a necessary condition for a point $\boldsymbol{x}^* \in \mathbf{dom}\, f^k$ to be the minimizer of f in the general case where the uncertainty region $\mathcal{A}$ of the minimizer of the unknown function is compact, but of arbitrary shape.

For any given point $\boldsymbol{x}^* \in \mathbb{R}^n \setminus \mathcal{A}$, define the set

$$\tilde{\mathcal{A}}(\mathcal{A}, \boldsymbol{x}^*) := \Big\{\boldsymbol{x} \in \partial\mathcal{A} : (1-\theta)\boldsymbol{x} + \theta\boldsymbol{x}^* \notin \mathcal{A}, \ \ \forall \theta \in (0,1)\Big\}. \tag{3.4}$$

In words, $\tilde{\mathcal{A}}(\mathcal{A}, \boldsymbol{x}^*)$ is the set of points $\boldsymbol{x}$ on the boundary of $\mathcal{A}$ such that the line joining $\boldsymbol{x}$ to $\boldsymbol{x}^*$ does not intersect $\mathcal{A}$ (except at $\boldsymbol{x}$).

Theorem 3.4.1. *Suppose $f^k \in \mathcal{F}$ and $\mathcal{A} \subseteq \mathbf{dom}\, f^k$ is a compact set. A necessary condition for a point $\boldsymbol{x}^* \in \mathbb{R}^n$ to be in $\mathcal{M}(f^k, \mathcal{A}, \sigma) \setminus \mathcal{A}$ is*

$$\min_{\boldsymbol{x}_u^* \in \mathcal{A},\, \boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)} \frac{\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma. \tag{3.5}$$

Furthermore, the above inequality (3.5) can be reduced to

$$\min_{(\boldsymbol{x}_u^*, \boldsymbol{g}^k) \in \mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*)} \frac{\langle \boldsymbol{g}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma, \tag{3.6}$$

where

$$\mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*) := \Big\{(\boldsymbol{x}, \boldsymbol{g}) \in \tilde{\mathcal{A}}(\mathcal{A}, \boldsymbol{x}^*) \times \partial f^k(\boldsymbol{x}^*) : \ \langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0\Big\}. \tag{3.7}$$

Proof. Suppose $f^u \in \mathcal{S}(\mathcal{A}, \sigma)$. For any $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$, let $\boldsymbol{g}^u_{\boldsymbol{x}} \in \partial f^u(\boldsymbol{x})$ and $\boldsymbol{g}^u_{\boldsymbol{y}} \in \partial f^u(\boldsymbol{y})$. From the definition of a strongly convex function, we have

$$\langle \boldsymbol{g}^u_{\boldsymbol{x}} - \boldsymbol{g}^u_{\boldsymbol{y}}, \boldsymbol{x} - \boldsymbol{y}\rangle \geq \sigma \|\boldsymbol{x} - \boldsymbol{y}\|^2$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Let $\boldsymbol{x}^*_u \in \mathcal{A}$ be the true minimizer of f^u and suppose $\boldsymbol{x}^*$ is the minimizer of $f = f^k + f^u$. Then, substitute $\boldsymbol{x}^*$ into $\boldsymbol{x}$ and $\boldsymbol{x}^*_u$ into $\boldsymbol{y}$ to get

$$\langle \boldsymbol{g}^u_{\boldsymbol{x}^*} - \boldsymbol{g}^u_{\boldsymbol{x}^*_u}, \boldsymbol{x}^* - \boldsymbol{x}^*_u\rangle \geq \sigma \|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|^2$$

for all $\boldsymbol{g}^u_{\boldsymbol{x}^*} \in \partial f^u(\boldsymbol{x}^*)$ and $\boldsymbol{g}^u_{\boldsymbol{x}^*_u} \in \partial f^u(\boldsymbol{x}^*_u)$. Since $\boldsymbol{x}^*_u$ is the minimizer of f^u, we have $\boldsymbol{0} \in \partial f^u(\boldsymbol{x}^*_u)$. Consider $\boldsymbol{x}^* \notin \mathcal{A}$, which implies $\boldsymbol{x}^* \neq \boldsymbol{x}^*_u$, and rewrite the inequality above (with $\boldsymbol{g}^u_{\boldsymbol{x}^*_u} = \boldsymbol{0}$) to get

$$\left\langle \boldsymbol{g}^u_{\boldsymbol{x}^*}, \frac{\boldsymbol{x}^* - \boldsymbol{x}^*_u}{\|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|} \right\rangle \geq \sigma \|\boldsymbol{x}^* - \boldsymbol{x}^*_u\| > 0.$$

Recall the definition of $\boldsymbol{u}(\cdot, \cdot)$ in (3.1). The inequality above becomes

$$\langle \boldsymbol{g}^u_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle \geq \sigma \|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|. \tag{3.8}$$

Using the fact that $\boldsymbol{x}^*$ is the minimizer of $f = f^k + f^u$, we get $\boldsymbol{0} \in \partial f^k(\boldsymbol{x}^*) + \partial f^u(\boldsymbol{x}^*)$, so there exists $\boldsymbol{g}^k_{\boldsymbol{x}^*} \in \partial f^k(\boldsymbol{x}^*)$ and $\boldsymbol{g}^u_{\boldsymbol{x}^*} \in \partial f^u(\boldsymbol{x}^*)$ such that $\boldsymbol{g}^k_{\boldsymbol{x}^*} + \boldsymbol{g}^u_{\boldsymbol{x}^*} = \boldsymbol{0}$. Since the inequality (3.8) is true for any $\boldsymbol{g}^u_{\boldsymbol{x}^*} \in \partial f^u(\boldsymbol{x}^*)$, we can apply $\boldsymbol{g}^u_{\boldsymbol{x}^*} = -\boldsymbol{g}^k_{\boldsymbol{x}^*}$ to (3.8) and get

$$\langle -\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle \geq \sigma \|\boldsymbol{x}^* - \boldsymbol{x}^*_u\| \quad \Leftrightarrow \quad \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|} \leq -\sigma.$$

Thus, if $f^u \in \mathcal{S}(\mathcal{A}, \sigma)$, we have a necessary condition that

$$\textit{if} \quad \boldsymbol{x}^* \in \mathcal{M} \setminus \mathcal{A} \quad \textit{then there exist} \quad \boldsymbol{x}_u^* \in \mathcal{A} \quad \textit{and} \quad \boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)$$
$$\textit{such that} \quad \frac{\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma.$$

Since the sets $\mathcal{A}$ and $\partial f^k(\boldsymbol{x}^*)$ are compact by the assumption that f^k is convex, the necessary condition above is equivalent to

$$\min_{\boldsymbol{x}_u^* \in \mathcal{A},\, \boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)} \frac{\langle g_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma. \tag{3.9}$$

Next, we will show that we can consider the minimum over the set $\mathcal{X}$ (defined in (3.7)) instead of $\mathcal{A} \times \partial f^k(\boldsymbol{x}^*)$. Define the set

$$\mathcal{D}(f^k, \boldsymbol{x}^*) := \{(\boldsymbol{x}, \boldsymbol{g}_{\boldsymbol{x}^*}^k) \in \mathbb{R}^n \times \partial f^k(\boldsymbol{x}^*) : \ \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0\}.$$

First, using the fact that σ and $\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|$ are positive, we have

$$\frac{\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma \quad \Rightarrow \quad \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle < 0.$$

This means that we can consider the pair $(\boldsymbol{x}_u^*, \boldsymbol{g}_{\boldsymbol{x}^*}^k)$ inside the set $(\mathcal{A} \times \mathbb{R}^n) \cap \mathcal{D}$ instead of $\mathcal{A} \times \partial f^k(\boldsymbol{x}^*)$. Next, let

$$\mathcal{E}(\mathcal{A}, \boldsymbol{x}^*) := \{\boldsymbol{x} \in \mathcal{A} : \exists \theta \in (0,1), \ \ (1-\theta)\boldsymbol{x} + \theta \boldsymbol{x}^* \in \mathcal{A}\}.$$

Suppose $(\boldsymbol{x}_u^{(1)}, \boldsymbol{g}_{\boldsymbol{x}^*}^k) \in \mathcal{D} \cap (\mathcal{E} \times \mathbb{R}^n)$. We choose $\bar{\theta}$ so that $\bar{\theta} \in (0,1)$ and $\boldsymbol{x}_u^{(2)} = (1-\bar{\theta})\boldsymbol{x}_u^{(1)} + \bar{\theta}\boldsymbol{x}^* \in \mathcal{A}$, i.e., $\boldsymbol{x}_u^{(2)}$ is in between $\boldsymbol{x}_u^{(1)}$ and $\boldsymbol{x}^*$, and also in the set $\mathcal{A}$. We have

$$\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^{(1)})\rangle = \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^{(2)})\rangle < 0$$

and so

$$\frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^{(2)}_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^{(2)}_u\|} < \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^{(1)}_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^{(1)}_u\|},$$

i.e., if $\boldsymbol{x}^{(1)}_u$ satisfies (3.10), then so does $\boldsymbol{x}^{(2)}_u$. This means that we can consider the pair $(\boldsymbol{x}^*_u, \boldsymbol{g}^k_{\boldsymbol{x}^*})$ inside the set $\Big((\mathcal{A} \setminus \mathcal{E}) \times \mathbb{R}^n\Big) \cap \mathcal{D}$ instead of $\mathcal{A} \times \partial f^k(\boldsymbol{x}^*)$. However, we will show that in fact the set

$$\mathcal{A} \setminus \mathcal{E} = \Big\{\boldsymbol{x} \in \mathcal{A} : \ (1-\theta)\boldsymbol{x} + \theta \boldsymbol{x}^* \notin \mathcal{A}, \quad \forall \theta \in (0,1)\Big\}$$

is contained in $\partial\mathcal{A}$, i.e., $\mathcal{A} \setminus \mathcal{E} \subseteq \partial\mathcal{A}$. Suppose $\boldsymbol{x} \in \mathcal{A}^\circ$ so there exists $\epsilon > 0$ such that $\mathcal{B}(\boldsymbol{x}, \epsilon) \subseteq \mathcal{A}$. By choosing $\hat{\theta} = \frac{\epsilon}{2\|\boldsymbol{x}^* - \boldsymbol{x}\|}$, we get $(1-\hat{\theta})\boldsymbol{x} + \hat{\theta}\boldsymbol{x}^* \in \mathcal{A}$ and $\hat{\theta} \in (0,1)$ since $\boldsymbol{x}^* \notin \mathcal{A}$. This implies that $\boldsymbol{x} \notin \mathcal{A} \setminus \mathcal{E}$ and therefore $\mathcal{A} \setminus \mathcal{E} \subseteq \partial\mathcal{A}$. Using the definition of $\tilde{\mathcal{A}}$ in (3.4), we can then rewrite the set $\mathcal{A} \setminus \mathcal{E}$ as follows:

$$\mathcal{A} \setminus \mathcal{E}(\mathcal{A}, \boldsymbol{x}^*) = \tilde{\mathcal{A}}(\mathcal{A}, \boldsymbol{x}^*).$$

From the definition of $\mathcal{X}$ in (3.7), we have

$$\Big(\tilde{\mathcal{A}}(\mathcal{A}, \boldsymbol{x}^*) \times \mathbb{R}^n\Big) \cap \mathcal{D}(f^k, \boldsymbol{x}^*) = \mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*).$$

Thus, the necessary condition (3.9) reduces to

$$\min_{(\boldsymbol{x}^*_u, \boldsymbol{g}^k_{\boldsymbol{x}^*}) \in \mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*)} \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|} \leq -\sigma.$$

□

We can interpret the necessary condition in Theorem 3.4.1 as follows. To check whether $\boldsymbol{x}^* \in \mathbb{R}^n$ can be a minimizer of $f(\boldsymbol{x})$, we can follow the inequality (3.5) and search for a pair $(\boldsymbol{x}_u^*, \boldsymbol{g}_{\boldsymbol{x}^*}^k)$ with $\boldsymbol{x}_u^* \in \mathcal{A}$ and $\boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)$ such that the pair satisfies the inequality

$$\frac{\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma. \tag{3.10}$$

However, the inequality (3.6) with $\mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*)$ defined in (3.7) suggests that we do not have to search throughout the space $\mathcal{A} \times \partial f^k(\boldsymbol{x}^*)$. Instead, we can restrict our attention to be in the set $\mathcal{X}$. Now we have the variables $\boldsymbol{x}_u^*$ and $\boldsymbol{g}_{\boldsymbol{x}^*}^k$ that are coupled through the inequality $\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle < 0$. That is, if we first choose $\boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)$, then we can consider $\boldsymbol{x}_u^*$ that is in the set $\{\boldsymbol{x} \in \partial\mathcal{A} : \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0, \quad (1-\theta)\boldsymbol{x} + \theta\boldsymbol{x}^* \notin \mathcal{A}, \quad \forall\theta \in (0,1)\}$. Similarly, if we first choose $\boldsymbol{x}_u^* \in \{\boldsymbol{x} \in \partial\mathcal{A} : \ (1-\theta)\boldsymbol{x} + \theta\boldsymbol{x}^* \notin \mathcal{A}, \quad \forall\theta \in (0,1)\}$, then we can consider $\boldsymbol{g}_{\boldsymbol{x}^*}^k$ that is in the set $\{\boldsymbol{g} \in \partial f^k(\boldsymbol{x}^*) : \langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0\}$.

If the function f^k is differentiable at $\boldsymbol{x}^*$, we have a single element in the set $\partial f^k(\boldsymbol{x}^*)$, namely $\nabla f^k(\boldsymbol{x}^*)$, and we can search for $\boldsymbol{x}_u^* \in \partial\mathcal{A}$ such that $\langle \nabla f^k(\boldsymbol{x}^*), \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle < 0$. However, if the set $\mathcal{A}$ is arbitrary, this search may be computationally expensive. In the next section, we consider additional structure on the set $\mathcal{A}$ to simplify the search.

Remark 4. Note that the set $\mathcal{A}^\circ \subseteq \mathcal{M}(f^k, \mathcal{A}, \sigma)$. To see this, note that for all $\boldsymbol{x}^* \in \mathcal{A}^\circ$, there exists $\epsilon > 0$ such that $\mathcal{B}(\boldsymbol{x}^*, \epsilon) \subset \mathcal{A}^\circ$. Suppose $\boldsymbol{g} \in \partial f^k(\boldsymbol{x}^*)$. We can choose $f^u(\boldsymbol{x}) = \frac{\sigma_u}{2}\left\|\boldsymbol{x} - \left(\boldsymbol{x}^* + \frac{\boldsymbol{g}}{\sigma_u}\right)\right\|^2$ where $\sigma_u = \frac{2k\|\boldsymbol{g}\|}{\epsilon}$ and $k = \max\left\{1, \frac{\sigma\epsilon}{2\|\boldsymbol{g}\|}\right\}$. One can verify that $\boldsymbol{x}_u^* \in \mathcal{B}(\boldsymbol{x}^*, \epsilon)$, $\nabla f^u(\boldsymbol{x}^*) = -\boldsymbol{g}$, and $\sigma_u \geq \sigma$.

3.5 Analysis for the Case where Uncertainty Region is a Ball

Here, we consider additional structure on the uncertainty set $\mathcal{A}$ in order to provide a more specific characterization of the region $\mathcal{M}$. In particular, we consider $\mathcal{A} = \bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$, where $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ is the best guess of what the true parameter $\boldsymbol{x}_u^*$ is, and ϵ_0 is the maximum possible deviation of the true minimizer from our best guess.

We begin by investigating a property of the necessary condition (3.6) under a coordinate transformation. Suppose $\boldsymbol{x} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)}) \in \mathbb{R}^n$ and $\boldsymbol{x}^* \notin \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$. Let $\mathbf{T}$ and $\mathbf{R}$ be

the translation and rotation operators such that $\mathbf{R}(\mathbf{T}(\bar{\boldsymbol{x}})) = \mathbf{0}$, $\mathbf{R}(\mathbf{T}(\boldsymbol{x}^*)) = (\tilde{x}^*_{(1)}, 0, \ldots, 0)$ with $\tilde{x}^*_{(1)} > 0$, and $\mathbf{R}(\boldsymbol{g}^k) = (\tilde{g}_{(1)}, \tilde{g}_{(2)}, 0, \ldots, 0)$ with $\tilde{g}_{(2)} \geq 0$ while preserving the distance between any two points. In other words, given the ball $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$, a point $\boldsymbol{x}^*$ and a vector $\boldsymbol{g}^k$, we transform the coordinates so that the ball is centered at the origin, the point $\boldsymbol{x}^*$ lies on the $x_{(1)}$-axis, and the vector $\boldsymbol{g}^k$ lies on the $x_{(1)}$-$x_{(2)}$ plane.

Next, consider the expression $\frac{\langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|}$. Notice that both numerator and denominator can be written as inner products. Since $\mathbf{R}$ is a unitary operator, we have

$$\frac{\Big\langle \mathbf{R}(\boldsymbol{g}), \boldsymbol{u}\Big(\mathbf{R}(\mathbf{T}(\boldsymbol{x}^*)), \mathbf{R}(\mathbf{T}(\boldsymbol{x}^*_u))\Big)\Big\rangle}{\|\mathbf{R}(\mathbf{T}(\boldsymbol{x}^*)) - \mathbf{R}(\mathbf{T}(\boldsymbol{x}^*_u))\|} = \frac{\langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}^*_u\|}.$$

This means that even though we use the coordinate transformation $\mathbf{R}(\mathbf{T}(\cdot))$, we can still apply Theorem 3.4.1. Therefore, for the purpose of deriving our main result, without loss of generality, we can consider $\bar{\boldsymbol{x}} = \mathbf{0}$, $\boldsymbol{x}^* = (x^*_{(1)}, 0, \ldots, 0)$ where $x^*_{(1)} > \epsilon_0$, and $\boldsymbol{g}$ $(= \boldsymbol{g}^k) = (g_{(1)}, g_{(2)}, 0, \ldots, 0)$, where $g_{(2)} \geq 0$.

Before going into the result, we introduce some definitions that will appear in the theorem. For any given $\boldsymbol{x}^* \in \mathbb{R}^n$, define $\boldsymbol{z}_1(\boldsymbol{x}^*) \in \mathbb{R}^n$ as

$$\boldsymbol{z}_1(\boldsymbol{x}^*) := \underset{\boldsymbol{x} \in \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)}{\arg\min} \|\boldsymbol{x} - \boldsymbol{x}^*\|. \tag{3.11}$$

By our assumption that $\boldsymbol{x}^* = (x^*_{(1)}, 0, \ldots, 0)$, we have $\boldsymbol{z}_1(\boldsymbol{x}^*) = (\epsilon_0, 0, \ldots, 0)$. Since $\boldsymbol{x}^* \notin \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$, the point $\boldsymbol{z}_1$ is unique and is on $\partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$. If $\boldsymbol{g} \neq \alpha(\boldsymbol{x}^* - \bar{\boldsymbol{x}}) = (\alpha x^*_{(1)}, 0, \ldots, 0)$ for all $\alpha \geq 0$ (i.e., $\angle(\boldsymbol{g}, \boldsymbol{x}^* - \bar{\boldsymbol{x}}) \neq 0$), we define the set $\mathcal{P}$ to be such that

$$\mathcal{P}(\boldsymbol{g}, \boldsymbol{x}^*) := \underset{\boldsymbol{x} \in \partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)}{\arg\min} \angle(\boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}^*),$$

the point $\boldsymbol{z}_2 \in \mathbb{R}^n$ to be such that

$$\boldsymbol{z}_2(\boldsymbol{g}, \boldsymbol{x}^*) := \underset{\boldsymbol{x} \in \mathcal{P}(\boldsymbol{g}, \boldsymbol{x}^*)}{\arg\min} \|\boldsymbol{x} - \boldsymbol{x}^*\|, \tag{3.12}$$

and the curve $\mathcal{C}_0(\bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{g}, \boldsymbol{x}^*)$ to be the shortest path on the surface $\partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ that connects $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ together, i.e., $\mathcal{C}_0$ is the geodesic path between $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ on $\partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$.

To clarify these definitions, we introduce two more objects. Let L be the ray that starts from the point $\boldsymbol{x}^*$ and runs parallel to the vector $\boldsymbol{g}$ i.e.,

$$L(\boldsymbol{g}, \boldsymbol{x}^*) = \{\boldsymbol{x} \in \mathbb{R}^n : \exists\, t \in [0, \infty),\ \boldsymbol{x} = \boldsymbol{x}^* + t\boldsymbol{g}\}.$$

If $\boldsymbol{g} \neq \alpha(\boldsymbol{x}^* - \bar{\boldsymbol{x}}) = (\alpha x^*_{(1)}, 0, \ldots, 0)$ for all $\alpha \in \mathbb{R}$, let P_2 be the 2-dimensional plane that contains the vectors $\boldsymbol{g}$ and $\boldsymbol{x}^* - \bar{\boldsymbol{x}}$ as its bases, and contains the point $\boldsymbol{x}^*$, i.e.,

$$\begin{aligned} P_2(\bar{\boldsymbol{x}}, \boldsymbol{g}, \boldsymbol{x}^*) &:= \{\boldsymbol{x} \in \mathbb{R}^n : \exists\, s, t \in \mathbb{R} \text{ such that } \boldsymbol{x} = \boldsymbol{x}^* + s\boldsymbol{g} + t(\boldsymbol{x}^* - \bar{\boldsymbol{x}})\} \\ &= \{\boldsymbol{x} \in \mathbb{R}^n : x_{(3)} = x_{(4)} = \ldots = x_{(n)} = 0\}, \end{aligned}$$

where the second equality follows from the fact that

$$\boldsymbol{x}^* = (x^*_{(1)}, 0, 0, \ldots, 0) \quad \text{and} \quad \boldsymbol{g} = (g_{(1)}, g_{(2)}, 0, \ldots, 0).$$

There are two possible cases:

(i) the ray L passes through the ball $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ and

(ii) the ray L does not pass through the ball $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$.

In the first case, we have

$$\min_{\boldsymbol{x} \in \partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)} \angle(\boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}^*) = 0,$$

and there are either one or two elements in the set $\mathcal{P}$. The point $\boldsymbol{z}_2$ is the one that closer to the point $\boldsymbol{x}^*$. Note that $\boldsymbol{z}_2 \in P_2$. The illustration of the first case is shown in Figure 3.1.

In the second case, we have

$$\min_{\boldsymbol{x} \in \partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)} \angle(\boldsymbol{g}, \boldsymbol{x} - \boldsymbol{x}^*) > 0.$$

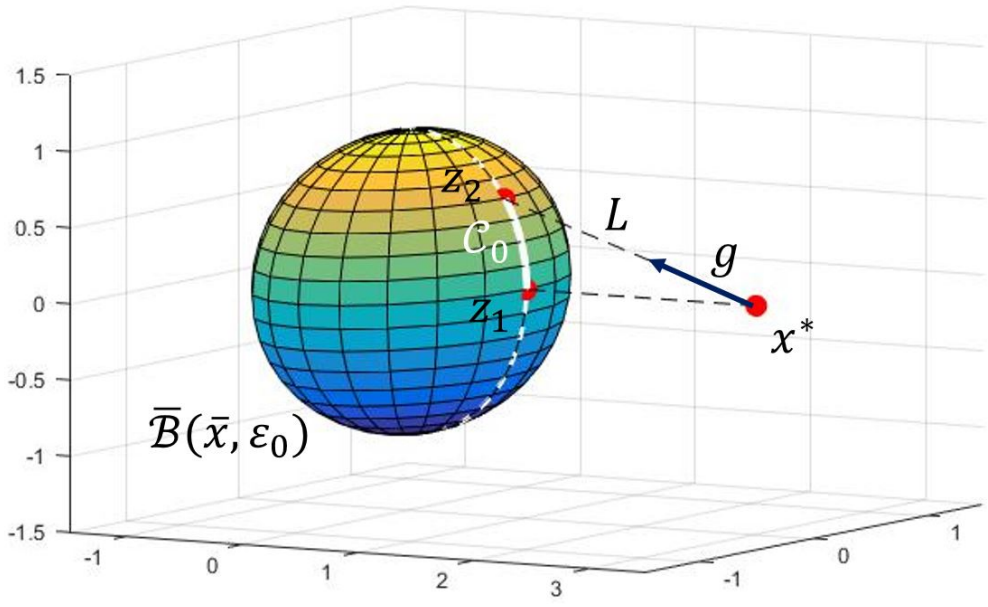


Figure 3.1. The points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, and the curve $\mathcal{C}_0$ on the surface $\partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ in the case that the ray L passes through the ball $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$.

The vector $\boldsymbol{z}_2 - \boldsymbol{x}^*$ is a tangent vector at the point $\boldsymbol{z}_2$ on the ball $\bar{\mathcal{B}}$ and has angle $\angle(\boldsymbol{z}_2 - \boldsymbol{x}^*, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) = \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$. Furthermore, the point $\boldsymbol{z}_2$ is on the plane P_2 since $\boldsymbol{z}_2 - \boldsymbol{x}^*$ and $\bar{\boldsymbol{x}} - \boldsymbol{x}^*$ must be on the same 2D-plane in order to minimize the angle between them. The illustration of the second case is shown in Figure 3.2.

Since P_2 passes through the center $\bar{\boldsymbol{x}}$ of the ball $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$, we can define the great circle $\mathcal{G} \subset P_2$ which is the intersection of $\partial\bar{\mathcal{B}}$ with P_2. Since $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are in $\mathcal{G}$ (and also in P_2), the geodesic path $\mathcal{C}_0$ is in P_2. The geodesic path in both cases is also shown in Figure 3.1 and Figure 3.2.

Before stating the theorem, define the open half-space

$$\mathcal{H}(\boldsymbol{g}, \boldsymbol{x}^*) := \{\boldsymbol{x} \in \mathbb{R}^n : \langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0\},$$

Note that $\mathcal{C}_0(\bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{g}, \boldsymbol{x}^*) \cap \mathcal{H}(\boldsymbol{g}, \boldsymbol{x}^*) \neq \emptyset$ as long as $\angle(\boldsymbol{g}, \boldsymbol{z}_2 - \boldsymbol{x}^*) < \frac{\pi}{2}$ or equivalently, $\angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$ as shown in Figure 3.3 and Figure 3.4.

We now come to the main result of this section.

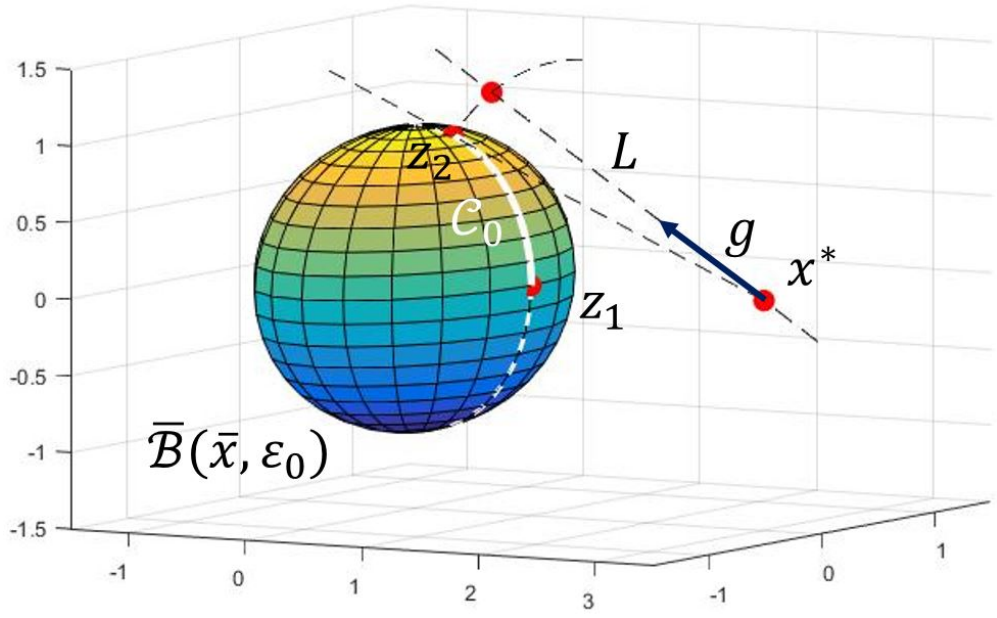


Figure 3.2. The points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, and the curve $\mathcal{C}_0$ on the surface $\partial\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ in the case that the ray L does not pass through the ball $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$.

Theorem 3.5.1. *Suppose $f^k \in \mathcal{F}$ and $\epsilon_0 > 0$. A necessary condition for a point $\boldsymbol{x}^* \in \mathbb{R}^n$ to be in $\mathcal{M}\Big(f^k, \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0), \sigma\Big) \setminus \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ is*

$$\min_{(\boldsymbol{x}_u^*, \boldsymbol{g}^k) \in \tilde{\mathcal{X}}(f^k, \bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{x}^*)} \frac{\langle \boldsymbol{g}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*) \rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma, \tag{3.13}$$

where

$$\tilde{\mathcal{X}}(f^k, \bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{x}^*) := \Big\{(\boldsymbol{x}, \boldsymbol{g}) \in \mathcal{C}_0(\bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{g}, \boldsymbol{x}^*) \times \partial f^k(\boldsymbol{x}^*)\}. \tag{3.14}$$

Proof. For a given $\boldsymbol{g}_{\boldsymbol{x}^*}^k \in \partial f^k(\boldsymbol{x}^*)$ with $\boldsymbol{g}_{\boldsymbol{x}^*}^k \neq \boldsymbol{0}$, we consider the angle $\angle(\boldsymbol{g}_{x^*}^k, \boldsymbol{x}^* - \bar{\boldsymbol{x}})$ in two disjoint cases:

(a) Suppose the gradient $\boldsymbol{g}_{\boldsymbol{x}^*}^k$ is colinear with the vector $\boldsymbol{x}^* - \bar{\boldsymbol{x}}$.

(i) If $\boldsymbol{g}_{\boldsymbol{x}^*}^k = \alpha(\boldsymbol{x}^* - \bar{\boldsymbol{x}})$ for some $\alpha > 0$ (i.e., $\boldsymbol{g}_{\boldsymbol{x}^*}^k$ is pointing directly away from $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ on the $x_{(1)}$-axis), then $\langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}) \rangle > 0$ for all $\boldsymbol{x} \in \overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$. Thus, no points in $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ can satisfy the inequality (3.10).

(ii) If $\boldsymbol{g}_{\boldsymbol{x}^*}^k = \alpha(\boldsymbol{x}^* - \bar{\boldsymbol{x}})$ for some $\alpha < 0$ (i.e., $\boldsymbol{g}_{\boldsymbol{x}^*}^k$ is pointing directly toward $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ on the $x_{(1)}$-axis), then the ray L passes through the ball $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ at $\boldsymbol{z}_1$, and thus $\{\boldsymbol{z}_1(\boldsymbol{x}^*)\} = \{\boldsymbol{z}_2(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{x}^*)\} = \mathcal{C}_0$. Furthermore, $\bar{\mathcal{B}} \subset \mathcal{H}$. For simplicity of notation, we will omit the arguments and write $\boldsymbol{z}_1(\boldsymbol{x}^*)$ and $\boldsymbol{z}_2(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{x}^*)$ as $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, respectively. From (3.12), for all $\boldsymbol{x} \in \partial\bar{\mathcal{B}}$, we have

$$\begin{aligned}
& \angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{z}_2, \boldsymbol{x}^*)) \leq \angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}^*)) \\
\Rightarrow \quad & \angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2)) \geq \angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})) \\
\Rightarrow \quad & \cos\angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2)) \leq \cos\angle(\boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})) \\
\Rightarrow \quad & \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2) \rangle \leq \langle \boldsymbol{g}_{\boldsymbol{x}^*}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}) \rangle.
\end{aligned} \tag{3.15}$$

Since $\boldsymbol{z}_2,\ \boldsymbol{x} \in \mathcal{H}$, we have $\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2)\rangle \leq \langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0$. In addition, from (3.11), for all $\boldsymbol{x} \in \partial\bar{\mathcal{B}}$, we have $0 < \|\boldsymbol{x}^* - \boldsymbol{z}_1\| \leq \|\boldsymbol{x}^* - \boldsymbol{x}\|$. Since $\boldsymbol{z}_1 = \boldsymbol{z}_2$ in this case, we obtain

$$\frac{-\|\boldsymbol{g}^k_{\boldsymbol{x}^*}\|}{\|\boldsymbol{x}^* - \boldsymbol{z}_1\|} = \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_1)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{z}_1\|} \leq \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}\|}$$

for all $\boldsymbol{x} \in \partial\bar{\mathcal{B}}$. Thus, it suffices to only check $\boldsymbol{z}_1 \in \mathcal{C}_0$ to see if (3.10) is satisfied.

(b) Suppose the gradient $\boldsymbol{g}^k_{\boldsymbol{x}^*}$ is not colinear with the vector $\boldsymbol{x}^* - \bar{\boldsymbol{x}}$. Then we can define the points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as described earlier. If

$$\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) \geq \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right),$$

then $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*) = \emptyset$ as shown in Figure 3.3, and no points in $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ can satisfy the inequality (3.10). If

$$\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right),$$

then $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*) \neq \emptyset$ and $\boldsymbol{z}_2 \in \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*)$ as shown in Figure 3.4. In this case, consider a point $\boldsymbol{x} \in \partial\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*)$ and $\boldsymbol{x} \notin \mathcal{C}_0$.

(i) Suppose $\|\boldsymbol{x} - \boldsymbol{x}^*\| > \|\boldsymbol{z}_2 - \boldsymbol{x}^*\|$. By the definition of $\boldsymbol{z}_2$ in (3.12), we have $\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{z}_2, \boldsymbol{x}^*)) \leq \angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}, \boldsymbol{x}^*))$. Since $\boldsymbol{z}_2,\ \boldsymbol{x} \in \mathcal{H}$, using the same argument as (3.15), we get $\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2)\rangle \leq \langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle < 0$. Therefore,

$$\frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{z}_2)\rangle}{\|\boldsymbol{z}_2 - \boldsymbol{x}^*\|} < \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x})\rangle}{\|\boldsymbol{x} - \boldsymbol{x}^*\|},$$

i.e., if $\boldsymbol{x}$ satisfies (3.10), then so does $\boldsymbol{z}_2$. This means that we can consider $\boldsymbol{z}_2 \in \mathcal{C}_0$ instead of any point in $\partial\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*)$ with greater distance from $\boldsymbol{x}^*$.

(ii) Suppose $\|\boldsymbol{x}-\boldsymbol{x}^*\| \leq \|\boldsymbol{z}_2-\boldsymbol{x}^*\|$. Since $\mathcal{C}_0$ is connected and $h(\boldsymbol{y}) = \|\boldsymbol{y}-\boldsymbol{x}^*\|$ is a continuous function, $\{h(\boldsymbol{y}) : \boldsymbol{y} \in \mathcal{C}_0\}$ is connected. Then, we have

$$\Big[\|\boldsymbol{z}_1-\boldsymbol{x}^*\|, \|\boldsymbol{z}_2-\boldsymbol{x}^*\|\Big] \subseteq \Big\{\|\boldsymbol{y}-\boldsymbol{x}^*\| : \boldsymbol{y} \in \mathcal{C}_0\Big\}.$$

Thus, there exists a $\boldsymbol{z} \in \mathcal{C}_0 \cap \mathcal{H}$ such that $\|\boldsymbol{z}-\boldsymbol{x}^*\| = \|\boldsymbol{x}-\boldsymbol{x}^*\|$. However, since $\mathcal{C}_0 \subset P_2$, we get that

$$\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{z},\boldsymbol{x}^*)) \leq \angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x},\boldsymbol{x}^*)).$$

Furthermore, since $\boldsymbol{z}$, $\boldsymbol{x} \in \mathcal{H}$, using the same argument as (3.15), we get

$$\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{z})\rangle < \langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{x})\rangle < 0.$$

In this case, we also have

$$\frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{z})\rangle}{\|\boldsymbol{z}-\boldsymbol{x}^*\|} \leq \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{x})\rangle}{\|\boldsymbol{x}-\boldsymbol{x}^*\|},$$

i.e., if $\boldsymbol{x}$ satisfies (3.10), then so does $\boldsymbol{z}$.

Thus, we conclude that for each point $\boldsymbol{x} \in \partial\bar{\mathcal{B}} \cap \mathcal{H}$, there is a point $\boldsymbol{z} \in \mathcal{C}_0$ such that $\frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{z})\rangle}{\|\boldsymbol{x}^*-\boldsymbol{z}\|} \leq \frac{\langle \boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{u}(\boldsymbol{x}^*,\boldsymbol{x})\rangle}{\|\boldsymbol{x}^*-\boldsymbol{x}\|}$. Therefore, to check if there is a point $\boldsymbol{x} \in \partial\bar{\mathcal{B}} \cap \mathcal{H}$ satisfying (3.10), we only need to check points in $\mathcal{C}_0$, yielding (3.13). □

In fact, we can replace $\mathcal{C}_0(\bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{g}, \boldsymbol{x}^*)$ in Theorem 3.5.1 by $\mathcal{C}_0(\bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{g}, \boldsymbol{x}^*) \cap \mathcal{H}(\boldsymbol{g}, \boldsymbol{x}^*)$. However, for simplicity of exposition, we forego the discussion of this further reduction in search space.

The set $\tilde{\mathcal{X}}(f^k, \bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{x}^*)$ defined in (3.14) suggests that we do not have to search for a pair $(\boldsymbol{x}^*, \boldsymbol{g}^k)$ that satisfies the inequality

$$\frac{\langle \boldsymbol{g}^k, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}^*_u)\rangle}{\|\boldsymbol{x}^*-\boldsymbol{x}^*_u\|} \leq -\sigma$$

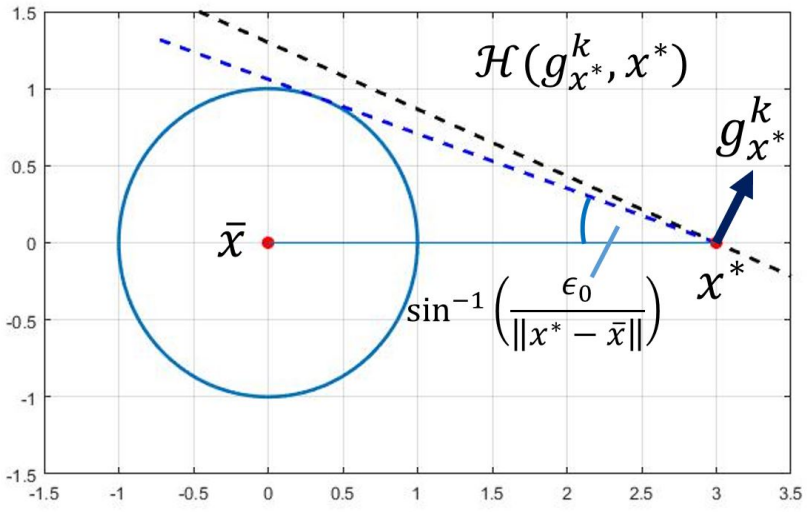


Figure 3.3. The area above the black dotted line is $\mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*)$ and the blue dotted line shows the angle $\arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$. In this case, the angle $\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) \geq \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$, so $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*) = \emptyset$.

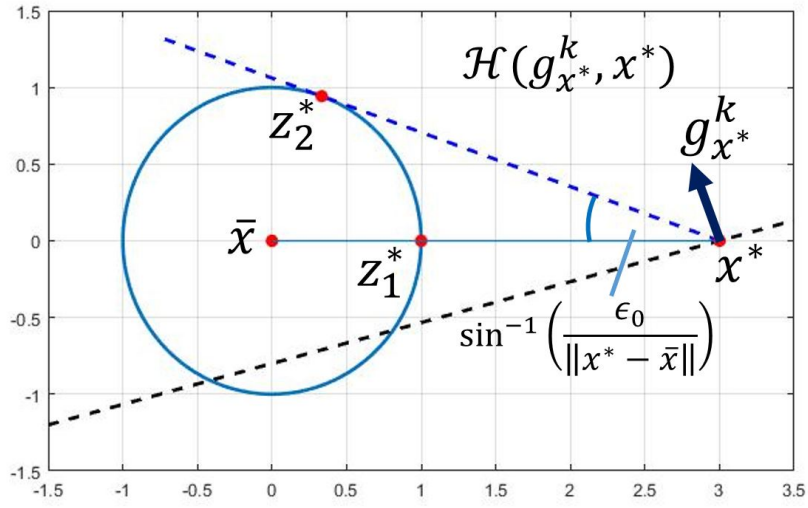


Figure 3.4. The area above the black dotted line is $\mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*)$ and the blue dotted line shows the angle $\arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$. In this case, the angle $\angle(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$, so $\bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0) \cap \mathcal{H}(\boldsymbol{g}^k_{\boldsymbol{x}^*}, \boldsymbol{x}^*) \neq \emptyset$.

throughout the set $\mathcal{X}(f^k, \mathcal{A}, \boldsymbol{x}^*)$ defined in (3.7) but can instead restrict our attention to be in the set $\tilde{\mathcal{X}}(f^k, \bar{\boldsymbol{x}}, \epsilon_0, \boldsymbol{x}^*)$ in (3.14). Since the curve $\mathcal{C}_0$ depends on the vector $\boldsymbol{g}^k$ that we choose from $\partial f^k(\boldsymbol{x}^*)$, we have to first select $\boldsymbol{g}^k \in \partial f^k(\boldsymbol{x}^*)$ and then we can consider the points on the curve $\mathcal{C}_0$ to see if they satisfy (3.10). We will use this in the algorithm for computing the region $\mathcal{M}$ in the next section.

3.6 Algorithm and Example

3.6.1 Algorithm

Consider the case from the previous section where the uncertainty set is a ball, i.e., $\mathcal{A} = \bar{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$. In this subsection, we will give an algorithm (Algorithm 1) to identify the region that satisfies the necessary condition (3.13). We provide a discussion of each of the steps below.

Algorithm 1 Region $\mathcal{M}$ Identification (Ball Case)

Let $X \subseteq \mathbf{dom}\, f^k$ be a set of points in the space
Input X, $f^k \in \mathcal{F}$, $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, $\epsilon_0 > 0$, and $\sigma > 0$
Output `minimizer`(X)

for $\boldsymbol{x}^* \in X$ **do** ▷ Loop through the space
  `minimizer`$(\boldsymbol{x}^*) \leftarrow$ false
  $d \leftarrow \|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|$
  $\boldsymbol{g} \leftarrow \nabla f^k(\boldsymbol{x}^*)$
  $\alpha \leftarrow \angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*)$
  if $\alpha < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{d}\right)$ **then**
    for $\theta \in \left[0, \arccos(\frac{\epsilon_0}{d})\right]$ **do**
      $\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\| \leftarrow \sqrt{d^2 + \epsilon_0^2 - 2\epsilon_0 d\cos\theta}$
      $\angle(\boldsymbol{g}, \boldsymbol{x}^* - \boldsymbol{x}_u^*) \leftarrow \alpha + \left[\boldsymbol{\pi} - \arcsin\left(\frac{\epsilon_0 \sin\theta}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|}\right)\right]$
      $\langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle \leftarrow \|\boldsymbol{g}\| \cos\angle(\boldsymbol{g}, \boldsymbol{x}^* - \boldsymbol{x}_u^*)$
      if $\frac{\langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*)\rangle}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|} \leq -\sigma$ **then**
        `minimizer`$(\boldsymbol{x}^*) \leftarrow$ true
      end if
    end for
  end if
end for
return `minimizer`(X)

Let X be a set of points; we wish to check whether each point in X is a potential minimizer of $f^k + f^u$. For simplicity, we assume that the function f^k is differentiable, i.e., $\partial f^k(\boldsymbol{x}^*) = \{\nabla f^k(\boldsymbol{x}^*)\}$ and the set of points $X \subseteq \mathbf{dom}\, f^k$. For example, we can use `linspace` in MATLAB to form a range for each axis, followed by using `meshgrid` to construct X. The object `minimizer` is an array that keeps a Boolean value for each point in X to indicate whether it is a potential minimizer. First, we loop through each point $\boldsymbol{x}^*$ in the set X and assign Boolean 'false' to that $\boldsymbol{x}^*$. In order to change the Boolean to be 'true', the point $\boldsymbol{x}^*$ has to satisfy the inequality (3.13). Before checking that inequality, we need to compute several intermediate variables. In the algorithm, we compute the distance between the center of the ball $\bar{\boldsymbol{x}}$ and the point $\boldsymbol{x}^*$ ($d \leftarrow \|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|$), the gradient of f^k at $\boldsymbol{x}^*$ ($g \leftarrow \nabla f^k(\boldsymbol{x}^*)$), and the angle between the gradient and reference ($\alpha \leftarrow \angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*)$). Note that we can compute α explicitly by

$$\alpha \leftarrow \angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) = \arccos\left(\frac{\langle \boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^* \rangle}{\|\boldsymbol{g}\| \|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|}\right).$$

We then verify the condition

$$\angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*) < \frac{\pi}{2} + \arcsin\left(\frac{\epsilon_0}{\|\boldsymbol{x}^* - \bar{\boldsymbol{x}}\|}\right)$$

(line 6); if this is not satisfied, no points in $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ can satisfy the inequality (3.10) as argued in the proof of Theorem 3.5.1 and illustrated in Figure 3.3. The next step is to compute the path $\mathcal{C}_0$, which we parametrize by using the variable θ. The variable θ in the algorithm corresponds to

$$\theta = \angle(\boldsymbol{x}_u^* - \bar{\boldsymbol{x}}, \boldsymbol{x}^* - \bar{\boldsymbol{x}}) \quad \text{where} \quad \boldsymbol{x}_u^* \in \mathcal{C}_0$$

as shown in Figure 3.5. So, we need to know the range of θ that characterizes the path $\mathcal{C}_0$. This range can be computed by considering the points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, at which the angle θ equals 0 and $\arccos(\frac{\epsilon_0}{\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|})$, respectively, as shown in Figure 3.6. Consider Figure 3.5. For each θ in the range (discretized to a sufficiently fine resolution), we can compute the distance

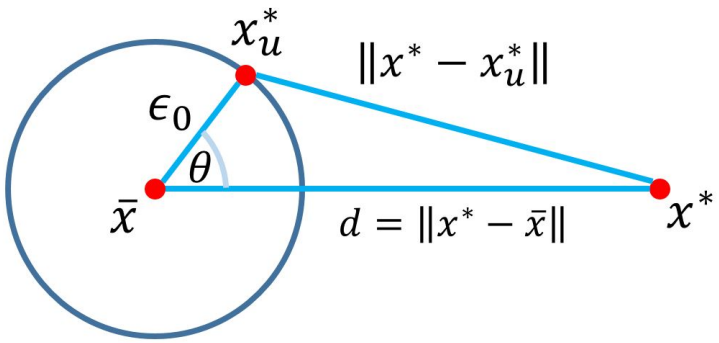


Figure 3.5. Given ϵ_0, d, and θ, we can compute $\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|$.

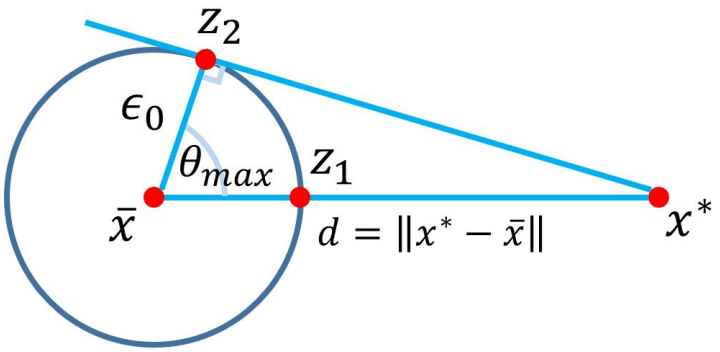


Figure 3.6. Given ϵ_0 and d, we can compute $\angle(\boldsymbol{z}_2 - \bar{\boldsymbol{x}}, \boldsymbol{x}^* - \bar{\boldsymbol{x}})$.

$\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|$ (line 8) by using the cosine law. Consider Figure 3.7. We can compute the angle $\angle(\boldsymbol{g}, \boldsymbol{x}^* - \boldsymbol{x}_u^*)$ (line 9) by using

$$\angle(\boldsymbol{g}, \boldsymbol{x}^* - \boldsymbol{x}_u^*) = \angle(\boldsymbol{g}, \boldsymbol{x}^* - \bar{\boldsymbol{x}}) + \arcsin\left(\frac{\epsilon_0 \sin\theta}{\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|}\right)$$
$$\text{and} \quad \angle(\boldsymbol{g}, \boldsymbol{x}^* - \bar{\boldsymbol{x}}) = (\boldsymbol{\pi} - \angle(\boldsymbol{g}, \bar{\boldsymbol{x}} - \boldsymbol{x}^*)).$$

After that we compute the inner product $\langle \boldsymbol{g}, \boldsymbol{u}(\boldsymbol{x}^*, \boldsymbol{x}_u^*) \rangle$ (line 10). Finally, we can compute the LHS of (3.13) and compare it to $-\sigma$. If the inequality (3.13) is satisfied by the current values $\boldsymbol{x}^*$ and θ, we set the Boolean associated to this $\boldsymbol{x}^*$ to be 'true'.

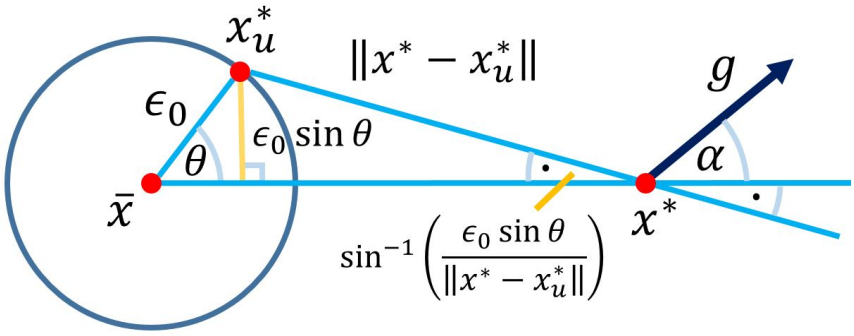


Figure 3.7. Given ϵ_0, θ, $\|\boldsymbol{x}^* - \boldsymbol{x}_u^*\|$, and α, we can compute $\angle(\boldsymbol{g}, \boldsymbol{x}^* - \boldsymbol{x}_u^*)$.

3.6.2 Example

Consider the known function $f^k(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2$, and suppose the unknown function f^u has minimizer in the ball centered at $\bar{\boldsymbol{x}} = (0, 0)$. We vary the radius of the ball of uncertainty (ϵ_0) among the values 0.1, 0.4, and 0.8, and the strong convexity parameter (σ) of the function f^u among the values 0.25, 2.0, and 5.0. Examples of the region that contains the possible minimizer of the sum $f^k + f^u$ are shown in Figure 3.8. In the figure, the function $f^k(\boldsymbol{x})$ is shown by using level curves and the uncertainty ball is shown by the beige circle. The region containing the possible minimizers of $f^k + f^u$ (i.e., the set of points $\boldsymbol{x} \in \mathbb{R}^n$ that satisfies (3.13)) is shown in blue (it contains the uncertainty set within it). Note that the solution region shrinks with increasing σ and grows with increasing ϵ_0.

3.7 Conclusions

In this chapter, we studied the properties of the minimizer of the sum of convex functions in which one of the functions is unknown but the others are known. However, we assumed that the unknown function is strongly convex with known convexity parameter, and that we have a region $\mathcal{A}$ where the minimizer of this function lies. We established a necessary condition for a given point to be a minimizer of the sum of known and unknown functions for general compact set $\mathcal{A}$. We then considered a special case where the region of the unknown function's minimizer is a ball. In this case, we simplified the necessary condition and provided an algorithm to determine the region that satisfies the necessary condition.

Future work could focus on providing sufficient conditions for a given point to be a minimizer (to complement our necessary condition). Alternatively, one could analyze properties of the set of solutions that satisfy the necessary condition.

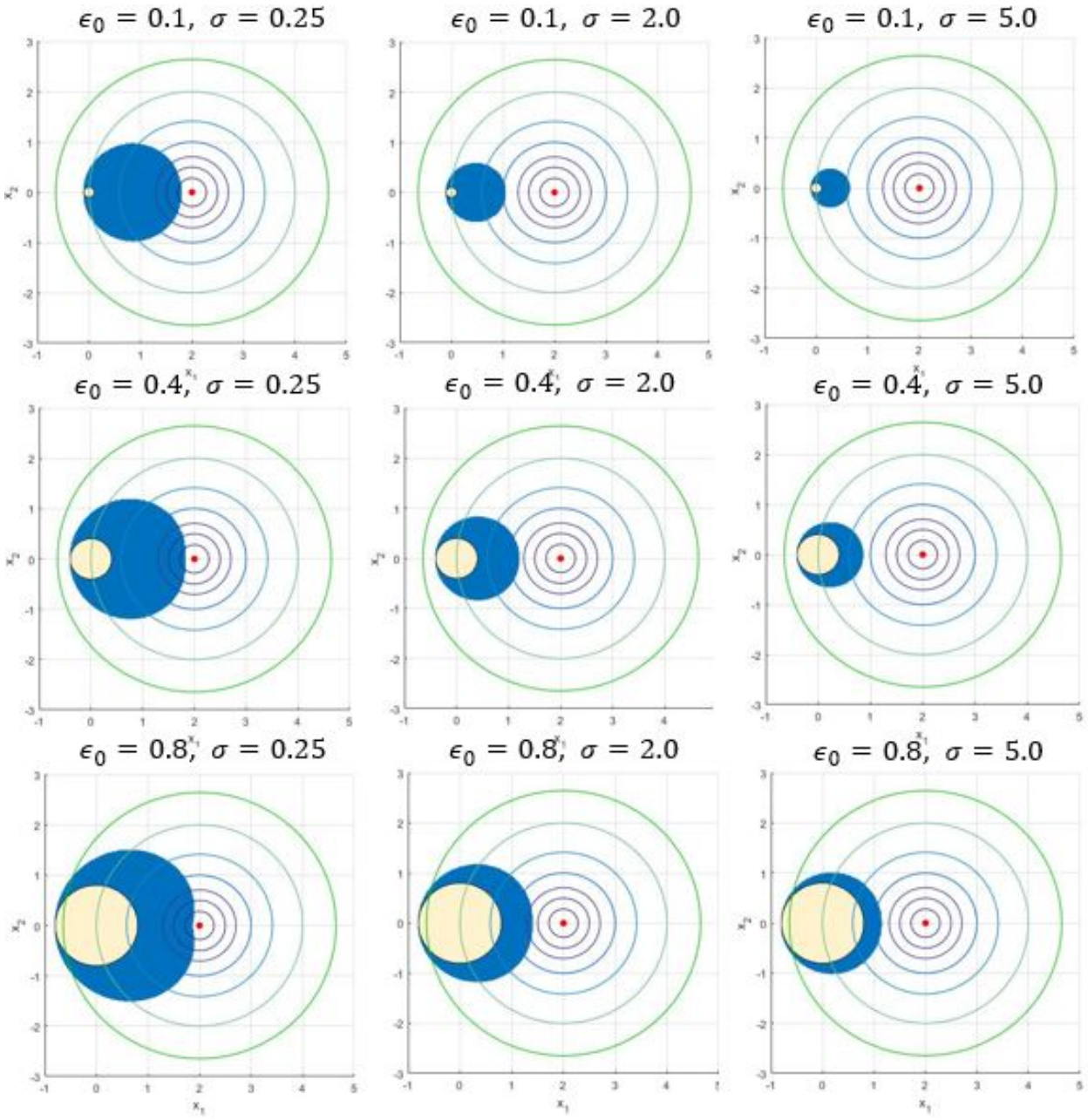


Figure 3.8. The function $f^k(\boldsymbol{x}) = (x_1 - 2)^2 + x_2^2$ is shown by the level curves while the balls $\overline{\mathcal{B}}(\bar{\boldsymbol{x}}, \epsilon_0)$ with the center at $(0, 0)$ are shown by the beige circle. The radius of the ball of uncertainty (ϵ_0) and the strong convexity parameter (σ) of the function f_m are varied and the solution sets are shown by the dark blue regions.

4. SCALABLE DISTRIBUTED OPTIMIZATION OF MULTI-DIMENSIONAL FUNCTIONS DESPITE BYZANTINE ADVERSARIES

4.1 Introduction

The design of distributed algorithms has received significant attention in the past few decades [61], [62]. In particular, for the problem of distributed optimization, a set of agents in a network are required to reach agreement on a parameter that minimizes the average of their local objective functions, using information received from their neighbors [9], [63]–[65]. A variety of approaches have been proposed to tackle different challenges of this problem, e.g., distributed optimization under constraints [25], distributed optimization under time-varying graphs [27], and distributed optimization for nonconvex nonsmooth functions [66]. However, these existing works typically make the assumption that all agents are trustworthy and cooperative (i.e., they follow the prescribed protocol); indeed, such protocols fail if even a single agent behaves in a malicious or incorrect manner [30].

As security becomes a more important consideration in large scale systems, it is crucial to develop algorithms that are resilient to agents that do not follow the prescribed algorithm. A handful of recent papers have considered fault tolerant algorithms for the case where agent misbehavior follows specific patterns [31], [32]. A more general (and serious) form of misbehavior is captured by the *Byzantine* adversary model from computer science, where misbehaving agents can send arbitrary (and conflicting) values to their neighbors at each iteration of the algorithm. Under such Byzantine behavior, it has been shown that it is impossible to guarantee computation of the true optimal point [30], [33]. Thus, researchers have begun formulating distributed optimization algorithms that allow the non-adversarial

nodes to converge to a certain region surrounding the true minimizer, regardless of the adversaries' actions [30], [34], [41].

It is worth noting that one major limitation of the above works is that they all make the assumption of scalar-valued objective functions, and the extension of the above ideas to general multi-dimensional convex functions remains largely open. In fact, one major challenge for minimizing multi-dimensional functions is that the region containing the minimizer of the sum of functions is itself difficult to characterize. Specifically, in contrast to the case of scalar functions, where the global minimizer[1] always lies within the smallest interval containing all local minimizers, the region containing the minimizer of the sum of multi-dimensional functions may not necessarily be in the convex hull of the minimizers (Chapter 2).

There exists a branch of literature focusing on secure distributed machine learning in a client-server architecture [42], [43], [67], where the server appropriately filters the information received from the clients. However, their extensions to a distributed (peer-to-peer) setting remains unclear. The papers [44], [45] consider a vector version of the resilient machine learning problem in a distributed (peer-to-peer) setting. These papers show that the states of regular nodes will converge to the statistical minimizer with high probability (as the amount of data of each node goes to infinity), but the analysis is restricted to i.i.d training data across the network. However, when each agent has a finite amount of data, these algorithms are still vulnerable to sophisticated attacks as shown in [46]. The work [47] considers a Byzantine distributed optimization problem for multi-dimensional functions, but relies on redundancy among the local functions, and also requires the underlying communication network to be complete. The recent work [68] studies resilient stochastic optimization problem. However, the assumptions made are quite different, in that it considers non-convex smooth functions, and the results do not ensure asymptotic consensus.

To the best of our knowledge, our conference paper [40] is the first one that provides a scalable algorithm with convergence guarantees in general networks under very general conditions on the multi-dimensional convex functions held by the agents in the presence of Byzantine faults. Different from existing works, the algorithm in [40] does not rely on any

[1]↑We will use the terms "global minimizer" and "minimizer of the sum" interchangeably since we only consider convex functions.

statistical assumptions or redundancy of local functions. Technically, the analysis addresses the challenge of finding a region that contains the global minimizer for multiple-dimensional functions, and shows that regular states are guaranteed to converge to that region under the proposed algorithm. The Distance-MinMax Filtering Dynamics in [40] requires each regular node to compute an auxiliary point using resilient asymptotic consensus techniques on their individual functions' minimizers in advance. After that, there are two filtering steps in the main algorithm that help regular nodes to discard extreme states. The first step is to remove extreme states (based on the distance to the auxiliary point), and the second step is to remove states that have extreme values in any of their components. On the other hand, the algorithm in [40] suffers from the need to compute the auxiliary point prior to running the main algorithm, since the fixed auxiliary point is only achieved by the resilient consensus algorithm asymptotically.

In this chapter, we eliminate this drawback. The algorithms and analysis we propose here expand upon the work in [40] in the following significant ways. First, the algorithms in this chapter bring the computation of the auxiliary point into the main algorithm, so that the local update of auxiliary point and local filtering strategies are performed simultaneously. This makes the analysis much more involved since we need to take into account the coupled dynamics of the estimated auxiliary point and the optimization variables. Second, the algorithms make better use of local information by including each regular node's own state as a metric. In practice, we observe that this performs better than the approach in [40], since each agent may discard fewer states and hence, there are more non-extreme states that can help the regular agents get close to the true global minimizer. Again, we characterize the convergence region that all regular states are guaranteed to converge to using the proposed algorithm. Third, we present an alternate algorithm in this chapter which only makes use of the distance filter (as opposed to both the distance and min-max filter); we show that this algorithm significantly reduces the requirements on the network topology for our convergence guarantees, at the cost of losing guarantees on consensus of the regular nodes' states. Importantly, our work represents the first attempt to provide convergence guarantees in a geometric sense, characterizing a region where all states are ensured to converge to, without relying on any statistical assumptions or redundancy of local functions.

This chapter is organized as follows. Section 4.2 introduces various mathematical preliminaries, and states the problem of resilient distributed optimization. We provide our proposed algorithms in Section 4.3. We then state the assumptions and some important results related to properties of the proposed algorithms in Section 4.4. In Section 4.5, we provide discussion on the results. Finally, we simulate our algorithms to numerically evaluate their performance in Section 4.6, and conclude in Section 4.7.

4.2 Mathematical Notation and Problem Formulation

Let $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ denote the set of natural numbers (including zero), integers, and real numbers, respectively. We also denote the set of positive integers by $\mathbb{Z}_+$. The cardinality of a set is denoted by $|\cdot|$. The set of subgradients of a convex function f at point $\boldsymbol{x}$ is called the subdifferential of f at $\boldsymbol{x}$, and is denoted $\partial f(\boldsymbol{x})$.

4.2.1 Linear Algebra

Vectors are taken to be column vectors, unless otherwise noted. We use $x^{(\ell)}$ to represent the ℓ-th component of a vector $\boldsymbol{x}$. The Euclidean norm on $\mathbb{R}^d$ is denoted by $\|\cdot\|$. We denote by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ the Euclidean inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^T \boldsymbol{v}$ and by $\angle(\boldsymbol{u}, \boldsymbol{v})$ the angle between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\angle(\boldsymbol{u}, \boldsymbol{v}) = \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right)$. We use $\mathcal{S}_d^+$ to denote the set of positive definite matrices in $\mathbb{R}^{d \times d}$. The Euclidean ball in d-dimensional space with center at $\boldsymbol{x}_0$ and radius $r \in \mathbb{R}_{>0}$ is denoted by $\mathcal{B}(\boldsymbol{x}_0, r) := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}$.

4.2.2 Graph Theory

We denote a network by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which consists of the set of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ and the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. If $(v_i, v_j) \in \mathcal{E}$, then node v_j can receive information from node v_i. The in-neighbor and out-neighbor sets are denoted by $\mathcal{N}_i^{\text{in}} = \{v_j \in \mathcal{V} : \ (v_j, v_i) \in \mathcal{E}\}$ and $\mathcal{N}_i^{\text{out}} = \{v_j \in \mathcal{V} : \ (v_i, v_j) \in \mathcal{E}\}$, respectively. A path from node $v_i \in \mathcal{V}$ to node $v_j \in \mathcal{V}$ is a sequence of nodes $v_{k_1}, v_{k_2}, \ldots, v_{k_l}$ such that $v_{k_1} = v_i$, $v_{k_l} = v_j$ and $(v_{k_r}, v_{k_{r+1}}) \in \mathcal{E}$ for $1 \leq r \leq l-1$. Throughout this chapter, the terms nodes and agents

will be used interchangeably. Given a set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, where each $\boldsymbol{x}_i \in \mathbb{R}^d$, we define for all $\mathcal{S} \subseteq \mathcal{V}$,

$$\{\boldsymbol{x}_i\}_{\mathcal{S}} := \{\boldsymbol{x}_i \in \mathbb{R}^d : v_i \in \mathcal{S}\}.$$

Definition 4.2.1. *A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be rooted at node $v_i \in \mathcal{V}$ if for all nodes $v_j \in \mathcal{V} \setminus \{v_i\}$, there is a path from v_i to v_j. A graph is said to be rooted if it is rooted at some node $v_i \in \mathcal{V}$.*

We will rely on the following definitions from [69].

Definition 4.2.2 (r-reachable set)**.** *For a given graph $\mathcal{G}$ and a positive integer $r \in \mathbb{Z}_+$, a subset of nodes $\mathcal{S} \subseteq \mathcal{V}$ is said to be r-reachable if there exists a node $v_i \in \mathcal{S}$ such that $|\mathcal{N}_i^{\text{in}} \setminus \mathcal{S}| \geq r$.*

Definition 4.2.3 (r-robust graph)**.** *For $r \in \mathbb{Z}_+$, a graph $\mathcal{G}$ is said to be r-robust if for all pairs of disjoint nonempty subsets $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{V}$, at least one of $\mathcal{S}_1$ or $\mathcal{S}_2$ is r-reachable.*

The above definitions capture the idea that sets of nodes should contain individual nodes that have a sufficient number of neighbors outside that set. This will be important for the *local* decisions made by each node in the network under our algorithm, and will allow information from the rest of the network to penetrate into different sets of nodes.

4.2.3 Adversarial Behavior

Definition 4.2.4. *A node $v_i \in \mathcal{V}$ is said to be Byzantine if during each iteration of the prescribed algorithm, it is capable of sending arbitrary (and perhaps conflicting) values to different neighbors. It is also allowed to update its local information arbitrarily at each iteration of any prescribed algorithm.*

The set of Byzantine nodes is denoted by $\mathcal{A} \subset \mathcal{V}$. The set of regular nodes is denoted by $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$.

The identities of the Byzantine agents are unknown to regular agents in advance. Furthermore, we allow the Byzantine agents to know the entire topology of the network, functions equipped by the regular nodes, and the deployed algorithm. In addition, Byzantine agents

are allowed to coordinate with other Byzantine agents and access the current and previous information contained by the nodes in the network (e.g. current and previous states of all nodes). Such extreme behavior is typical in the study of the adversarial models [30], [33], [44]. In exchange for allowing such extreme behavior, we will consider a limitation on the number of such adversaries in the neighborhood of each regular node, as follows.

Definition 4.2.5 (F-local model)**.** *For $F \in \mathbb{Z}_+$, we say that the set of adversaries $\mathcal{A}$ is an F-local set if $|\mathcal{N}_i^{\text{in}} \cap \mathcal{A}| \leq F$, for all $v_i \in \mathcal{R}$.*

Thus, the F-local model captures the idea that each regular node has at most F Byzantine in-neighbors.

4.2.4 Problem Formulation

Consider a group of N agents $\mathcal{V}$ interconnected over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each agent $v_i \in \mathcal{V}$ has a local convex cost function $f_i : \mathbb{R}^d \to \mathbb{R}$. The objective is to collaboratively solve the minimization problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \frac{1}{N} \sum_{v_i \in \mathcal{V}} f_i(\boldsymbol{x}), \tag{4.1}$$

where $\boldsymbol{x} \in \mathbb{R}^d$ is the common decision variable. A common approach to solve such problems is for each agent to maintain a local estimate of the solution to the above problem, which it iteratively updates based on communications with its immediate neighbors. However, since Byzantine nodes are allowed to send arbitrary values to their neighbors at each iteration of any algorithm, it is not possible to solve Problem (4.1) under such misbehavior (since one is not guaranteed to infer any information about the true functions of the Byzantine agents) [30], [33]. Thus, the optimization problem is recast into the following form:

$$\min_{\boldsymbol{x} \in \mathbb{R}^d} \frac{1}{|\mathcal{R}|} \sum_{v_i \in \mathcal{R}} f_i(\boldsymbol{x}), \tag{4.2}$$

i.e., we restrict our attention only to the functions held by regular nodes.

Remark 5. The challenge in solving the above problem lies in the fact that no regular agent is aware of the identities or actions of the Byzantine agents. Furthermore, in the worst-case scenario, it is not feasible to achieve an exact solution to Problem 4.2, as the Byzantine agents can modify the functions while still adhering to the algorithm, making it impossible to differentiate them [30], [33].

In the next section, we propose two scalable algorithms that allow the regular nodes to approximately solve the above problem, regardless of the identities or actions of the Byzantine agents (as proven later in the chapter).

4.3 Resilient Distributed Optimization Algorithms

4.3.1 Proposed Algorithms

The algorithms that we propose are stated as Algorithm 2 and Algorithm 3. We start with Algorithm 1. At each time-step k, each regular node[2] $v_i \in \mathcal{R}$ maintains and updates a vector $\boldsymbol{x}_i[k] \in \mathbb{R}^d$, which is its estimate of the solution to Problem (4.2), and a vector $\boldsymbol{y}_i[k] \in \mathbb{R}^d$, which is its estimate of an auxiliary point that provides a general sense of direction for each agent to follow. Specifically, the auxiliary points $\boldsymbol{y}_i[k]$ will be used to perform the distance-based filtering step (**Line 7**) in which the neighbors' states $\{\boldsymbol{x}_j[k]\}_{v_j \in \mathcal{N}_i^{\text{in}}}$ far from $\boldsymbol{y}_i[k]$ are removed at time-step k. We now explain each step used in Algorithm 2 in detail.[3]

- **Line 1:** $\hat{\boldsymbol{x}}_i^* \leftarrow$ `optimize` (f_i)

  Each node $v_i \in \mathcal{R}$ uses any appropriate optimization algorithm to get an approximate minimizer $\hat{\boldsymbol{x}}_i^* \in \mathbb{R}^d$ of its local function f_i. We assume that there exists $\epsilon^* \in \mathbb{R}_{\geq 0}$ such that the algorithm achieves $\|\hat{\boldsymbol{x}}_i^* - \boldsymbol{x}_i^*\| \leq \epsilon^*$ for all $v_i \in \mathcal{R}$ where $\boldsymbol{x}_i^* \in \mathbb{R}^d$ is a true minimizer of the function f_i; we assume formally that such a true (but not necessary unique) minimizer exists for each $v_i \in \mathcal{R}$ in the next section.

[2]↑Byzantine nodes do not necessarily need to follow the above algorithm, and can update their states however they wish.

[3]↑In the algorithm, $\mathcal{X}_i[k]$, $\mathcal{X}_i^{\text{dist}}[k]$, $\mathcal{X}_i^{\text{mm}}[k]$, $\mathcal{Y}_i[k]$ and $\mathcal{Y}_i^{\text{mm}}[k]$ are multisets.

Algorithm 2 Simultaneous Distance-MinMax Filtering Dynamics

Input Network $\mathcal{G}$, functions $\{f_i\}_{i=1}^N$, parameter F

Each $v_i \in \mathcal{R}$ sets $\hat{\boldsymbol{x}}_i^* \leftarrow \texttt{optimize}(f_i)$
Each $v_i \in \mathcal{R}$ sets $\boldsymbol{x}_i[0] \leftarrow \hat{\boldsymbol{x}}_i^*$ and $\boldsymbol{y}_i[0] \leftarrow \hat{\boldsymbol{x}}_i^*$
for $k = 0, 1, 2, 3, \ldots$ **do**
    for $v_i \in \mathcal{R}$ **do** ▷ Implement in parallel
        Step I: Broadcast and Receive
        $\texttt{broadcast}(\mathcal{N}_i^{\text{out}},\ \boldsymbol{x}_i[k],\ \boldsymbol{y}_i[k])$
        $\mathcal{X}_i[k],\ \mathcal{Y}_i[k] \leftarrow \texttt{receive}(\mathcal{N}_i^{\text{in}})$
        Step II: Resilient Consensus Step
        $\mathcal{X}_i^{\text{dist}}[k] \leftarrow \texttt{dist\_filt}(F,\ \boldsymbol{y}_i[k],\ \mathcal{X}_i[k])$
        $\mathcal{X}_i^{\text{mm}}[k] \leftarrow \texttt{x\_minmax\_filt}(F,\ \mathcal{X}_i^{\text{dist}}[k])$
        $\boldsymbol{z}_i[k] \leftarrow \texttt{x\_weighted\_average}(\mathcal{X}_i^{\text{mm}}[k])$
        Step III: Gradient Update
        $\boldsymbol{x}_i[k+1] \leftarrow \texttt{gradient}(f_i, \boldsymbol{z}_i[k])$
        Step IV: Update the Estimated Auxiliary Point
        $\mathcal{Y}_i^{\text{mm}}[k] \leftarrow \texttt{y\_minmax\_filt}(F,\ \mathcal{Y}_i[k])$
        $\boldsymbol{y}_i[k+1] \leftarrow \texttt{y\_weighted\_average}(\mathcal{Y}_i^{\text{mm}}[k])$
    end for
end for

- **Line 2:** $\boldsymbol{x}_i[0] \leftarrow \hat{\boldsymbol{x}}_i^*$ and $\boldsymbol{y}_i[0] \leftarrow \hat{\boldsymbol{x}}_i^*$

  Each node $v_i \in \mathcal{R}$ initializes its own estimated solution to Problem (4.2) ($\boldsymbol{x}_i[0] \in \mathbb{R}^d$) and estimated auxiliary point ($\boldsymbol{y}_i[0] \in \mathbb{R}^d$) to be $\hat{\boldsymbol{x}}_i^*$.

- **Line 5:** `broadcast` ($\mathcal{N}_i^{\text{out}}$, $\boldsymbol{x}_i[k]$, $\boldsymbol{y}_i[k]$)

  Node $v_i \in \mathcal{R}$ broadcasts its current state $\boldsymbol{x}_i[k]$ and estimated auxiliary point $\boldsymbol{y}_i[k]$ to its out-neighbors $\mathcal{N}_i^{\text{out}}$.

- **Line 6:** $\mathcal{X}_i[k],\ \mathcal{Y}_i[k] \leftarrow$ `receive`$(\mathcal{N}_i^{\text{in}})$

  Node $v_i \in \mathcal{R}$ receives the current states $\boldsymbol{x}_j[k]$ and $\boldsymbol{y}_j[k]$ from its in-neighbors $\mathcal{N}_i^{\text{in}}$. So, at time step k, node v_i possesses the sets of states[4]

$$\mathcal{X}_i[k] := \left\{\boldsymbol{x}_j[k] \in \mathbb{R}^d : v_j \in \mathcal{N}_i^{\text{in}} \cup \{v_i\}\right\} \quad \text{and}$$
$$\mathcal{Y}_i[k] := \left\{\boldsymbol{y}_j[k] \in \mathbb{R}^d : v_j \in \mathcal{N}_i^{\text{in}} \cup \{v_i\}\right\}.$$

  The sets $\mathcal{X}_i[k]$ and $\mathcal{Y}_i[k]$ have an indirect relationship through the distance-based filter (**Line 7**) as only $\boldsymbol{y}_i[k] \in \mathcal{Y}_i[k]$ is used as the reference to remove states in $\mathcal{X}_i[k]$.

- **Line 7:** $\mathcal{X}_i^{\text{dist}}[k] \leftarrow$ `dist_filt`$(F,\ \boldsymbol{y}_i[k],\ \mathcal{X}_i[k])$

  Intuitively, regular node v_i ignores the states that are far away from its own auxiliary state $\boldsymbol{y}_i[k]$ in L^2 sense. Formally, node $v_i \in \mathcal{R}$ computes the distance between each vector in $\mathcal{X}_i[k]$ and its own estimated auxiliary point $\boldsymbol{y}_i[k]$:

$$\mathcal{D}_{ij}[k] := \|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \ \text{ for } \ \boldsymbol{x}_j[k] \in \mathcal{X}_i[k]. \tag{4.3}$$

  Then, node $v_i \in \mathcal{R}$ sorts the values in the set $\{\mathcal{D}_{ij}[k] : v_j \in \mathcal{N}_i^{\text{in}} \cup \{v_i\}\}$ and removes the F largest values that are larger than its own value $\mathcal{D}_{ii}[k]$. If there are fewer than F values higher than its own value, v_i removes all of those values. Ties in values are broken arbitrarily. The corresponding states of the remaining values are stored in

[4]↑In case a regular node v_i has a Byzantine neighbor v_j, we abuse notation and take the value $\boldsymbol{x}_j[k]$ to be the value received from node v_j (i.e., it does not have to represent the true state of node v_j).

$\mathcal{X}_i^{\text{dist}}[k]$. In other words, regular node v_i removes up to F of its neighbors' vectors that are furthest away from the auxiliary point $\boldsymbol{y}_i[k]$.

- **Line 8:** $\mathcal{X}_i^{\text{mm}}[k] \leftarrow \texttt{x\_minmax\_filt}(F,\ \mathcal{X}_i^{\text{dist}}[k])$

  Intuitively, regular node v_i ignores the states that contains extreme values in any of their components in the ordering sense. Formally, for each time-step $k \in \mathbb{N}$ and dimension $\ell \in \{1, 2, \ldots, d\}$, define the set $\mathcal{V}_i^{\text{remove}}(\ell)[k] \subseteq \mathcal{N}_i^{\text{in}}$, where a node v_j is in $\mathcal{V}_i^{\text{remove}}(\ell)[k]$ if and only if

  - $x_j^{(\ell)}[k]$ is within the F-largest values of the set $\left\{x_r^{(\ell)}[k] \in \mathbb{R} : \boldsymbol{x}_r[k] \in \mathcal{X}_i^{\text{dist}}[k]\right\}$ and $x_j^{(\ell)}[k] > x_i^{(\ell)}[k]$, or
  - $x_j^{(\ell)}[k]$ is within the F-smallest values of the set $\left\{x_r^{(\ell)}[k] \in \mathbb{R} : \boldsymbol{x}_r[k] \in \mathcal{X}_i^{\text{dist}}[k]\right\}$ and $x_j^{(\ell)}[k] < x_i^{(\ell)}[k]$.

  Ties in values are broken arbitrarily. Node v_i then removes the state of all nodes in $\bigcup_{\ell \in \{1,2,\ldots,d\}} \mathcal{V}_i^{\text{remove}}(\ell)[k]$ and the remaining states are stored in $\mathcal{X}_i^{\text{mm}}[k]$:

$$\mathcal{X}_i^{\text{mm}}[k] = \left\{\boldsymbol{x}_j[k] \in \mathbb{R}^d : v_j \in \mathcal{V}_i^{\text{dist}}[k] \setminus \bigcup_{\ell \in \{1,\ldots,d\}} \mathcal{V}_i^{\text{remove}}(\ell)[k]\right\}, \tag{4.4}$$

  where $\mathcal{V}_i^{\text{dist}}[k] = \left\{v_j \in \mathcal{R} : \boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{dist}}[k]\right\}$.

- **Line 9:** $\boldsymbol{z}_i[k] \leftarrow \texttt{x\_weighted\_average}(\mathcal{X}_i^{\text{mm}}[k])$

  Each node $v_i \in \mathcal{R}$ computes

$$\boldsymbol{z}_i[k] = \sum_{\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{mm}}[k]} w_{x,ij}[k]\ \boldsymbol{x}_j[k], \tag{4.5}$$

  where $w_{x,ij}[k] > 0$ for all $\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{mm}}[k]$ and $\sum_{\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{mm}}[k]} w_{x,ij}[k] = 1$.

- **Line 10:** $\boldsymbol{x}_i[k+1] \leftarrow \texttt{gradient}\ (f_i, \boldsymbol{z}_i[k])$

  Node $v_i \in \mathcal{R}$ computes the gradient update as follows:

$$\boldsymbol{x}_i[k+1] = \boldsymbol{z}_i[k] - \eta[k]\ \boldsymbol{g}_i[k], \tag{4.6}$$

where $\boldsymbol{g}_i[k] \in \partial f_i(\boldsymbol{z}_i[k])$ and $\eta[k]$ is the step-size at time k. The conditions corresponding to the step-size are given in the next section.

- **Line 11:** $\mathcal{Y}_i^{\text{mm}}[k] \leftarrow$ `y_minmax_filt`$(F,\ \mathcal{Y}_i[k])$
  For each dimension $\ell \in \{1, 2, \ldots, d\}$, node $v_i \in \mathcal{R}$ removes the F highest and F lowest values of its neighbors' auxiliary points along that dimension. More specifically, for each dimension $\ell \in \{1, 2, \ldots, d\}$, node v_i sorts the values in the set of scalars $\{y_j^{(\ell)}[k] : \boldsymbol{y}_j[k] \in \mathcal{Y}_i[k]\}$ and then removes the F largest and F smallest values that are larger and smaller than its own value, respectively. If there are fewer than F values higher (resp. lower) than its own value, v_i removes all of those values. Ties in values are broken arbitrarily. The remaining values are stored in $\mathcal{Y}_i^{\text{mm}}[k](\ell)$ and the set $\mathcal{Y}_i^{\text{mm}}[k]$ is the collection of $\mathcal{Y}_i^{\text{mm}}[k](\ell)$, i.e., $\mathcal{Y}_i^{\text{mm}}[k] = \left\{\mathcal{Y}_i^{\text{mm}}[k](\ell) : \ell \in \{1, 2, \ldots, d\}\right\}$.

- **Line 12:** $\boldsymbol{y}_i[k+1] \leftarrow$ `y_weighted_average`$(\mathcal{Y}_i^{\text{mm}}[k])$
  For each dimension $\ell \in \{1, 2, \ldots, d\}$, each node $v_i \in \mathcal{R}$ computes

$$y_i^{(\ell)}[k+1] = \sum_{y_j^{(\ell)}[k] \in \mathcal{Y}_i^{\text{mm}}[k](\ell)} w_{y,ij}^{(\ell)}[k]\ y_j^{(\ell)}[k], \tag{4.7}$$

  where $w_{y,ij}^{(\ell)}[k] > 0$ for all $y_j^{(\ell)}[k] \in \mathcal{Y}_i^{\text{mm}}[k](\ell)$ and $\sum_{y_j^{(\ell)}[k] \in \mathcal{Y}_i^{\text{mm}}[k](\ell)} w_{y,ij}^{(\ell)}[k] = 1$.

Note that the filtering processes `x_minmax_filt` (**Line 8**) and `y_minmax_filt` (**Line 11**) are different. In `x_minmax_filt`, each node removes the **whole** state vector for a neighbor if it contains an extreme value in any component, while in `y_minmax_filt`, each node only removes the extreme components in each vector. In addition, `x_weighted_average` (**Line 9**) and `y_weighted_average_2` (**Line 12**) are also different in that `x_weighted_average` designates agent v_i at time-step k to utilize the same set of weights $\{w_{x,ij} \in \mathbb{R} : \boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{mm}}[k]\}$ for all components while `y_weighted_average` allows agent v_i at time-step k to use a different set of weights $\{w_{y,ij}^{(\ell)} \in \mathbb{R} : y_j^{(\ell)}[k] \in \mathcal{Y}_i^{\text{mm}}[k](\ell)\}$ for each coordinate ℓ (since the number of remaining values in each component $|\mathcal{Y}_i^{\text{mm}}[k](\ell)|$ is not necessarily the same). These differences will become clear when considering the example provided in the next subsection.

We also consider a variant of Algorithm 2 defined as follows.

Algorithm 3 Simultaneous Distance Filtering Dynamics

Algorithm 3 is the same as Algorithm 2 except that

- **Line 8** is removed, and
- $\mathcal{X}_i^{\mathrm{mm}}[k]$ in **Line 9** is replaced by $\mathcal{X}_i^{\mathrm{dist}}[k]$.

Although Algorithms 2 and 3 are very similar (differing only in the use of an additional filter in Algorithm 2), our subsequent analysis will reveal the relative costs and benefits of each algorithm. We emphasize that both algorithms involve only simple operations in each iteration, and that the regular agents do not need to know the network topology, or functions possessed by another agents. Furthermore, the regular agents do not need to know the identities of adversaries; they only need to know the upper bound for the number of local adversaries. However, we assume that all regular agents use the same step-size $\eta[k]$ (**Line 10**, equation (4.6)).

4.3.2 Example of Algorithm 2

Before we prove the convergence properties of the algorithms, we first demonstrate Algorithm 2, which is more complicated due to the min-max filtering step (**Line 8**), step by step using an example.

Suppose there are 8 agents forming the complete graph (for the purpose of illustration). Let node v_i have the local objective function $f_i : \mathbb{R}^2 \to \mathbb{R}$ defined as $f_i(\boldsymbol{x}) = (x^{(1)} + i)^2 + (x^{(2)} - i)^2$ for all $i \in \{1, 2, \ldots, 8\}$. Let the set of adversarial nodes be $\mathcal{A} = \{v_4, v_8\}$ and thus, we have $\mathcal{R} = \{v_1, v_2, v_3, v_5, v_6, v_7\}$. Note that only the regular nodes execute the algorithm

(and they do not know which agents are adversarial). Let $F = 2$ and at some time-step $\hat{k} \in \mathbb{N}$, each regular node has the following state and the estimated auxiliary point:[5]

$$\begin{aligned}
\boldsymbol{x}_1[\hat{k}] &= \begin{bmatrix} 4 & 2 \end{bmatrix}^T, \quad \boldsymbol{y}_1[\hat{k}] = \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \\
\boldsymbol{x}_2[\hat{k}] &= \begin{bmatrix} 4 & 1 \end{bmatrix}^T, \quad \boldsymbol{y}_2[\hat{k}] = \begin{bmatrix} -1 & -2 \end{bmatrix}^T, \\
\boldsymbol{x}_3[\hat{k}] &= \begin{bmatrix} 3 & 3 \end{bmatrix}^T, \quad \boldsymbol{y}_3[\hat{k}] = \begin{bmatrix} -2 & 1 \end{bmatrix}^T, \\
\boldsymbol{x}_5[\hat{k}] &= \begin{bmatrix} 2 & 1 \end{bmatrix}^T, \quad \boldsymbol{y}_5[\hat{k}] = \begin{bmatrix} 0 & 2 \end{bmatrix}^T, \\
\boldsymbol{x}_6[\hat{k}] &= \begin{bmatrix} 1 & 4 \end{bmatrix}^T, \quad \boldsymbol{y}_6[\hat{k}] = \begin{bmatrix} 1 & 3 \end{bmatrix}^T, \\
\boldsymbol{x}_7[\hat{k}] &= \begin{bmatrix} 0 & 0 \end{bmatrix}^T, \quad \boldsymbol{y}_7[\hat{k}] = \begin{bmatrix} 1 & 3 \end{bmatrix}^T.
\end{aligned}$$

Let $\boldsymbol{x}_{a\to b}[k]$ (resp. $\boldsymbol{y}_{a\to b}[k]$) be the state (resp. estimated auxiliary point) that is sent from the adversarial node $v_a \in \mathcal{A}$ to the regular node $v_b \in \mathcal{R}$ at time-step k. Suppose that in time-step $\hat{k}$, each adversarial agent sends the same states and the same estimated auxiliary points to its neighbors (although this is not necessary) as follows:

$$\begin{aligned}
\boldsymbol{x}_{4\to i}[\hat{k}] &= \begin{bmatrix} 3 & 2 \end{bmatrix}^T, \quad \boldsymbol{y}_{4\to i}[\hat{k}] = \begin{bmatrix} -1 & 1 \end{bmatrix}^T, \\
\boldsymbol{x}_{8\to i}[\hat{k}] &= \begin{bmatrix} 0 & 5 \end{bmatrix}^T, \quad \boldsymbol{y}_{8\to i}[\hat{k}] = \begin{bmatrix} 2 & 2 \end{bmatrix}^T
\end{aligned}$$

for all $i \in \{1, 2, 3, 5, 6, 7\}$. We will demonstrate the calculation of $\boldsymbol{x}_1[\hat{k}+1]$ and $\boldsymbol{y}_1[\hat{k}+1]$, computed by regular node v_1.

Since the network is the complete graph, the set of in-neighbors and out-neighbors of node v_1 is $\mathcal{N}_1^{\text{in}} = \mathcal{N}_1^{\text{out}} = \mathcal{V} \setminus \{v_1\}$ and $\mathcal{X}_i[\hat{k}]$ (resp. $\mathcal{Y}_i[\hat{k}]$) includes all the states (resp. estimated auxiliary points). Then, node v_1 performs the distance filtering step (**Line 7**) as

[5]↑The number of agents in this demonstration is not enough to satisfy the robustness condition (Assumption 4.4.4) presented in the next section. However, for our purpose here, it is enough to consider a small number of agents to gain an understanding for each step of the algorithm.

follows. First, it calculates the squared distances $\mathcal{D}_{1j}^2[\hat{k}]$ (since squaring does not alter the order) for all $\boldsymbol{x}_j[\hat{k}] \in \mathcal{X}_i[\hat{k}]$ as in (4.3). Node v_1 has

$$\mathcal{D}_{11}^2[\hat{k}] = 20,\ \mathcal{D}_{12}^2[\hat{k}] = 17,\ \mathcal{D}_{13}^2[\hat{k}] = 18,\ \mathcal{D}_{14}^2[\hat{k}] = 13,$$
$$\mathcal{D}_{15}^2[\hat{k}] = 5,\ \mathcal{D}_{16}^2[\hat{k}] = 17,\ \mathcal{D}_{17}^2[\hat{k}] = 0,\ \mathcal{D}_{18}^2[\hat{k}] = 25.$$

Since $\mathcal{D}_{11}^2[\hat{k}]$ is the second largest, node v_1 discards only node v_8's state (which is the furthest away from v_1's auxiliary point) and $\mathcal{X}_1^{\text{dist}}$ contains all states except $\boldsymbol{x}_8[\hat{k}] = \boldsymbol{x}_{8\to1}[\hat{k}]$.

Then node v_1 performs the min-max filtering process (**Line 8**) as follows. First, consider the first component of the states in $\mathcal{X}_1^{\text{dist}}$. The states of nodes v_1 and v_2 contain the highest value in the first component (which is 4). Since the tie can be broken arbitrarily, we choose $x_1^{(1)}[\hat{k}]$ to come first followed by $x_2^{(1)}[\hat{k}]$ in the ordering, so none of these values are discarded. On the other hand, the state of node v_7 contains the lowest value in its first component, while node v_6's state contains the second lowest value in that component (since node v_8 has already been discarded by the distance filtering process). Node v_1 thus sets $\mathcal{V}_1^{\text{remove}}(1)[\hat{k}] = \{v_6, v_7\}$. Next, consider the second component in which the states of v_6 and v_3 contain the highest and second highest values, respectively, and the states of v_7 and v_5 contain the lowest and second lowest values, respectively. Thus, node v_1 sets $\mathcal{V}_1^{\text{remove}}(2)[\hat{k}] = \{v_3, v_5, v_6, v_7\}$. Since node v_1 removes the entire state from all the nodes in both $\mathcal{V}_1^{\text{remove}}(1)[\hat{k}]$ and $\mathcal{V}_1^{\text{remove}}(2)[\hat{k}]$, according to equation (4.4), we have $\mathcal{X}_1^{\text{mm}}[\hat{k}] = \left\{\boldsymbol{x}_1[\hat{k}], \boldsymbol{x}_2[\hat{k}], \boldsymbol{x}_4[\hat{k}]\right\} = \left\{[4\ \ 2]^T, [4\ \ 1]^T, [3\ \ 2]^T\right\}$.

Next, node v_1 performs the weighted average step (**Line 9**) as follows, Suppose node v_1 assigns the weights $w_{x,11}[\hat{k}] = 0.5$, $w_{x,12}[\hat{k}] = 0.25$ and $w_{x,14}[\hat{k}] = 0.25$. Node v_1 calculates the weighted average according to (4.5) yielding $z_1^{(1)}[\hat{k}] = 3.75$ and $z_1^{(2)}[\hat{k}] = 1.75$. In the gradient step (**Line 10**), suppose $\eta[\hat{k}] = 0.1$. Node v_1 calculates the gradient of its local function f_1 at $\boldsymbol{z}_1[\hat{k}]$ which yields $\boldsymbol{g}_1[\hat{k}] = [9.5\ \ 1.5]^T$ and then calculates the state $\boldsymbol{x}_1[\hat{k}+1]$ as described in (4.6) which yields $\boldsymbol{x}_1[\hat{k}+1] = [2.8\ \ 1.6]^T$.

Next, we consider the estimated auxiliary point update of node v_1. In fact, we can perform the update (**Line 11** and **Line 12**) for each component separately. First, consider the first component in which v_8 and v_7 contain the largest and second largest values, respectively, and v_3 and v_2 contain the smallest and second smallest values, respectively. Node v_1 removes these

values and thus, $\mathcal{Y}_1^{\text{mm}}[\hat{k}](1) = \{y_1^{(1)}[\hat{k}], y_4^{(1)}[\hat{k}], y_5^{(1)}[\hat{k}], y_6^{(1)}[\hat{k}]\} = \{0, -1, 0, 1\}$. Suppose node v_1 assigns the weights $w_{y,11}^{(1)}[\hat{k}] = w_{y,14}^{(1)}[\hat{k}] = w_{y,15}^{(1)}[\hat{k}] = w_{y,16}^{(1)}[\hat{k}] = 0.25$. Then, the weighted average of the first component according to (4.7) becomes $y_1^{(1)}[\hat{k}+1] = 0$. Finally, for the second component, v_6 and v_7 contain the largest values, and v_2 and v_1 contain the smallest and second smallest values, respectively. Node v_1 removes the value obtained from v_2, v_6 and v_7 and thus, the set $\mathcal{Y}_1^{\text{mm}}[\hat{k}](2) = \{y_1^{(2)}[\hat{k}], y_3^{(2)}[\hat{k}], y_4^{(2)}[\hat{k}], y_5^{(2)}[\hat{k}], y_8^{(2)}[\hat{k}]\} = \{0, 1, 1, 2, 2\}$. Suppose node v_1 assigns the weights to each value in $\mathcal{Y}_1^{\text{mm}}[\hat{k}](2)$ equally. The weighted average of the second component becomes $y_1^{(2)}[\hat{k}+1] = 1.2$. Thus, we have $\boldsymbol{y}_1[\hat{k}+1] = [0 \;\; 1.2]^T$.

4.4 Assumptions and Main Results

Having defined the steps in Algorithms 2 and 3, we now turn to proving their resilience and convergence properties.

4.4.1 Assumptions

Assumption 4.4.1. For all $v_i \in \mathcal{V}$, the functions $f_i(\boldsymbol{x})$ are convex, and the sets $\arg\min f_i(\boldsymbol{x})$ are non-empty and bounded.

Since the set $\arg\min f_i(\boldsymbol{x})$ is non-empty, let $\boldsymbol{x}_i^*$ be an arbitrary minimizer of the function f_i.

Assumption 4.4.2. There exists $L \in \mathbb{R}_{>0}$ such that $\|\tilde{\boldsymbol{g}}_i(\boldsymbol{x})\|_2 \leq L$ for all $\boldsymbol{x} \in \mathbb{R}^d$, $v_i \in \mathcal{V}$, and $\tilde{\boldsymbol{g}}_i(\boldsymbol{x}) \in \partial f_i(\boldsymbol{x})$.

The bounded subgradient assumption above is common in the distributed convex optimization literature [70]–[72].

Assumption 4.4.3. The step-size sequence $\{\eta[k]\}_{k=0}^{\infty} \subset \mathbb{R}_{>0}$ used in **Line 11** of Algorithm 2 is of the form

$$\eta[k] = \frac{c_1}{k + c_2} \quad \text{for some} \;\; c_1, c_2 \in \mathbb{R}_{>0}. \tag{4.8}$$

Note that the step-size in (4.8) satisfies $\eta[k+1] < \eta[k]$ for all $k \in \mathbb{N}$, and

$$\lim_{k\to\infty} \eta[k] = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} \eta[k] = \infty \tag{4.9}$$

for any choices of c_1, $c_2 \in \mathbb{R}_{>0}$.

Assumption 4.4.4. Given a positive integer $F \in \mathbb{Z}_+$, the Byzantine agents form a F-local set.

Assumption 4.4.5. For all $k \in \mathbb{N}$ and $\ell \in \{1, 2, \ldots, d\}$, the weights $w_{x,ij}[k]$ and $w_{y,ij}^{(\ell)}[k]$ (used in **Line 9** and **Line 12** of Algorithm 2) are positive if and only if $\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\mathrm{mm}}[k]$ for Algorithm 2 (and $\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\mathrm{dist}}[k]$ for Algorithm 3) and $y_j^{(\ell)}[k] \in \mathcal{Y}_i^{\mathrm{mm}}[k](\ell)$, respectively. Furthermore, there exists $\omega \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}$ and $\ell \in \{1, 2, \ldots, d\}$, the non-zero weights are lower bounded by ω.

Remark 6. In fact, the parameter F in Assumption 4.4.4 can be an upper bound on the number of local Byzantine agents. In exchange for guarding the system from powerful Byzantine adversaries, we need to provide an upper bound on the number of them. In particular, it is typical that in the design stage, one needs to determine the specifications of the system, e.g. the type and the number of adversaries that the system can tolerate. Hence, the assumption of knowing an upper bound on the number of adversaries is very common in the literature on resilient distributed algorithms, e.g., [43], [44].

4.4.2 Analysis of Auxiliary Point Update

Since the dynamics of the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{R}}$ are independent of the dynamics of the estimated solutions $\{\boldsymbol{x}_i[k]\}_{\mathcal{R}}$, we begin by analyzing the convergence properties of the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{R}}$.

In order to establish this result, we need to define the following scalar quantities. For $k \in \mathbb{N}$ and $\ell \in \{1, 2, \ldots, d\}$, let $M^{(\ell)}[k] := \max_{v_i \in \mathcal{R}} y_i^{(\ell)}[k]$, $m^{(\ell)}[k] := \min_{v_i \in \mathcal{R}} y_i^{(\ell)}[k]$, and $D^{(\ell)}[k] := M^{(\ell)}[k] - m^{(\ell)}[k]$. Define the vector $\boldsymbol{D}[k] := \left[D^{(1)}[k], D^{(2)}[k], \cdots, D^{(d)}[k]\right]^T$.

The proposition below shows that the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{R}}$ converge **exponentially fast** to a single point called $\boldsymbol{y}[\infty]$.

Proposition 4.4.1. *Suppose Assumption 4 hold, the graph $\mathcal{G}$ is $(2F+1)$-robust, and the weights $w_{y,ij}^{(\ell)}[k]$ satisfy Assumption 4.4.5. Suppose the estimated auxiliary points of the regular agents $\{\boldsymbol{y}_i[k]\}_{\mathcal{R}}$ follow the update rule described as **Line 11** and **Line 12** in Al-*

gorithm 2. Then, in both Algorithm 2 and Algorithm 3, there exists $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ *with* $y^{(\ell)}[\infty] \in \left[m^{(\ell)}[k], M^{(\ell)}[k]\right]$ *for all* $k \in \mathbb{N}$ *and* $\ell \in \{1, 2, \ldots, d\}$ *such that for all* $v_i \in \mathcal{R}$*, we have*

$$\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\| < \beta e^{-\alpha k},$$

where $\alpha := \frac{1}{|\mathcal{R}|-1} \log \frac{1}{\gamma} > 0$*,* $\beta := \frac{1}{\gamma}\|\boldsymbol{D}[0]\|$*, and* $\gamma := 1 - \frac{\omega^{|\mathcal{R}|-1}}{2}$*.*

The proof of the above proposition follows by noting that the updates for $\{\boldsymbol{y}_i[k]\}_{\mathcal{R}}$ essentially boil down to a set of d scalar consensus updates (one for each dimension of the vector), Thus, one can directly leverage the proof for scalar consensus (with filtering of extreme values) from [30, Proposition 6.3]. We provide the proof of Proposition 4.4.1 in Appendix B.3.

Recall that $\{\hat{\boldsymbol{x}}_i^*\}_{\mathcal{R}}$ is the set containing the approximate minimizers of the regular nodes' local functions. Let $\boldsymbol{x}$ be a matrix in $\mathbb{R}^{d \times |\mathcal{R}|}$, where each column of $\boldsymbol{x}$ is a different vector from $\{\hat{\boldsymbol{x}}_i^*\}_{\mathcal{R}}$. In addition, let $\overline{\boldsymbol{x}}$ and $\underline{\boldsymbol{x}}$ be the vectors in $\mathbb{R}^d$ defined by $\overline{x}_i = \max_{1 \le j \le |\mathcal{R}|} [\boldsymbol{x}]_{ij}$ and $\underline{x}_i = \min_{1 \le j \le |\mathcal{R}|} [\boldsymbol{x}]_{ij}$, respectively. Since we set $\boldsymbol{y}_i[0] = \hat{\boldsymbol{x}}_i^*$ for all $v_i \in \mathcal{R}$ according to **Line 2** in Algorithm 2, we can write

$$\beta = \frac{1}{\gamma}\|\boldsymbol{D}[0]\| = \frac{1}{\gamma}\|\overline{\boldsymbol{x}} - \underline{\boldsymbol{x}}\|.$$

4.4.3 Convergence to Consensus of States

Having established convergence of the auxiliary points to a common value (for the regular nodes), we now consider the state update and show that the states of all regular nodes $\{\boldsymbol{x}_i[k]\}_{\mathcal{R}}$ asymptotically reach consensus under Algorithm 2. Before stating the main theorem, we provide a result from [30, Lemma 2.3] which is important for proving the main theorem.

Lemma 4.4.6. *Suppose the graph* $\mathcal{G}$ *satisfies Assumption 4.4.4 and is* $((2d+1)F+1)$*-robust. Let* $\mathcal{G}'$ *be a graph obtained by removing* $(2d+1)F$ *or fewer incoming edges from each node in* $\mathcal{G}$*. Then* $\mathcal{G}'$ *is rooted.*

This means that if we have enough redundancy in the network (in this case, captured by the $((2d+1)F+1)$-robustness condition), information from at least one node can still flow to the other nodes in the network even after each regular node discards up to F neighboring states in the distance filtering step (**Line 7**) and up to $2dF$ neighboring states in the min-max filtering step (**Line 8**). This transmissibility of information is a crucial condition for reaching consensus among regular nodes.

Theorem 4.4.7 (Consensus)**.** *Suppose Assumptions 4.4.2, 4.4.3, 4.4.4, and 4.4.5 hold, and the graph $\mathcal{G}$ is $((2d+1)F+1)$-robust. If the regular agents follow Algorithm 2 then for all $v_i, v_j \in \mathcal{R}$, it holds that*

$$\lim_{k\to\infty} \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| = 0.$$

Proof. It is sufficient to show that all regular nodes $v_i \in \mathcal{R}$ reach consensus on each component of their vectors $\boldsymbol{x}_i[k]$ as $k \to \infty$. For all $\ell \in \{1, 2, \ldots, d\}$ and for all $v_i \in \mathcal{R}$, from (4.5) and (4.6), the ℓ-th component of the vector $\boldsymbol{x}_i[k]$ evolves as

$$x_i^{(\ell)}[k+1] = \sum_{\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\mathrm{mm}}[k]} w_{x,ij}[k]\; x_j^{(\ell)}[k] - \eta[k]\; g_i^{(\ell)}[k].$$

From [30, Proposition 5.1], the above equation can be rewritten as

$$x_i^{(\ell)}[k+1] = \sum_{v_j \in (\mathcal{N}_i^{\mathrm{in}} \cap \mathcal{R}) \cup \{v_i\}} \bar{w}_{x,ij}^{(\ell)}[k]\; x_j^{(\ell)}[k] - \eta[k]\; g_i^{(\ell)}[k], \tag{4.10}$$

where $\bar{w}_{x,ii}^{(\ell)}[k] + \sum_{v_j \in \mathcal{N}_i^{\mathrm{in}} \cap \mathcal{R}} \bar{w}_{x,ij}^{(\ell)}[k] = 1$, and $\bar{w}_{x,ii}^{(\ell)}[k] > \omega$ and at least $|\mathcal{N}_i^{\mathrm{in}}| - 2F$ of the other weights are lower bounded by $\frac{\omega}{2}$.

Consider the set $\mathcal{X}_i^{\mathrm{mm}}[k]$ which is obtained by removing at most $F + 2dF$ states received from v_i's neighbors (up to F states removed by the distance filtering process in **line 7**, and up to $2F$ additional states removed by the min-max filtering process on each of the d components in **line 8**). Since the graph is $((2d+1)F+1)$-robust and the Byzantine agents form an F-local set by Assumption 4.4.4, from Lemma 4.4.6, the subgraph consisting of regular nodes will be rooted. Using the fact that the term $\eta[k]\; g_i^{(\ell)}[k]$ asymptotically goes to

zero (by Assumptions 4.4.2 and (4.9)) and equation (4.10), we can proceed as in the proof of [30, Theorem 6.1] to show that

$$\lim_{k\to\infty} |x_i^{(\ell)}[k] - x_j^{(\ell)}[k]| = 0,$$

for all $v_i, v_j \in \mathcal{R}$, which completes the proof. □

Theorem 4.4.7 established consensus of the states of the regular agents, leveraging (and extending) similar analysis for scalar functions from [30], only for Algorithm 2. However, this does not hold for Algorithm 3 since there might exist a regular agent $v_i \in \mathcal{R}$, time-step $k \in \mathbb{N}$ and dimension $\ell \in \{1, 2, \ldots, d\}$ such that an adversarial state $x_s^{(\ell)}[k] \in \{x_j^{(\ell)}[k] \in \mathbb{R} : \boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{dist}}[k], v_j \in \mathcal{A}\}$ cannot be written as a convex combination of $\left\{x_j^{(\ell)}[k] \in \mathbb{R} : v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}\right\}$, and thus we cannot obtain equation (4.10). On the other hand, Proposition 4.4.1 established consensus of the auxiliary points, which will be now used to characterize the convergence *region* of both Algorithm 2 and Algorithm 3.

4.4.4 The Region To Which The States Converge

We now analyze the trajectories of the states of the agents under Algorithm 2 and Algorithm 3. We start with the following result regarding the intermediate state $\boldsymbol{z}_i[k]$ calculated in **Lines 7-9** of Algorithm 2.

Lemma 4.4.8. *Suppose Assumptions 4.4.4 and 4.4.5 hold. Furthermore:*

- *if the regular agents follow Algorithm 2, suppose the graph $\mathcal{G}$ is $((2d+1)F+1)$-robust;*
- *otherwise, if the regular agents follow Algorithm 3, suppose the graph $\mathcal{G}$ is $(2F+1)$-robust.*

For all $k \in \mathbb{N}$ and $v_i \in \mathcal{R}$, if there exists $R_i[k] \in \mathbb{R}_{\geq 0}$ such that $\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq R_i[k]$ for all $v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}$ then $\|\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k]\| \leq R_i[k]$.

Proof. Consider the distance filtering step in **Line 7** of Algorithm 2. Recall the definition of $\mathcal{D}_{ij}[k]$ from (4.3). We will first prove the following claim. For each $k \in \mathbb{N}$ and $v_i \in \mathcal{R}$, there exists $v_r \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}$ such that for all $\boldsymbol{x}_j[k] \in \mathcal{X}_i^{\text{dist}}[k]$,

$$\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq \|\boldsymbol{x}_r[k] - \boldsymbol{y}_i[k]\|,$$

or equivalently, $\mathcal{D}_{ij}[k] \leq \mathcal{D}_{ir}[k]$.

There are two possible cases. First, if the set $\mathcal{X}_i^{\text{dist}}[k]$ contains only regular nodes, we can simply choose $v_r \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}$ to be the node whose state $\boldsymbol{x}_r[k]$ is furthest away from $\boldsymbol{y}_i[k]$. Next, consider the case where $\mathcal{X}_i^{\text{dist}}[k]$ contains the state of one or more Byzantine nodes. Since node $v_i \in \mathcal{R}$ removes the F states from $\mathcal{N}_i^{\text{in}}$ that are furthest away from $\boldsymbol{y}_i[k]$ (**Line 7**), and there are at most F Byzantine nodes in $\mathcal{N}_i^{\text{in}}$, there is at least one regular state removed by node v_i. Let v_r be one of the regular nodes whose state is removed. We then have $\mathcal{D}_{ir}[k] \geq \mathcal{D}_{ij}[k]$, for all $v_j \in \{v_s \in \mathcal{V} : \boldsymbol{x}_s[k] \in \mathcal{X}_i^{\text{dist}}[k]\}$ which proves the claim.

If Algorithm 2 is implemented, let $\hat{\mathcal{X}}_i[k] = \mathcal{X}_i^{\text{mm}}[k]$ and we have that $\mathcal{X}_i^{\text{mm}}[k] \subseteq \mathcal{X}_i^{\text{dist}}[k]$ due to the min-max filtering step in **Line 8**. If Algorithm 3 is implemented, let $\hat{\mathcal{X}}_i[k] = \mathcal{X}_i^{\text{dist}}[k]$ since **Line 8** is removed. Then, consider the weighted average step in **Line 9**. From (4.5), we have

$$\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k] = \sum_{\boldsymbol{x}_j[k] \in \hat{\mathcal{X}}_i[k]} w_{x,ij}[k] \left(\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\right).$$

Since $\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq \|\boldsymbol{x}_r[k] - \boldsymbol{y}_i[k]\|$ for all $\boldsymbol{x}_j[k] \in \hat{\mathcal{X}}_i[k]$ (where v_r is the node identified in the claim at the start of the proof), we obtain

$$\begin{aligned}\|\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k]\| &\leq \sum_{\boldsymbol{x}_j[k] \in \hat{\mathcal{X}}_i[k]} w_{x,ij}[k] \, \|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \\ &\leq \sum_{\boldsymbol{x}_j[k] \in \hat{\mathcal{X}}_i[k]} w_{x,ij}[k] \, \|\boldsymbol{x}_r[k] - \boldsymbol{y}_i[k]\|.\end{aligned}$$

Since $v_r \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}$, by our assumption, we have $\|\boldsymbol{x}_r[k] - \boldsymbol{y}_i[k]\| \leq R_i[k]$. Thus, using the above inequality and Assumption 4.4.5, we obtain that $\|\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k]\| \leq R_i[k]$. □

Lemma 4.4.8 essentially states that if the set of states $\{\boldsymbol{x}_j[k] : v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{R}) \cup \{v_i\}\}$ is a subset of the local ball $\mathcal{B}(\boldsymbol{y}_i[k], R_i[k])$ then the intermediate state $\boldsymbol{z}_i[k]$ is still in the ball. This is a consequence of using the distance filter (and adding the min-max filter in Algorithm 2 does not destroy this property), and this will play an important role in proving the convergence theorem.

Next, we will establish certain quantities that will be useful for our analysis of the convergence region. For $v_i \in \mathcal{R}$ and $\epsilon > 0$, define

$$\mathcal{C}_i(\epsilon) := \{\boldsymbol{x} \in \mathbb{R}^d : f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{x}_i^*) + \epsilon\}. \tag{4.11}$$

For all $v_i \in \mathcal{R}$, since the set $\arg\min f_i(\boldsymbol{x})$ is bounded (by Assumption 4.4.1), there exists $\delta_i(\epsilon) \in (0, \infty)$ such that

$$\mathcal{C}_i(\epsilon) \subseteq \mathcal{B}(\boldsymbol{x}_i^*, \delta_i(\epsilon)). \tag{4.12}$$

The following proposition, whose proof is provided in the supplementary material, introduces an angle θ_i which is an upper bound on the angle between the negative of the gradient of f_i at a given point $\boldsymbol{x}$ and the vector $\boldsymbol{x}_i^* - \boldsymbol{x}$.

Proposition 4.4.2. *If Assumptions 4.4.1 and 4.4.2 hold then for all $v_i \in \mathcal{R}$ and $\epsilon > 0$, there exists $\theta_i(\epsilon) \in \left[0, \frac{\pi}{2}\right)$ such that for all $\boldsymbol{x} \notin \mathcal{C}_i(\epsilon)$ and $\tilde{\boldsymbol{g}}_i(\boldsymbol{x}) \in \partial f_i(\boldsymbol{x})$,*

$$\angle(-\tilde{\boldsymbol{g}}_i(\boldsymbol{x}),\ \boldsymbol{x}_i^* - \boldsymbol{x}) \leq \theta_i(\epsilon). \tag{4.13}$$

Before stating the main theorem, we define

$$\tilde{R}_i := \|\boldsymbol{x}_i^* - \boldsymbol{y}[\infty]\|. \tag{4.14}$$

Furthermore, for all $\xi \in \mathbb{R}_{\geq 0}$ and $\epsilon \in \mathbb{R}_{>0}$, we define the *convergence radius*

$$s^*(\xi, \epsilon) := \max_{v_i \in \mathcal{R}} \Big\{ \max\{\tilde{R}_i \sec\theta_i(\epsilon), \tilde{R}_i + \delta_i(\epsilon)\} \Big\} + \xi. \tag{4.15}$$

where $\tilde{R}_i$, $\theta_i(\epsilon)$ and $\delta_i(\epsilon)$ are defined in (4.14), (4.13) and (4.12), respectively. Based on the definition above, we refer to $\mathcal{B}(\boldsymbol{y}[\infty], s^*(\xi, \epsilon))$ as the *convergence ball*.

We now come to the main result of this chapter, showing that the states of all the regular nodes will converge to a ball of radius $\inf_{\epsilon>0} s^*(0,\epsilon)$ around the auxiliary point $\boldsymbol{y}[\infty]$ under Algorithm 2 and Algorithm 3.

Theorem 4.4.9 (Convergence)**.** *Suppose Assumptions 4.4.1-4.4.5 hold. Furthermore:*

- *if the regular agents follow Algorithm 2, suppose the graph $\mathcal{G}$ is $((2d+1)F+1)$-robust;*
- *otherwise, if the regular agents follow Algorithm 3, suppose the graph $\mathcal{G}$ is $(2F+1)$-robust.*

Then regardless of the actions of any F-local set of Byzantine adversaries, for all $v_i \in \mathcal{R}$, we have

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq \inf_{\epsilon>0} s^*(0,\epsilon).$$

The proof of the theorem requires several technical lemmas and propositions, and thus, we provide a proof sketch in Section 4.4.5 and a formal proof in the supplementary material.

The following theorem provides possible locations of the true minimizer $\boldsymbol{x}^*$, which is in fact inside the convergence region, even in the presence of adversarial agents.

Theorem 4.4.10. *Let $\boldsymbol{x}^*$ be a solution of Problem (4.2). If Assumptions 4.4.1 and 4.4.2 hold, then $\boldsymbol{x}^* \in \mathcal{B}\Big(\boldsymbol{y}[\infty],\ \inf_{\epsilon>0} s^*(0,\epsilon)\Big)$.*

Proof. We will show that the summation of any subgradients of the regular nodes' functions at any point outside the region $\mathcal{B}\Big(\boldsymbol{y}[\infty],\ \inf_{\epsilon>0} s^*(0,\epsilon)\Big)$ cannot be zero.

Let $\boldsymbol{x}_0$ be a point outside $\mathcal{B}\Big(\boldsymbol{y}[\infty],\ \inf_{\epsilon>0} s^*(0,\epsilon)\Big)$. Since $\|\boldsymbol{x}_0 - \boldsymbol{y}[\infty]\| > \max_{v_i\in\mathcal{R}}\{\tilde{R}_i + \delta_i(\epsilon)\}$ for some $\epsilon > 0$, we have that $\boldsymbol{x}_0 \notin \mathcal{C}_i(\epsilon)$ for all $v_i \in \mathcal{R}$. By the definition of $\mathcal{C}_i(\epsilon)$ in (4.11), we have $f_i(\boldsymbol{x}_0) > f_i(\boldsymbol{x}_i^*) + \epsilon$ for all $v_i \in \mathcal{R}$. Since the functions f_i are convex, we obtain $\boldsymbol{g}_i(\boldsymbol{x}_0) \neq \boldsymbol{0}$ for all $v_i \in \mathcal{R}$ where $\boldsymbol{g}_i(\boldsymbol{x}_0) \in \partial f_i(\boldsymbol{x}_0)$.

Consider the angle between the vectors $\boldsymbol{x}_0 - \boldsymbol{x}_i^*$ and $\boldsymbol{x}_0 - \boldsymbol{y}[\infty]$. If $\tilde{R}_i = 0$, from (4.14), we have $\boldsymbol{x}_i^* = \boldsymbol{y}[\infty]$ which implies that $\angle(\boldsymbol{x}_0 - \boldsymbol{x}_i^*,\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) = 0$. Suppose $\tilde{R}_i > 0$. Using Lemma B.1.1, we can bound the angle as follows:

$$\angle(\boldsymbol{x}_0 - \boldsymbol{x}_i^*,\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) \leq \arcsin\left(\frac{\tilde{R}_i}{\|\boldsymbol{x}_0 - \boldsymbol{y}[\infty]\|}\right).$$

Since $\|\boldsymbol{x}_0 - \boldsymbol{y}[\infty]\| > \max_{v_i \in \mathcal{R}}\{\tilde{R}_i \sec\theta_i(\epsilon)\}$ for some $\epsilon > 0$ and $\arcsin(x)$ is an increasing function in $x \in [-1, 1]$, we have

$$\angle(\boldsymbol{x}_0 - \boldsymbol{x}_i^*,\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) < \arcsin\left(\frac{\tilde{R}_i}{\tilde{R}_i \sec\theta_i(\epsilon)}\right)$$

and that $\arcsin(\cos\theta_i(\epsilon)) = \frac{\pi}{2} - \theta_i(\epsilon)$. Using Proposition 4.4.2 and the inequality above, we can bound the angle between the vectors $\boldsymbol{g}_i(\boldsymbol{x}_0)$ and $\boldsymbol{x}_0 - \boldsymbol{y}[\infty]$ as follows:

$$\begin{aligned}\angle(\boldsymbol{g}_i(\boldsymbol{x}_0),\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) &\leq \angle(\boldsymbol{g}_i(\boldsymbol{x}_0),\ \boldsymbol{x}_0 - \boldsymbol{x}_i^*) + \angle(\boldsymbol{x}_0 - \boldsymbol{x}_i^*,\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) \\ &< \theta_i(\epsilon) + \left(\frac{\pi}{2} - \theta_i(\epsilon)\right) = \frac{\pi}{2}.\end{aligned}$$

Note that the first inequality is obtained from [57, Corollary 12]. Let $\boldsymbol{u} = \frac{\boldsymbol{x}_0 - \boldsymbol{y}[\infty]}{\|\boldsymbol{x}_0 - \boldsymbol{y}[\infty]\|}$. Compute the inner product

$$\left\langle \sum_{v_i \in \mathcal{R}} \boldsymbol{g}_i(\boldsymbol{x}_0),\ \boldsymbol{u} \right\rangle = \sum_{v_i \in \mathcal{R}} \|\boldsymbol{g}_i(\boldsymbol{x}_0)\| \cos\angle(\boldsymbol{g}_i(\boldsymbol{x}_0),\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]).$$

The RHS of the above equation is strictly greater than zero since $\cos\angle(\boldsymbol{g}_i(\boldsymbol{x}_0),\ \boldsymbol{x}_0 - \boldsymbol{y}[\infty]) > 0$ and $\|\boldsymbol{g}_i(\boldsymbol{x}_0)\| > 0$ for all $v_i \in \mathcal{R}$. This implies that $\sum_{v_i \in \mathcal{R}} \boldsymbol{g}_i(\boldsymbol{x}_0) \neq \boldsymbol{0}$. Since we can arbitrarily choose $\boldsymbol{g}_i(\boldsymbol{x}_0)$ from the set $\partial f_i(\boldsymbol{x}_0)$, we have $\boldsymbol{0} \notin \partial f(\boldsymbol{x}_0)$ where $f(\boldsymbol{x}) = \frac{1}{|\mathcal{R}|}\sum_{v_i \in \mathcal{R}} f_i(\boldsymbol{x})$. □

Theorem 4.4.9 and Theorem 4.4.10 show that both Algorithms 2 and 3 cause all regular nodes to converge to a region that also contains the true solution, regardless of the actions of any F-local set of Byzantine adversaries. The size of this region scales with the quantity $\inf_{\epsilon > 0} s^*(0, \epsilon)$. Loosely speaking, this quantity becomes smaller as the minimizers of the local functions of the regular agents get closer together. More specifically, consider a fixed

$\epsilon \in \mathbb{R}_{>0}$. If the functions $f_i(\boldsymbol{x})$ are translated so that the minimizers $\boldsymbol{x}_i^*$ get closer together (i.e., $\tilde{R}_i$ is smaller while $\theta_i(\epsilon)$ and $\delta_i(\epsilon)$ are fixed), then $s^*(0,\epsilon)$ also decreases. Consequently, the state $\boldsymbol{x}_i[k]$ is guaranteed to become closer to the true minimizer $\boldsymbol{x}^*$ as k goes to infinity. Figure 4.1 illustrates the key quantities appearing in the main theorems.

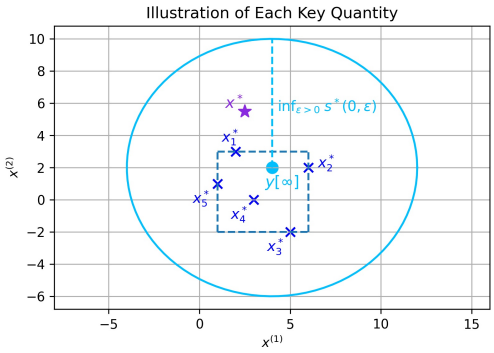


Figure 4.1. The local minimizers $\boldsymbol{x}_i^*$ and the global minimizer $\boldsymbol{x}^*$ are shown in the plot. The estimated auxiliary point $\boldsymbol{y}[\infty]$ is in the rectangle formed by the local minimizers (Proposition 4.4.1) whereas the global minimizer $\boldsymbol{x}^*$ is not necessarily in the rectangle [55]. However, the ball centered at $\boldsymbol{y}[\infty]$ with radius $\inf_{\epsilon>0} s^*(0,\epsilon)$ contains both the supremum limit of the state vectors $\boldsymbol{x}_i[k]$ and the global minimizer $\boldsymbol{x}^*$ (Theorem 4.4.9 and 4.4.10).

4.4.5 Proof Sketch of the Convergence Theorem

We work towards the proof of Theorem 4.4.9 in several steps, which we provide an overview below. The proofs of the intermediate results presented in this section are provided in the supplementary material.

For the subsequent analysis, we suppose that the graph $\mathcal{G}$

- is $((2d+1)F+1)$-robust for Algorithm 2, and
- is $(2F+1)$-robust for Algorithm 3.

Furthermore, unless stated otherwise, we will fix $\xi \in \mathbb{R}_{>0}$ and $\epsilon \in \mathbb{R}_{>0}$, and hide the dependence of ξ and ϵ in $\delta_i(\epsilon)$ and $s^*(\xi,\epsilon)$ by denoting them as δ_i and s^*, respectively.

Gradient Update Step Analysis

First, we consider the update from the intermediate states $\{\boldsymbol{z}_i[k]\}_{\mathcal{R}}$ to the states $\{\boldsymbol{x}_i[k+1]\}_{\mathcal{R}}$ via the gradient step (4.6) (i.e., **Line 10**). In particular, we provide a relationship between $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|$ and $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|$ for three different cases:

- $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \in \Big[0,\ \max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\}\Big]$,
- $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \in \Big(\max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\},\ s^*\Big]$,
- $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \in (s^*, \infty)$.

The corresponding formal statements are presented as follows. Lemma 4.4.11 below essentially says that if k is sufficiently large and $\boldsymbol{z}_i[k] \in \mathcal{B}(\boldsymbol{y}[\infty], \max_{v_i \in \mathcal{R}}\{\tilde{R}_i + \delta_i\})$, then after applying the gradient update (4.6), the state $\boldsymbol{x}_i[k+1]$ will still be in the convergence ball. To establish the result, let $k_1^* \in \mathbb{N}$ be a time-step such that $\eta[k_1^*] \leq \frac{\xi}{L}$.

Lemma 4.4.11. *Suppose Assumptions 4.4.2-4.4.5 hold. For all $v_i \in \mathcal{R}$ and $k \geq k_1^*$, if $\boldsymbol{z}_i[k] \in \mathcal{B}\Big(\boldsymbol{y}[\infty],\ \max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\}\Big)$ then $\boldsymbol{x}_i[k+1] \in \mathcal{B}(\boldsymbol{y}[\infty], s^*)$.*

Lemma 4.4.12, based on Proposition 4.4.2, analyzes the relationship between $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|$ and $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|$ when $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| > \tilde{R}_i + \delta_i$. The result will be used to prove Lemma 4.4.13.

For $v_i \in \mathcal{R}$, define $\Delta_i : [\tilde{R}_i, \infty) \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ to be the function

$$\Delta_i(p, l) := 2l\Big(\sqrt{p^2 - \tilde{R}_i^2}\cos\theta_i - \tilde{R}_i \sin\theta_i\Big) - l^2. \tag{4.16}$$

Lemma 4.4.12. *Suppose Assumptions 4.4.1, 4.4.2, 4.4.4 and 4.4.5 hold. For all $v_i \in \mathcal{R}$ and $k \in \mathbb{N}$, if $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| > \tilde{R}_i + \delta_i$ then*

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 \leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|^2 - \Delta_i(\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|), \tag{4.17}$$

where $\boldsymbol{g}_i[k] \in \mathbb{R}^d$ is defined in (4.6).

Similar to Lemma 4.4.11, Lemma 4.4.13 below states that if k is sufficiently large and $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \in \left(\max_{v_i \in \mathcal{R}}\{\tilde{R}_i + \delta_i\},\ s^*\right]$ then by applying the gradient step (4.6), we have that the state $\boldsymbol{x}_i[k+1]$ is still in the convergence ball.

To simplify the notations, define

$$\begin{aligned} a_i^{\pm} &:= -\tilde{R}_i \sin\theta_i \pm \sqrt{(s^*)^2 - \tilde{R}_i^2 \cos^2\theta_i} \quad \text{and} \\ b_i &:= 2\Big(\sqrt{(s^*)^2 - \tilde{R}_i^2}\cos\theta_i - \tilde{R}_i \sin\theta_i\Big). \end{aligned} \tag{4.18}$$

Let $k_2^* \in \mathbb{N}$ be a time-step such that $\eta[k_2^*] \leq \frac{1}{L}\min_{v_i \in \mathcal{R}}\Big\{\min\{a_i^+,\ b_i\}\Big\}$.

Lemma 4.4.13. *Suppose Assumptions 4.4.1-4.4.5 hold. For all $v_i \in \mathcal{R}$ and $k \geq k_2^*$, if $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \in \left(\max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\},\ s^*\right]$ then $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \in [0,\ s^*]$.*

The following lemma is useful for bounding the term Δ_i appeared in (4.17) for the case that $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| > s^*$.

Define the set of agents

$$\mathcal{I}_z[k] := \{v_i \in \mathcal{R} : \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| > s^*\}, \tag{4.19}$$

and let $k_3^* \in \mathbb{N}$ be a time-step such that $\eta[k_3^*] \leq \frac{1}{2L}\min_{v_i \in \mathcal{R}} b_i$.

Lemma 4.4.14. *If Assumptions 4.4.1-4.4.5 hold then for all $k \geq k_3^*$ and $v_i \in \mathcal{I}_z[k]$,*

$$\Delta_i(\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|) > \frac{1}{2} b_i \kappa_i \eta[k],$$

where Δ_i and $\boldsymbol{g}_i[k]$ are defined in (4.16) and (4.6), respectively, and $\kappa_i := \frac{\epsilon}{\delta_i(\epsilon)} > 0$.

Lemmas 4.4.11-4.4.14 collectively establish the complete relationship governing the update from $\{\boldsymbol{z}_i[k]\}_{\mathcal{R}}$ to $\{\boldsymbol{x}_i[k+1]\}_{\mathcal{R}}$, which will be used to prove Lemma 4.4.15.

Bounds on States of Regular Agents

Next, we consider the update from the states $\{\boldsymbol{x}_i[k]\}_{\mathcal{R}}$ to the intermediate states $\{\boldsymbol{z}_i[k]\}_{\mathcal{R}}$ via two filtering steps (**Lines 7 and 8**) and the weighted average step (**Line 9**). In particular, utilizing Lemma 4.4.8, we derive the following relationship.

Proposition 4.4.3. *If Assumptions 4.4.4 and 4.4.5 hold, then for all $k \in \mathbb{N}$ and $v_i \in \mathcal{R}$, it holds that*

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq \max_{v_j \in \mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + 2\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|.$$

By combining the above inequality with the relationship between $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|$ and $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|$ from Lemmas 4.4.11-4.4.14, and bounding the second term on the RHS, $\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|$, using Proposition 4.4.1, we obtain a relationship between $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|$ and $\max_{v_j \in \mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\|$. As a result, we can bound the distance $\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\|$ by a particular bounded sequence defined below.

Define the time-step $k_0 \in \mathbb{N}$ as $k_0 := \max_{\ell \in \{1,2,3\}} k_\ell^*$. Recall the definition of α and β from Proposition 4.4.1. Let

$$\phi[k_0] = \max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[0] - \boldsymbol{y}[\infty]\| + 2\beta \sum_{k=0}^{k_0-1} e^{-\alpha k} + L \sum_{k=0}^{k_0-1} \eta[k], \tag{4.20}$$

and define a sequence $\{\phi[k]\}_{k=k_0}^{\infty}$ satisfying the update rule

$$\phi^2[k+1] = \max\left\{(s^*)^2, \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \frac{1}{2}\eta[k] \min_{v_i \in \mathcal{R}} b_i \kappa_i\right\}. \tag{4.21}$$

Lemma 4.4.15. *Suppose Assumptions 4.4.1-4.4.5 hold. For all $k \geq k_0$, it holds that*

$$\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq \phi[k].$$

Furthermore, there exists $\bar{\phi} \in \mathbb{R}_{\geq 0}$ such that for all $k \geq k_0$, the sequence $\phi[k]$ can be uniformly bounded as $\phi[k] < \bar{\phi}$.

Convergence Analysis

Finally, we will utilize the following lemma to further analyze the sequence $\{\phi[k]\}$ defined in (4.21).

Lemma 4.4.16. *Consider a sequence $\{\hat{\eta}[k]\}_{k=0}^{\infty} \subset \mathbb{R}_{\geq 0}$ that satisfies $\sum_{k=0}^{\infty} \hat{\eta}[k] = \infty$. If $\gamma_1 \in \mathbb{R}_{\geq 0}$, $\gamma_2 \in \mathbb{R}_{>0}$ and $\lambda \in (-1, 1)$, then there is no sequence $\{u[k]\}_{k=0}^{\infty} \subset \mathbb{R}_{\geq 0}$ that satisfies the update rule*

$$u^2[k+1] = (u[k] + \gamma_1 \lambda^k)^2 - \gamma_2 \hat{\eta}[k].$$

By employing Lemmas 4.4.15 and 4.4.16, Proposition 4.4.4 demonstrates that any repulsion of the state $\boldsymbol{z}_i[k]$ from the convergence ball $\mathcal{B}(\boldsymbol{y}[\infty], s^*)$ due to inconsistency of the estimates of the auxiliary point (Proposition 4.4.1 and 4.4.3) is compensated by the gradient term pulling the state $\boldsymbol{x}_i[k]$ to the convergence ball. Consequently, the quantity $\phi[k]$ decreases until it does not exceed s^*. In other words, the sequence analysis results in

$$\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq \phi[k] \leq s^* \tag{4.22}$$

for a sufficiently large time-step k. The crucial finite time convergence result is formally stated as follows.

Proposition 4.4.4. *Suppose Assumptions 4.4.1-4.4.5 hold. Then, there exists $K \in \mathbb{N}$ such that for all $v_i \in \mathcal{R}$ and $k \geq K$, we have $\boldsymbol{x}_i[k] \in \mathcal{B}(\boldsymbol{y}[\infty], s^*)$.*

Since all the prior analyses valid for all $\xi \in \mathbb{R}_{>0}$ and $\epsilon \in \mathbb{R}_{>0}$, the convergence result in Theorem 4.4.9 follows from taking $\inf_{\xi>0,\epsilon>0}$ and $\limsup_k$ to (4.22).

4.5 Discussion

4.5.1 Redundancy and Guarantees Trade-off

An appropriate notion of network redundancy is necessary for any Byzantine resilient optimization algorithm [30]; for both Algorithm 2 and Algorithm 3, this is captured by the corresponding robustness conditions in Theorem 4.4.9. In particular, Algorithm 2 requires

the graph to be $((2d+1)F+1)$-robust since it implements two filters (a distance-based filter (**Line 7**) and a min-max filter (**Line 8**)) while Algorithm 3 requires the graph to only be $(2F+1)$-robust as a result of only using the distance-based filter. Since each of these filtering steps discards a set of state vectors, the robustness condition allows the graph to retain some flow of information. Thus, while Algorithm 2 requires significantly stronger conditions on the network topology (i.e., requiring the robustness parameter to scale linearly with the dimension of the functions), it provides the benefit of guaranteeing consensus. Algorithm 3 only requires the robustness parameter to scale with the number of adversaries in each neighborhood, and thus can be used for optimizing high-dimensional functions with relatively sparse networks, at the cost of losing the guarantee on consensus.

Remark 7. In the vector resilient consensus problem [73], [74], it has been established that guaranteeing a non-empty interior for the convex hull, formed by the states of the regular nodes, requires $\Omega(dF)$ neighboring nodes, which aligns with Algorithm 2. Moreover, for the case where $d \geq 4$, the corresponding time complexity has been shown to be exponential in d [74]. However, our algorithms offer the advantage of significantly reducing computational requirements, as will be discussed next.

4.5.2 Time Complexity

Suppose the network is r-robust and the number of in-neighbors $|\mathcal{N}_i^{\text{in}}|$ is linearly proportional to r for all $v_i \in \mathcal{V}$. For the distance-based filter (**Line 7**), each regular agent $v_i \in \mathcal{R}$ computes the L^2-norm between its auxiliary state and in-neighbor states and then finds the agent that attains the maximum value which take $\mathcal{O}(dr)$ operations. On the other hand, for the min-max filter (**Line 8**), each regular agent $v_i \in \mathcal{R}$ is required to sort the in-neighbor states for each dimension which takes $\mathcal{O}(dr \log r)$ operations. For Algorithms 2 and 3, the time complexities for filtering process are $\tilde{\mathcal{O}}(d^2)$ and $\tilde{\mathcal{O}}(d)$, respectively.

4.5.3 Convergence Ball

Consider univariate functions (i.e., one-dimensional space) case. To facilitate the discussion, we denote $\min_{v_i \in \mathcal{R}} x_i^*$ and $\max_{v_i \in \mathcal{R}} x_i^*$ by $\underline{x}$ and $\overline{x}$, respectively. For simplicity, assume

that the local minimizer x_i^* is unique for all $v_i \in \mathcal{R}$ so that δ_i defined in (4.12) can be chosen arbitrary close to zero for all $v_i \in \mathcal{R}$. In this case, we have that for all $v_i \in \mathcal{R}$, θ_i defined in (4.13) is zero. Therefore, the convergence radius s^* in (4.15) simplifies to $\max_{v_i \in \mathcal{R}} \tilde{R}_i$ (where $\tilde{R}_i$ defined in (4.14)). In the best case, we can have $y[\infty] = \frac{1}{2}(\underline{x} + \overline{x})$ which results in the convergence region $[\underline{x}, \overline{x}]$ as derived in [30]. In the worst case, (assuming approximation error ϵ^* in **Line 1** is zero) we can have $y[\infty] = \underline{x}$ or $\overline{x}$ which results in the convergence region $[2\underline{x} - \overline{x}, \overline{x}]$ or $[\underline{x}, 2\overline{x} - \underline{x}]$, respectively. In such worst case, the region is two times bigger than the region derived in [30]. These results are due to our "radius analysis" which is uniform in all directions from $\boldsymbol{y}[\infty]$.

4.5.4 Maximum Tolerance

Based on the robustness condition for each algorithm and a formula from [75], given the number of agents N in the complete graph and the number of dimensions for the optimization variables d, the upper bound on the number of local Byzantine agents F such that the corresponding guarantees still hold, is as follows:

- $F = \left\lfloor \frac{N-1}{2(2d+1)} \right\rfloor$ for Algorithm 2, and
- $F = \left\lfloor \frac{1}{4}(N-1) \right\rfloor$ for Algorithm 3.

From a practical perspective, the robustness property demonstrates a natural trade-off for the system designer. A network that has a stronger robustness property can tolerate more adversaries, but can also induce more costs.

4.5.5 Importance of Main States Computation

If we simply implement a resilient consensus protocol on local minimizers similar to the auxiliary states, $\boldsymbol{y}_i[k]$, computation (in **Lines 11-12**) and remove the main states, $\boldsymbol{x}_i[k]$, computation (in **Lines 7-9**), we would obtain that the states of the regular agents converge to the hyper-rectangle formed by the local minimizers (for resilient component-wise consensus algorithms [76]), or the convex hull of the local minimizers (for resilient vector consensus algorithms [74], [77]). Even though this method works for the single dimension case due to the

convergence to the same set as deploying a resilient distributed optimization algorithm [30], [33], it might not give a desired result for the multi-dimensional case owing to several reasons. First, it is possible that the minimizer of the sum lies outside both the hyper-rectangle and convex hull [55], [56] as shown in Figure 4.1. Second, using only a resilient consensus protocol, one ignores the gradient information which steers the regular agents' states toward the true minimizer. Third, we empirically show in Section 4.6 that implementing a resilient distributed optimization algorithm (especially Algorithm 2) usually gives better results in terms of both optimality gap and distance to the global minimizer.

4.6 Numerical Experiment

We now provide two numerical experiments to illustrate Algorithm 2 and Algorithm 3. In the first experiment, we generate quadratic functions for the local objective functions. Using these functions, we demonstrate the performance (e.g., optimality gaps, distances to the global minimizer) of our algorithms. We also compare the optimality gaps of the function value obtained using the states $\boldsymbol{x}_i[k]$ and the value obtained using the auxiliary points $\boldsymbol{y}_i[k]$, and plot the trajectories of the states of a subset of regular nodes. In the second experiment, we demonstrate the performance of our algorithm on a machine learning task (banknote authentication task). Specifically, we compare the accuracy of the models obtained from our algorithm (resilient distributed model) and that of a centralized model.

4.6.1 Synthetic Quadratic Functions

Preliminary Settings

- **Main Parameters:** We set the number of nodes to be $n = 25$ and the dimension of each function to be $d = 2$.
- **Adversary Parameters:** We consider the F-local model, and set $F = 2$ for Algorithm 2 and $F = 5$ for Algorithm 3.

Network Settings

- **Topology Generation:** We construct an 11-robust graph on $n = 25$ nodes following the approach from [69], [75]. This graph can tolerate up to 2 local adversaries for Algorithm 2, and up to 5 local adversaries for Algorithm 3 according to Theorem 4.4.9. Note that the same graph is used to perform numerical experiments for both Algorithms 2 and 3.

Adversaries' Strategy

- **Adversarial Nodes:** We construct the set of adversarial nodes $\mathcal{A}$ by randomly choosing nodes in $\mathcal{V}$ so that the set of adversarial nodes form a F-local set. Note that in general, constructing $\mathcal{A}$ depends on the topology of the network. In our experiment, we have $\mathcal{A} = \{v_9, v_{16}\}$ for Algorithm 2 and $\mathcal{A} = \{v_5, v_{11}, v_{12}, v_{17}, v_{22}, v_{24}\}$ for Algorithm 3.

- **Adversarial Values Transmitted:** Here, we use a sophisticated approach rather than simply choosing the transmitted values at random. Suppose v_s is an adversary node and v_i is a regular node which is an out-neighbor of v_s, i.e., $v_s \in \mathcal{N}_i^{\text{in}}$. First, consider the state of nodes in the network at time-step k. The adversarial node v_s uses an oracle to determine the region in the state space for the regular node v_i in which if the adversarial node selects the transmitted value to be outside the region then the value will be discarded by that regular agent v_i. Then, v_s chooses $\boldsymbol{x}_{s\to i}[k]$ (the forged state sent from v_s to v_i at time k) so that it is in the safe region and far from the global minimizer. In this way, the adversaries' values will not be discarded and also try to prevent the regular nodes from getting close to the minimizer. Similarly, for the auxiliary point update, the adversarial node v_s uses an oracle to determine the safe region in the auxiliary point's space for the regular node v_i. Since the safe region is a hyper-rectangle in general, v_s chooses $\boldsymbol{y}_{s\to i}[k]$ (the forged estimated auxiliary point sent from v_s to v_i at time k) to be near a corner (chosen randomly) of the hyper-rectangle.

Objective Functions Settings

- **Local Functions:** For $v_i \in \mathcal{V}$, we set the local objective functions $f_i : \mathbb{R}^d \to \mathbb{R}$ to be

$$f_i(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}_i \boldsymbol{x} + \boldsymbol{b}_i^T \boldsymbol{x},$$

where $\boldsymbol{Q}_i \in \mathcal{S}_d^+$ and $\boldsymbol{b}_i \in \mathbb{R}^d$ are chosen randomly. Note that the same local functions are used to perform numerical experiments for both Algorithms 2 and 3.

- **Global Objective Function:** According to our objective (4.2), we then have the global objective function $f : \mathbb{R}^d \to \mathbb{R}$ as follows:

$$f(\boldsymbol{x}) = \frac{1}{|\mathcal{R}|}\Big(\frac{1}{2}\boldsymbol{x}^T\Big(\sum_{v_i \in \mathcal{R}} \boldsymbol{Q}_i\Big)\boldsymbol{x} + \Big(\sum_{v_i \in \mathcal{R}} \boldsymbol{b}_i\Big)^T \boldsymbol{x}\Big),$$

where the set of regular nodes $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$.

Algorithm Settings

- **Initialization:** For each regular node $v_i \in \mathcal{R}$, we compute the exact minimizer $\boldsymbol{x}_i^* = -\boldsymbol{Q}_i^T \boldsymbol{b}_i$ and use it as the initial state and auxiliary point of v_i as suggested in **Line 1-2** of Algorithm 2.

- **Weights Selection:** For each time-step $k \in \mathbb{N}$ and regular node $v_i \in \mathcal{R}$, we randomly choose the weights $w_{x,ij}[k], w_{y,ij}^{(\ell)}[k]$ so that they follow the description of **Line 9** and **Line 12**, and Assumption 4.4.5.

- **Step-size Selection:** We choose the step-size schedule (in **Line 11** of Algorithm 2) to be $\eta[k] = \frac{1}{k+1}$.

- **Gradient Norm Bound:** We choose the upper bound of the gradient norm to be $L = 10^5$. If the norm exceeds the bound, we scale the gradient down so that its norm is equal to L, i.e.,

$$\boldsymbol{g}_i[k] = \begin{cases} \nabla f_i(\boldsymbol{z}_i[k]) & \text{if } \|\nabla f_i(\boldsymbol{z}_i[k])\| \leq L, \\ \frac{L}{\|\nabla f_i(\boldsymbol{z}_i[k])\|} \cdot \nabla f_i(\boldsymbol{z}_i[k]) & \text{otherwise.} \end{cases}$$

Simulation Settings and Results

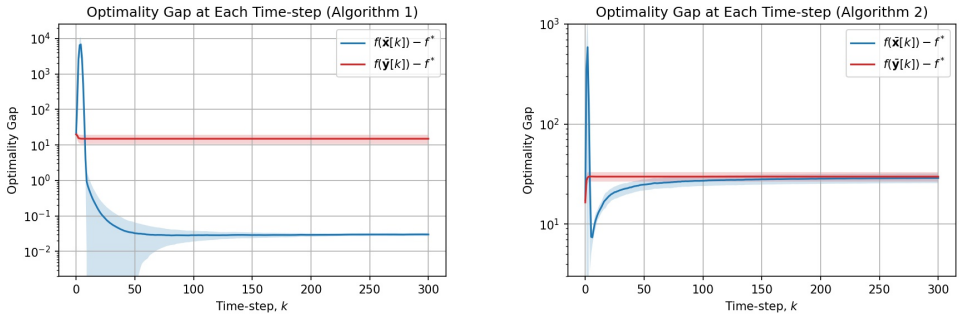
- **Time Horizon:** We set the time horizon of our simulations to be $K = 300$ (starting from $k = 0$).

- **Experiments Detail:** For both Algorithms 2 and 3, we fix the graph, local functions, and step-size schedule. However, since the set of adversaries are different, the global objective functions, and hence the global minimizers are different. For each algorithm, we run the experiment 10 times setting the same states initialization across the runs. The results from the runs are different due to the randomness in the adversaries' strategy.

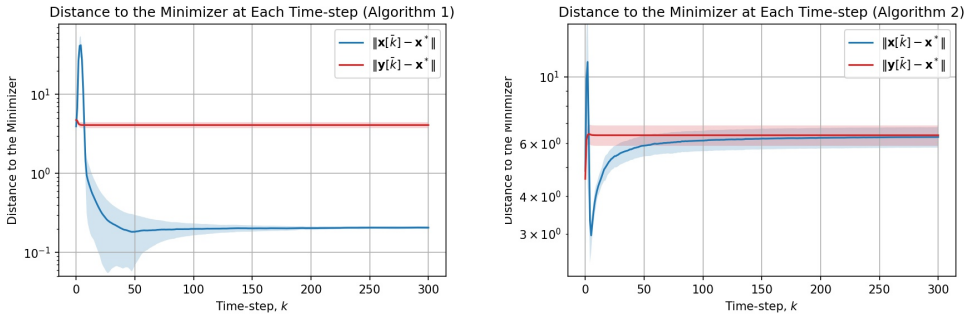
- **Performance Metrics:** We examine the performance of our algorithms by considering the optimality gaps evaluated at different points (Figure 4.2a), distances to the global minimizer evaluated at different points (Figure 4.2b), and trajectories of randomly selected regular agents (Figure 4.2c).

- **Algorithm 2's Results:** The lines corresponding to the optimality gap and distance to the global minimizer evaluated using auxiliary points are almost horizontal since the convergence to consensus is very fast. However, one can see that the optimality gap and distance to the minimizer obtained from the regular states are significantly smaller than that from the auxiliary points due to the use of gradient information (**Line 10**) and extreme states filtering (**Line 8**) in the regular state update. In particular, at $k = 300$, the optimality gap and distance to the global minimizer at the regular states' average are only about 0.030 and 0.206, respectively. Moreover, the state trajectories converge together and stay close to the global minimizer even in the presence of sophisticated adversaries. Note that, from our observations, Algorithm 2 yields better results than Algorithm 3 given the same settings.

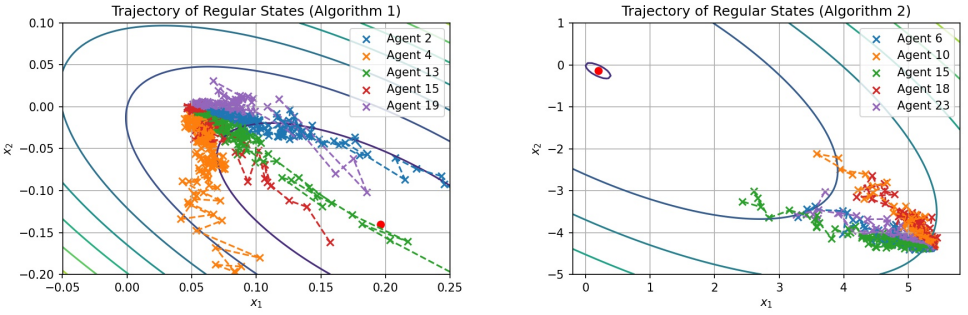
- **Algorithm 3's Results:** The optimality gaps and distances to the global minimizer evaluated using the states are slightly better than the values obtained using the auxiliary points, and the state trajectories remain reasonably close to the global minimizer showing that the algorithm can tolerate $F = 5$ local adversaries (which is more than Algorithm 2). Interestingly, the state trajectories seem to converge together even though the consensus guarantee is lacking due to the absence of the distance-based filter.



(a) The optimality gap evaluated at the average of the regular nodes' states $f(\bar{\boldsymbol{x}}) - f^*$ averaged over 10 runs (blue line), and the optimality gap evaluated at the average of the regular nodes' auxiliary points $f(\bar{\boldsymbol{y}}) - f^*$ averaged over 10 runs (red line).



(b) The distance between the average of the regular nodes' states and the global minimizer $\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|$ averaged over 10 runs (blue line), and the distance between the average of the regular nodes' auxiliary points and the global minimizer $\|\bar{\boldsymbol{y}} - \boldsymbol{x}^*\|$ averaged over 10 runs (red line).



(c) The trajectory of the states of a subset of the regular nodes. Different colors of the trajectory represent different regular agents v_i in the network.

Figure 4.2. The plots show the results obtained from (left) Algorithm 2 and (right) Algorithm 3. In the first four plots, the shaded regions represent +1/-1 standard deviation from the mean. In the last two plots, the contour lines show the level sets of the global objective function (in this case, a quadratic function) and the red dots represent the global minimizer.

4.6.2 Banknote Authentication using Regularized Logistic Regression

Dataset Information[6]

- **Description:** The data were extracted from images that were taken from genuine and forged banknote-like specimens.
- **Data Points:** The total number of data is 1,372.
- **Features:** The dataset consists of four features: (1) the variance of a wavelet transformed image, (2) the skewness of a wavelet transformed image, (3) the curtosis of a wavelet transformed image, and (4) the entropy of an image.
- **Labels:** There are two classes: '0' (genuine) and '1' (counterfeit).

Preliminary Settings

- **Main Parameters:** We set the number of nodes to be $n = 50$. Since there are four features, the dimension of the states is $d = 5$ (one for each feature and the other one for the bias).
- **Adversarial Parameters:** We use the F-local model with $F = 2$.
- **Dataset Partitioning:** We randomly partition the dataset into three chunks: 1,000 training data points, 186 validation data points, and 186 test data points. We then distribute the training dataset to the nodes in the network equally. Thus, each node contains $m = 20$ training data points.

Network and Weights Settings

We construct the network and corresponding weight matrix using the same approach as in the synthetic quadratic functions case.

Adversaries' Strategy

We choose the set of adversarial nodes $\mathcal{A}$ and adversarial values transmission strategy using the same method as in the synthetic quadratic functions case.

Objective Functions Settings

[6]↑https://archive.ics.uci.edu/ml/datasets/banknote+authentication

- **Notations:** Let $\boldsymbol{x}_{ij} \in \mathbb{R}^{d-1}$ be the feature vector of the j-th data points at node $v_i \in \mathcal{V}$, and $Y_{ij} \in \{0, 1\}$ be the corresponding label. We let $\tilde{\boldsymbol{x}}_{ij} = \begin{bmatrix} \boldsymbol{x}_{ij}^T & 1 \end{bmatrix}^T$ to account for the bias term.

- **Local Functions:** Since this is a classification task, we choose the logistic regression model with L_2-regularization in which its loss function is strongly convex. For $v_i \in \mathcal{V}$, we set the local objective functions $f_i : \mathbb{R}^d \to \mathbb{R}$ to be

$$f_i(\boldsymbol{W}) = |\mathcal{R}| \sum_{j=1}^{m} \log\Big(\exp(-Y_{ij}\tilde{\boldsymbol{x}}_{ij}^T \boldsymbol{W}) + 1\Big) + \frac{\varsigma}{2}\|\boldsymbol{W}\|^2,$$

  where the set of regular nodes $\mathcal{R} = \mathcal{V} \setminus \mathcal{A}$ and $\varsigma \in \mathbb{R}_{>0}$ is the regularization parameter which will be chosen later.

- **Global Objective Function:** According to our objective (4.2), we then have the global objective function $f : \mathbb{R}^d \to \mathbb{R}$ as follows:

$$f(\boldsymbol{W}) = \sum_{v_i \in \mathcal{R}} \sum_{j=1}^{m} \log\Big(\exp(-Y_{ij}\tilde{\boldsymbol{x}}_{ij}^T \boldsymbol{W}) + 1\Big) + \frac{\varsigma}{2}\|\boldsymbol{W}\|^2.$$

- **Regularization Parameter Selection:** We consider $\varsigma \in \{10^{-4}, 10^{-3}, \ldots, 10^{5}\}$. We train our (centralized) logistic model using the global objective function above for each value of ς and then we select the value of ς that gives the best validation accuracy.

Algorithm Settings

- **Initialization:** As suggested in **Line 1** of Algorithm 2, we numerically find the minimizer of the local functions using the default optimizer of `sklearn.linear_model.LogisticRegression`. Then, we use the minimizer of each regular node to be the initial state and auxiliary point as in **Line 2**.

The methodology of step-size selection and gradient norm bound is the same as in the synthetic quadratic functions case.

Simulation Settings and Results

- **Benchmark:** We evaluate the performance (accuracy) of the (centralized) logistic model with the selected regularization parameter, ς.

- **Time Horizon:** We set the time horizon of our simulations of our distributed algorithm to be $K = 200$ (starting from $k = 0$).

- **Simulation:** We run the simulations of Algorithm 2 by varying the parameter η_0 from -2 to 4 with increasing step of 1. We evaluate the performance of each model (i.e., each η_0) by considering the accuracy obtained by using the state $\bar{\boldsymbol{W}}[K] = \frac{1}{|\mathcal{R}|} \sum_{v_i \in \mathcal{R}} \boldsymbol{W}_i[K]$ for each η_0 and the validation data. Then, we select the parameter η_0 which provides the best accuracy. Finally, with the selected value of η_0, we evaluate the performance (accuracy) of the corresponding model with the test data.

- **Result:** We repeat the whole process 5 times. In other words, each run uses different realization of data partitioning (hence, different local functions and global function), network topology, and adversaries set. The result of each run is shown in Table 4.1. The first three rows show the adversaries set, regularization parameter and step-size parameter of each run. The next (resp. last) three rows show the training (resp. test) accuracy of the centralized model, distributed model evaluated at $\bar{\boldsymbol{W}}[K] = \frac{1}{|\mathcal{R}|} \sum_{v_i \in \mathcal{R}} \boldsymbol{W}_i[K]$, and the minimum accuracy among the local model of regular nodes evaluated at its own state $\boldsymbol{W}_i[K]$. We can see that despite the presence of adversaries with sophisticated behavior, the performance of our algorithm is just slightly lower than the centralized model's performance for this task.

4.7 Conclusion and Future work

In this chapter, we considered the distributed optimization problem in the presence of Byzantine agents. We developed two resilient distributed optimization algorithms for multi-dimensional functions. The key improvement over our previous work in [40] is that the algorithms proposed in this chapter do not require a fixed auxiliary point to be computed in advance (which will not happen under finite time in general). Our algorithms have low

Table 4.1. Training/Test Accuracy of Centralized (C), Distributed (D) Models and Minimum among Regular Agents' Models (MIN) for each Run of Banknote Authentication Task

	1st Run	2nd Run	3rd Run	4th Run	5th Run
$v_i \in \mathcal{A}$	$9, 23$	$9, 15$	$10, 11$	$5, 29$	$24, 47$
ς	1.0	1.0	10	1.0	1.0
η_0	1	1	2	4	1
Train (C)	99.40	99.20	99.00	98.90	99.10
Train (D)	98.10	97.90	97.70	98.30	98.00
Train (MIN)	97.80	97.60	97.50	98.30	97.70
Test (C)	99.46	98.39	99.46	99.46	98.92
Test (D)	97.85	95.70	98.92	97.85	98.39
Test (MIN)	97.85	95.70	98.92	97.85	97.85

complexity and each regular node only needs local information to execute the steps. Algorithm 2 (with the min-max state filter), which requires more network redundancy, guarantees that the regular states can asymptotically reach consensus and enter a bounded region that contains the global minimizer, irrespective of the actions of Byzantine agents. On the other hand, Algorithm 3 (without the min-max filter) has a more relaxed condition on the network topology and can guarantee asymptotic convergence to the same region, but cannot guarantee consensus. For both algorithms, we explicitly characterized the size of the convergence region, and showed through simulations that Algorithm 2 appears to yield results that are closer to optimal, as compared to Algorithm 3.

As noted earlier, the consensus guarantee for Algorithm 2 comes at the cost of requiring that the robustness of the network scale linearly with the dimension of the local functions, which can be restrictive in practice. This seems to be a common challenge for resilient consensus-based algorithms in systems with multi-dimensional states, e.g., [47], [74], [78]. Finding a relaxed condition on the network topology for high-dimensional resilient distributed optimization problems (with guaranteed consensus) would be a rich area for future research.

5. ON THE GEOMETRIC CONVERGENCE OF BYZANTINE-RESILIENT DISTRIBUTED OPTIMIZATION ALGORITHMS

5.1 Introduction

As discussed in Chapter 1 and Chapter 4, distributed optimization problems pertain to a setting where each node in a network has a local cost function, and the goal is for all agents in the network to agree on a minimizer of the average of the local cost functions. In the distributed optimization literature, there are two main paradigms: client-server and peer-to-peer. Motivated by settings where the client-server paradigm may suffer from a single point of failure or communication bottleneck, there is a growing amount of work on the peer-to-peer setting where the agents in the network are required to send and receive information only from their neighbors. A variety of algorithms have been proposed to solve such problems in peer-to-peer architectures (e.g., see [70], [71], [79]–[81]). The works [82], [29] and [83] summarize the recent advances in the field of (peer-to-peer) distributed optimization.

These aforementioned works typically make the assumption that all agents are trustworthy and cooperative (i.e., they follow the prescribed protocol). However, it has been shown that the regular agents fail to reach an optimal solution even if a single misbehaving (or "Byzantine") agent is present [30], [33]. Thus, designing distributed optimization algorithms that allow all the regular agents' states in the network to stay close to the minimizer of the sum of regular agents' functions regardless of the adversaries' actions has become a prevailing problem. Nevertheless, as discussed in Chapter 4, the number of works addressing Byzantine-resilient algorithms in the peer-to-peer setup is relatively limited when compared to the client-server setting.

In contrast to Chapter 4, which addressed Byzantine-resilient distributed deterministic optimization problems under a general convexity assumption, this chapter explores the problem under a strong convexity assumption. Our contributions are as follows. (i) We introduce an algorithmic framework called RedGraf, a generalization of BRIDGE in [45], which includes some state-of-the-art Byzantine-resilient distributed optimization algorithms

as special cases. (ii) We propose a novel contraction property which we show provides a general method for proving geometric convergence of algorithms in RedGraf. To the best of our knowledge, our work is the first to provide a geometric rate of convergence of all regular agents' states to a ball containing the true minimizer for a class of resilient algorithms under the strong convexity assumption. In addition, we explicitly characterize the convergence rate and the size of the convergence region. (iii) We introduce a novel mixing dynamics property which is used to derive approximate consensus results for algorithms in RedGraf in which both the convergence rate and the final consensus diameter are explicitly characterized. (iv) Using our framework, we analyze the contraction and mixing dynamics properties of some state-of-the-art algorithms, leading to convergence and consensus results for each algorithm. Our work is the first to show that these algorithms satisfy such properties. (v) We demonstrate and compare the performance of the algorithms through numerical simulations to corroborate the theoretical results for convergence and approximate consensus.

5.2 Related Work

The survey paper [84] provides an overview of some Byzantine-resilient algorithms for both the client-server and peer-to-peer paradigms. Since we are focusing on resilient algorithms for peer-to-peer settings, we discuss the following research papers attempting to solve such problems. The papers [33], [30] and [50] show that using the distributed gradient descent (DGD) equipped with a *trimmed mean filter* guarantees convergence to the convex hull of the local minimizers under scalar-valued objective functions. Adopting a similar algorithm, the paper [51] gives the same guarantee for scalar-valued problems under *deception attacks*. The work [52] also considers the scalar version of such problems but relies on *trusted agents* which cannot be compromised by adversarial attacks. To tackle vector-valued problems, the paper [44] proposes ByRDiE, a coordinate descent method for machine learning problems leveraging the algorithm in [33], while the paper [45] presents BRIDGE, an algorithm framework for Byzantine-resilient distributed optimization problems. Even though [44] and [45] show the convergence to the minimizer with high probability (for certain specific algorithms), they require that the training data is i.i.d. across the agents in the network.

While resilient algorithms with the trimmed mean filter are widely used, e.g., [30], [33], [45], [50]–[52], the convergence analysis for multivariate functions under general assumptions is still lacking. The work [48] proposes RSGP, a resilient algorithm based on a Subgradient-Push method [27] equipped with a *maliciousness score* for detecting adversaries. However, the work requires that the regular agents' functions have common statistical characteristics, and does not provide any guarantees on the proposed algorithm. In Chapter 4, we propose Algorithms 2 and 3 (referred in this chapter as SDMMFD and SDFD, respectively), resilient algorithms for deterministic distributed convex optimization problems for multi-dimensional functions, which have an asymptotic convergence guarantee to a proximity of the true minimum. However, the work does not provide the convergence rate for the proposed algorithms. The work [47] provides an algorithm with provable *exact fault-tolerance*, but relies on redundancy among the local functions, and also requires the underlying communication network to be complete.

For distributed stochastic optimization problems, the paper [85] introduces a resilient algorithm based on a total variation norm penalty motivated from [86]. The recent paper [49] also considers stochastic problems, and proposes an algorithm utilizing a distance-based filter and objective value-based filter, but does not provide any performance guarantees. The recent paper [87] which also considers stochastic problems especially for machine learning, proposes a validation-based algorithm for both i.i.d. and non-i.i.d. settings. In particular, the work theoretically shows a convergence guarantee for the proposed algorithm under convex loss functions and i.i.d. data. The recent papers [68] and [88] propose algorithms which converge to a neighborhood of a stationary point for distributed stochastic non-convex optimization problems.

As described earlier in our contributions section, our work in this chapter addresses the gaps in the existing literature by showing geometric rate of convergence of all regular agents' states to a ball containing the true minimizer for a class of resilient algorithms under the strong convexity assumption, and explicitly characterizing the size of the ball. In particular, our work provides the convergence analysis of resilient algorithms with the trimmed mean filter studied in [30], [33], [45], [50]–[52] under mild assumptions.

5.3 Background

Let $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ denote the set of natural numbers (including zero), integers, and real numbers, respectively. Let $\mathbb{Z}_+$, $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{>0}$ denote the set of positive integers, non-negative real numbers, and positive real numbers, respectively. For convenience, for an integer $N \in \mathbb{Z}_+$, we define $[N] := \{1, 2, \ldots, N\}$. The cardinality of a set is denoted by $|\cdot|$. Given positive integers $F \in \mathbb{Z}_+$ and $s \geq F$, and a set of scalars $\mathcal{X} = \{x_1, x_2, \ldots, x_s\}$, define $M_F(\mathcal{X})$ and $m_F(\mathcal{X})$ to be the F-th largest element and F-th smallest element, respectively, of the set $\mathcal{X}$.

5.3.1 Linear Algebra

Vectors are taken to be column vectors, unless otherwise noted. We use $x^{(\ell)}$ to represent the ℓ-th component of a vector $\boldsymbol{x}$. The Euclidean norm on $\mathbb{R}^d$ is denoted by $\|\cdot\|$. We use $\mathbf{1}$ and $\boldsymbol{I}$ to denote the vector of all ones and the identity matrix, respectively, with appropriate dimensions. We denote by $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ the Euclidean inner product of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^T \boldsymbol{v}$ and by $\angle(\boldsymbol{u}, \boldsymbol{v})$ the angle between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, i.e., $\angle(\boldsymbol{u}, \boldsymbol{v}) = \arccos\left(\frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right)$. The Euclidean ball in $\mathbb{R}^d$ with center at $\boldsymbol{x}_0 \in \mathbb{R}^d$ and radius $r \in \mathbb{R}_{\geq 0}$ is denoted by $\mathcal{B}(\boldsymbol{x}_0, r) := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}$. For $N \in \mathbb{Z}_+$, a matrix $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ is (row-)stochastic if $\boldsymbol{W}\mathbf{1} = \mathbf{1}$ and $w_{ij} \geq 0$ for all $i, j \in [N]$. For $N \in \mathbb{Z}_+$, we use $\mathbb{S}^N$ to denote the set of all $N \times N$ (row-)stochastic matrices.

5.3.2 Functions Properties

For a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ and a point $\boldsymbol{x} \in \mathbb{R}^d$, define the vector $\nabla f(\boldsymbol{x}) \in \mathbb{R}^d$ to be the gradient of f at point $\boldsymbol{x}$.

Definition 5.3.1 (strongly convex function)**.** *Given a non-negative real number $\mu \in \mathbb{R}_{\geq 0}$ and differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, f is μ-strongly convex if for all $\boldsymbol{x}_1$, $\boldsymbol{x}_2 \in \mathbb{R}^d$,*

$$f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2) + \langle \nabla f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle + \frac{\mu}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{5.1}$$

Note that a differentiable function is convex if it is 0-strongly convex.

Definition 5.3.2 (Lipschitz gradient)**.** *Given a non-negative real number $L \in \mathbb{R}_{\geq 0}$ and differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, f has an L-Lipschitz gradient if for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$,*

$$\|\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2)\| \leq L\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|. \tag{5.2}$$

5.3.3 Graph Theory

We denote a network by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which consists of the set of N nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ and the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. If $(v_i, v_j) \in \mathcal{E}$, then node v_j can receive information from node v_i. The in-neighbor and out-neighbor sets are denoted by $\mathcal{N}_i^{\text{in}} = \{v_j \in \mathcal{V} : \ (v_j, v_i) \in \mathcal{E}\}$ and $\mathcal{N}_i^{\text{out}} = \{v_j \in \mathcal{V} : \ (v_i, v_j) \in \mathcal{E}\}$, respectively. A path from node $v_i \in \mathcal{V}$ to node $v_j \in \mathcal{V}$ is a sequence of nodes $v_{k_1}, v_{k_2}, \ldots, v_{k_l}$ such that $v_{k_1} = v_i$, $v_{k_l} = v_j$ and $(v_{k_r}, v_{k_{r+1}}) \in \mathcal{E}$ for $r \in [l-1]$. Throughout this chapter, the terms nodes and agents will be used interchangeably. Given a set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, where each $\boldsymbol{x}_i \in \mathbb{R}^d$, we use the following shorthand notation for all $\mathcal{S} \subseteq \mathcal{V}$: $\{\boldsymbol{x}_i\}_{\mathcal{S}} = \{\boldsymbol{x}_i \in \mathbb{R}^d : v_i \in \mathcal{S}\}$.

Definition 5.3.3. *A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be rooted at $v_i \in \mathcal{V}$ if for all $v_j \in \mathcal{V} \setminus \{v_i\}$, there is a path from v_i to v_j. A graph is said to be rooted if it is rooted at some $v_i \in \mathcal{V}$.*

We will rely on the following definitions from [69].

Definition 5.3.4 (r-reachable set)**.** *For a given graph $\mathcal{G}$ and a positive integer $r \in \mathbb{Z}_+$, a subset of nodes $\mathcal{S} \subseteq \mathcal{V}$ is said to be r-reachable if there exists a node $v_i \in \mathcal{S}$ such that $|\mathcal{N}_i^{\text{in}} \setminus \mathcal{S}| \geq r$.*

Definition 5.3.5 (r-robust graphs)**.** *For a positive integer $r \in \mathbb{Z}_+$, a graph $\mathcal{G}$ is said to be r-robust if for all pairs of disjoint nonempty subsets $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{V}$, at least one of $\mathcal{S}_1$ or $\mathcal{S}_2$ is r-reachable.*

The above definitions capture the idea that sets of nodes should contain individual nodes that have a sufficient number of neighbors outside that set. This will be important for the *local* decisions made by each node in resilient distributed algorithms, and will allow information from the rest of the network to penetrate into different sets of nodes.

Next, following [89], we define the composition of two graphs and conditions on a sequence of graphs which will be useful for stating a mild condition for achieving approximate consensus guarantees later as follows.

Definition 5.3.6 (composition)**.** *The composition of a directed graph $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ with a directed graph $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ written as $\mathcal{G}_2 \circ \mathcal{G}_1$, is the directed graph $(\mathcal{V}, \mathcal{E})$ with $(v_i, v_j) \in \mathcal{E}$ if there is a $v_k \in \mathcal{V}$ such that $(v_i, v_k) \in \mathcal{E}_1$ and $(v_k, v_j) \in \mathcal{E}_2$.*

Definition 5.3.7 (jointly rooted)**.** *A finite sequence of directed graphs $\{\mathcal{G}_k\}_{k\in[K]}$ is jointly rooted if the composition $\mathcal{G}_K \circ \mathcal{G}_{K-1} \circ \cdots \circ \mathcal{G}_1$ is rooted.*

Definition 5.3.8 (repeatedly jointly rooted)**.** *An infinite sequence of graphs $\{\mathcal{G}_k\}_{k\in\mathbb{Z}_+}$ is repeatedly jointly rooted if there is a positive integer $q \in \mathbb{Z}_+$ for which each finite sequence $\mathcal{G}_{q(k-1)+1}, \ldots, \mathcal{G}_{qk}$ is jointly rooted for all $k \in \mathbb{Z}_+$.*

For a stochastic matrix $S \in \mathbb{S}^N$, let $\mathbb{G}(S)$ denote the graph $\mathcal{G}$ whose adjacency matrix is the transpose of the matrix obtained by replacing all of S's nonzero entries with 1's.

5.3.4 Adversarial Behavior

Definition 5.3.9. *A node $v_i \in \mathcal{V}$ is said to be Byzantine if during each iteration of the prescribed algorithm, it is capable of sending arbitrary values to different neighbors. It is also allowed to update its local information arbitrarily at each iteration of any prescribed algorithm.*

The set of Byzantine agents is denoted by $\mathcal{V}_{\mathcal{B}} \subset \mathcal{V}$. The set of regular agents is denoted by $\mathcal{V}_{\mathcal{R}} = \mathcal{V} \setminus \mathcal{V}_{\mathcal{B}}$. The identities of the Byzantine agents are unknown to the regular agents in advance. Furthermore, we allow the Byzantine agents to know the entire topology of the network, functions equipped by the regular nodes, and the deployed algorithm. In addition, Byzantine agents are allowed to coordinate with other Byzantine agents and access the current and previous information contained by the nodes in the network (e.g. current and previous states of all nodes). Such extreme behavior is typical in the field of distributed computing [90] and in adversarial distributed optimization [30], [33], [44], [91], [92]. In

exchange for allowing such extreme behavior, we will consider a limitation on the number of such adversaries in the neighborhood of each regular node, as follows.

Definition 5.3.10 (F-local model)**.** *For a positive integer $F \in \mathbb{Z}_+$, we say that the set of adversaries $\mathcal{V}_\mathcal{B}$ is an F-local set if $|\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{B}| \leq F$, for all $v_i \in \mathcal{V}_\mathcal{R}$.*

Thus, the F-local model captures the idea that each regular node has at most F Byzantine in-neighbors.

5.4 Problem Formulation

Consider a group of N agents $\mathcal{V}$ interconnected over a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each agent $v_i \in \mathcal{V}$ has a local cost function $f_i : \mathbb{R}^d \to \mathbb{R}$. Since Byzantine nodes are allowed to send arbitrary values to their neighbors at each iteration of any algorithm, it is not possible to minimize the quantity $\frac{1}{N}\sum_{v_i \in \mathcal{V}} f_i(\boldsymbol{x})$ that is typically sought in distributed optimization (since one is not guaranteed to infer any information about the true functions of the Byzantine agents) [30], [33]. Thus, we restrict the summation only to the regular agents' functions, i.e., the objective is to minimize

$$f(\boldsymbol{x}) := \frac{1}{|\mathcal{V}_\mathcal{R}|} \sum_{v_i \in \mathcal{V}_\mathcal{R}} f_i(\boldsymbol{x}). \tag{5.3}$$

A key challenge in solving the above problem is that no regular agent is aware of the identities or actions of the Byzantine agents. In particular, solving Problem 5.3 exactly is not possible under Byzantine behavior, since the identities and local functions of the Byzantine nodes are not known (the Byzantine agents can simply change their local functions and pretend to be a regular agent in the algorithms and these can never be detected). Therefore, one must settle for computing an approximate solution to Problem 5.3 (see [30], [33] for a more detailed discussion of this fundamental limitation).

Establishing the convergence (especially obtaining the rate of convergence) for resilient distributed optimization algorithms under general assumptions on the local functions (i.e., not assuming i.i.d. or redundancy) is non-trivial as evidenced by the lack of such results in the literature. We close this gap by introducing a "proper" intermediate step which is showing the states contraction property (Definition 5.6.1) before proceeding to show the convergence

Algorithmic Framework 1 Resilient Distributed Gradient-Descent Algorithmic Framework (RedGraf)

Input: Network $\mathcal{G}$, functions $\{f_i\}_{\mathcal{V}_\mathcal{R}}$, parameter F

Step I: Initialization
Each $v_i \in \mathcal{V}_\mathcal{R}$ sets $\boldsymbol{z}_i[0] \leftarrow \texttt{init}(f_i)$
for $k = 0, 1, 2, 3, \ldots$ **do**
  for $v_i \in \mathcal{V}_\mathcal{R}$ **do** ▷ Implement in parallel
    Step II: Broadcast and Receive
    v_i broadcasts $\boldsymbol{z}_i[k]$ to $\mathcal{N}_i^{\text{out}}$ and receives $\boldsymbol{z}_j[k]$ from $v_j \in \mathcal{N}_i^{\text{in}}$
    Let $\mathcal{Z}_i[k] = \left\{\boldsymbol{z}_j[k] : v_j \in \mathcal{N}_i^{\text{in}} \cup \{v_i\}\right\}$
    Step III: Filtering Step
    $\tilde{\boldsymbol{z}}_i[k] \leftarrow \texttt{filt}(\mathcal{Z}_i[k], F)$ ▷ Note: $\tilde{\boldsymbol{z}}_i[k] = \left[\tilde{\boldsymbol{x}}_i^T[k], \tilde{\boldsymbol{y}}_i^T[k]\right]^T$
    Step IV: Gradient Update

$$\begin{aligned} \boldsymbol{x}_i[k+1] &= \tilde{\boldsymbol{x}}_i[k] - \alpha_k \nabla f_i(\tilde{\boldsymbol{x}}_i[k]), \\ \boldsymbol{y}_i[k+1] &= \tilde{\boldsymbol{y}}_i[k] \quad \text{(if exists)}, \end{aligned} \tag{5.4}$$

    where $\alpha_k \in \mathbb{R}_{>0}$ is the step-size
  end for
  $\boldsymbol{z}_i[k+1] = \left[\boldsymbol{x}_i^T[k+1], \boldsymbol{y}_i^T[k+1]\right]^T$ for $v_i \in \mathcal{V}_\mathcal{R}$
end for

(Theorem 5.6.4). Importantly, the contraction property not only captures some of state-of-the-art resilient distributed optimization algorithms in the literature (Theorem 5.6.8) but also facilitates the (geometric) convergence analysis.

5.5 Resilient Distributed Optimization Algorithms

5.5.1 Our Framework

In this subsection, we introduce a class of resilient distributed optimization algorithms represented by the ***Re**silient **D**istributed **Gr**adient-Descent **A**lgorithmic **F**ramework (RedGraf)* shown in Algorithmic Framework 1. At each time-step $k \in \mathbb{N}$, each regular agent[1] $v_i \in \mathcal{V}_\mathcal{R}$ maintains and updates a state vector $\boldsymbol{x}_i[k] \in \mathbb{R}^d$, which is its estimate of the solution to Problem 5.3, and optionally an auxiliary vector $\boldsymbol{y}_i[k] \in \mathbb{R}^{d'}$ where the dimension $d' \in \mathbb{N}$ depends on the specific algorithm. In our algorithmic framework, we let

[1]↑Byzantine agents do not necessarily need to follow the above algorithm, and can update their states however they wish.

$\boldsymbol{z}_i[k] = \left[\boldsymbol{x}_i^T[k], \boldsymbol{y}_i^T[k]\right]^T \in \mathbb{R}^{d+d'}$, and similarly, $\tilde{\boldsymbol{z}}_i[k] = \left[\tilde{\boldsymbol{x}}_i^T[k], \tilde{\boldsymbol{y}}_i^T[k]\right]^T \in \mathbb{R}^{d+d'}$. In fact, RedGraf is a generalization of BRIDGE proposed in [45] in the sense that our framework allows the state vector $\boldsymbol{z}_i[k]$ to include the auxiliary vector $\boldsymbol{y}_i[k]$. In Algorithmic Framework 1, the operation $\texttt{init}(f_i)$ initializes $\boldsymbol{z}_i[0] = \left[\boldsymbol{x}_i^T[0], \boldsymbol{y}_i^T[0]\right]^T$, and the operation $\texttt{filt}(\mathcal{Z}_i[k], F)$ performs a filtering procedure (to remove potentially adversarial states received from neighbors) and returns a vector $\tilde{\boldsymbol{z}}_i[k]$. These functions will vary across algorithms, and will be discussed for specific algorithms later.

5.5.2 Definition of Some Standard Operations for Resilient Distributed Optimization

To show that our framework (RedGraf) captures several existing resilient distributed optimization algorithms as special cases, we first define some operations that are used by existing algorithms. Throughout, let $\mathcal{V}_i[k] \subseteq \mathcal{N}_i^{\text{in}} \cup \{v_i\}$, $\mathcal{X}_i[k] = \{\boldsymbol{x}_j[k]\}_{\mathcal{N}_i^{\text{in}} \cup \{v_i\}}$ and $\mathcal{Y}_i[k] = \{\boldsymbol{y}_j[k]\}_{\mathcal{N}_i^{\text{in}} \cup \{v_i\}}$.

- $\tilde{\mathcal{V}}_i[k] \leftarrow \texttt{dist\_filt}(\mathcal{V}_i[k], \mathcal{Z}_i[k], F)$:
  Regular agent $v_i \in \mathcal{V}_{\mathcal{R}}$ removes F states that are far away from $\boldsymbol{y}_i[k]$. More specifically, an agent $v_j \in \mathcal{V}_i[k]$ is in $\tilde{\mathcal{V}}_i[k]$ if and only if

  $$\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq \max\{q_M, \|\boldsymbol{x}_i[k] - \boldsymbol{y}_i[k]\|\},$$

  where $q_M = M_F\Big(\{\|\boldsymbol{x}_s[k] - \boldsymbol{y}_i[k]\|\}_{v_s \in \mathcal{V}_i[k]}\Big)$.

- $\tilde{\mathcal{V}}_i[k] \leftarrow \texttt{full\_mm\_filt}(\mathcal{V}_i[k], \mathcal{X}_i[k], F)$:
  Regular agent $v_i \in \mathcal{V}_{\mathcal{R}}$ removes states that have extreme values in any of their components. For a given $k \in \mathbb{N}$ and $\ell \in [d]$, let $q_m^{(\ell)} = m_F\Big(\{x_s^{(\ell)}[k]\}_{\mathcal{V}_i[k]}\Big)$ and $q_M^{(\ell)} = M_F\Big(\{x_s^{(\ell)}[k]\}_{\mathcal{V}_i[k]}\Big)$. An agent $v_j \in \mathcal{V}_i[k]$ is in $\tilde{\mathcal{V}}_i[k]$ if and only if for all $\ell \in [d]$,

  $$\min\{q_m^{(\ell)}, x_i^{(\ell)}[k]\} \leq x_j^{(\ell)}[k] \leq \max\{q_M^{(\ell)}, x_i^{(\ell)}[k]\}.$$

- $\{\tilde{\mathcal{V}}_i^{(\ell)}[k]\}_{\ell\in[d]} \leftarrow$ `cw_mm_filt`$(\mathcal{V}_i[k],\, \mathcal{X}_i[k],\, F)$:
  For each dimension $\ell \in [d]$, regular agent $v_i \in \mathcal{V}_\mathcal{R}$ removes the F highest and F lowest values of the states of agents in $\mathcal{V}_i[k]$ along that dimension. More specifically, for a given $k \in \mathbb{N}$ and $\ell \in [d]$, let $q_m^{(\ell)} = m_F\Big(\{x_s^{(\ell)}[k]\}_{\mathcal{V}_i[k]}\Big)$ and $q_M^{(\ell)} = M_F\Big(\{x_s^{(\ell)}[k]\}_{\mathcal{V}_i[k]}\Big)$. An agent $v_j \in \mathcal{V}_i[k]$ is in $\tilde{\mathcal{V}}_i^{(\ell)}[k]$ if and only if

$$\min\{q_m^{(\ell)}, x_i^{(\ell)}[k]\} \leq x_j^{(\ell)}[k] \leq \max\{q_M^{(\ell)}, x_i^{(\ell)}[k]\}.$$

- $\tilde{\boldsymbol{x}}_i[k] \leftarrow$ `full_average`$(\mathcal{V}_i[k],\, \mathcal{X}_i[k])$:
  Regular agent $v_i \in \mathcal{V}_\mathcal{R}$ computes

$$\tilde{\boldsymbol{x}}_i[k] = \sum_{v_j\in\mathcal{V}_i[k]} w_{ij}[k]\; \boldsymbol{x}_j[k], \tag{5.5}$$

  where $\sum_{v_j\in\mathcal{V}_i[k]} w_{ij}[k] = 1$ and $w_{ij}[k] \in \mathbb{R}_{>0}$ for all $v_j \in \mathcal{V}_i[k]$.

- $\tilde{\boldsymbol{x}}_i[k] \leftarrow$ `cw_average`$(\{\mathcal{V}_i^{(\ell)}[k]\}_{\ell\in[d]},\, \mathcal{X}_i[k])$:
  For each dimension $\ell \in [d]$, regular agent $v_i \in \mathcal{V}_\mathcal{R}$ computes

$$\tilde{x}_i^{(\ell)}[k] = \sum_{v_j\in\mathcal{V}_i^{(\ell)}[k]} w_{ij}^{(\ell)}[k]\; x_j^{(\ell)}[k], \tag{5.6}$$

  where $w_{ij}^{(\ell)}[k] \in \mathbb{R}_{>0}$ for all $v_j \in \mathcal{V}_i^{(\ell)}[k]$ and $\sum_{v_j\in\mathcal{V}_i^{(\ell)}[k]} w_{ij}^{(\ell)}[k] = 1$.

- $\tilde{\boldsymbol{x}}_i[k] \leftarrow$ `safe_point`$(\mathcal{V}_i[k],\, \mathcal{X}_i[k],\, F)$:
  Regular agent $v_i \in \mathcal{V}_\mathcal{R}$ returns a state $\tilde{\boldsymbol{x}}_i[k]$ which can be written as

$$\tilde{\boldsymbol{x}}_i[k] = \sum_{v_j\in\mathcal{V}_i[k]\cap\mathcal{V}_\mathcal{R}} w_{ij}[k]\; \boldsymbol{x}_j[k], \tag{5.7}$$

  where $w_{ij}[k] \in \mathbb{R}_{>0}$ for all $v_j \in \mathcal{V}_i[k] \cap \mathcal{V}_\mathcal{R}$ and $\sum_{v_j\in\mathcal{V}_i[k]\cap\mathcal{V}_\mathcal{R}} w_{ij}[k] = 1$. The works [74], [77] discuss methods used to compute $\tilde{\boldsymbol{x}}_i[k]$.

5.5.3 Mapping Existing Algorithms into RedGraf

Using the operations defined above, we now discuss some algorithms in the literature that fall into our algorithmic framework.

Simultaneous Distance-MixMax Filtering Dynamics (SDMMFD) and *Simultaneous Distance Filtering Dynamics (SDFD)* [93]: These two algorithms are captured in our framework by defining $\boldsymbol{z}_i[k] = \left[\boldsymbol{x}_i^T[k], \boldsymbol{y}_i^T[k]\right]^T$ where $\boldsymbol{y}_i[k] \in \mathbb{R}^d$. In the initialization step $\boldsymbol{z}_i[0] \leftarrow \texttt{init}(f_i)$ (**Line 1**) of both algorithms, each regular agent $v_i \in \mathcal{V}_\mathcal{R}$ computes an approximate minimizer $\hat{\boldsymbol{x}}_i^* \in \mathbb{R}^d$ of its local function f_i (using any appropriate optimization algorithm) and then sets $\boldsymbol{x}_i[0] \in \mathbb{R}^d$ arbitrarily and $\boldsymbol{y}_i[0] = \hat{\boldsymbol{x}}_i^*$. In the filtering step $\tilde{\boldsymbol{z}}_i[k] \leftarrow \texttt{filt}(\mathcal{Z}_i[k], F)$ (**Line 5**), SDMMFD executes the following sequence of methods:

1. $\mathcal{V}_i^{\text{dist}}[k] \leftarrow \texttt{dist\_filt}(\mathcal{N}_i^{\text{in}} \cup \{v_i\}, \mathcal{Z}_i[k], F)$,
2. $\mathcal{V}_i^{\text{x,mm}}[k] \leftarrow \texttt{full\_mm\_filt}(\mathcal{V}_i^{\text{dist}}[k], \mathcal{X}_i[k], F)$,
3. $\tilde{\boldsymbol{x}}_i[k] \leftarrow \texttt{full\_average}(\mathcal{V}_i^{\text{x,mm}}[k], \mathcal{X}_i[k])$,
4. $\{\tilde{\mathcal{V}}_i^{(\ell)}[k]\}_{\ell \in [d]} \leftarrow \texttt{cw\_mm\_filt}(\mathcal{N}_i^{\text{in}} \cup \{v_i\}, \mathcal{Y}_i[k], F)$,
5. $\tilde{\boldsymbol{y}}_i[k] \leftarrow \texttt{cw\_average}(\{\tilde{\mathcal{V}}_i^{(\ell)}[k]\}_{\ell \in [d]}, \mathcal{Y}_i[k])$.

The first three steps compute the intermediate main state $\tilde{\boldsymbol{x}}_i[k]$ while the last two steps compute the intermediate auxiliary vector $\tilde{\boldsymbol{y}}_i[k]$. On the other hand, SDFD executes the same sequence of methods except that step (ii) is removed and $\mathcal{V}_i^{\text{x,mm}}[k]$ in step (iii) is replaced by $\mathcal{V}_i^{\text{dist}}[k]$. Then, for both algorithms, we set $\tilde{\boldsymbol{z}}_i[k] = [\tilde{\boldsymbol{x}}_i^T[k], \tilde{\boldsymbol{y}}_i^T[k]]^T$.

Coordinate-wise Trimmed Mean (CWTM) [30], [33], [45], [50]–[52] and *Resilient Vector Optimization (RVO)* based on *resilient vector consensus* [74], [77]: These algorithms are captured by setting $\boldsymbol{z}_i[k] = \boldsymbol{x}_i[k]$ (i.e., $\boldsymbol{y}_i[k] = \emptyset$). In the initialization step $\boldsymbol{z}_i[0] \leftarrow \texttt{init}(f_i)$ (**Line 1**) of both algorithms, the regular agents $v_i \in \mathcal{V}_\mathcal{R}$ arbitrarily initialize $\boldsymbol{x}_i[0] \in \mathbb{R}^d$. In the filtering step $\tilde{\boldsymbol{z}}_i[k] \leftarrow \texttt{filt}(\mathcal{Z}_i[k], F)$ (**Line 5**), CWTM executes the following sequence of methods:

1. $\{\tilde{\mathcal{V}}_i^{(\ell)}[k]\}_{\ell \in [d]} \leftarrow \texttt{cw\_mm\_filt}(\mathcal{N}_i^{\text{in}} \cup \{v_i\}, \mathcal{X}_i[k], F)$,

2. $\tilde{\boldsymbol{x}}_i[k] \leftarrow \texttt{cw\_average}(\{\mathcal{V}_i^{(\ell)}[k]\}_{\ell\in[d]}, \mathcal{X}_i[k])$,

whereas RVO executes

1. $\tilde{\boldsymbol{x}}_i[k] \leftarrow \texttt{safe\_point}(\mathcal{N}_i^{\text{in}} \cup \{v_i\}, \mathcal{X}_i[k], F)$.

In fact, the algorithms proposed in [49], [85]–[87] also fall into our framework. However, in this work, we focus on analyzing the four algorithms above since they share some common property (stated formally in Theorem 5.6.8), and we will provide a discussion on the algorithms in the work [49], [85]–[87] in Section 5.6.4.

5.6 Assumptions and Main Results

We now turn to stating assumptions and definitions in Section 5.6.1 which will be used to prove convergence properties in Section 5.6.2 and consensus properties in Section 5.6.3. Finally, in Section 5.6.4, we analyze certain properties of each algorithm mentioned in the previous section.

5.6.1 Assumptions and Definitions

Assumption 5.6.1. For all $v_i \in \mathcal{V}$, given positive numbers $\mu_i \in \mathbb{R}_{>0}$ and $L_i \in \mathbb{R}_{>0}$, the functions f_i are μ_i-strongly convex and differentiable. Furthermore, the functions f_i have L_i-Lipschitz continuous gradients.

The strongly convexity and Lipschitz continuous gradient assumptions given above are common in the distributed convex optimization literature [29], [80], [94]–[96]. We define $\tilde{L} := \max_{v_i\in\mathcal{V}_\mathcal{R}} L_i$ and $\tilde{\mu} := \min_{v_i\in\mathcal{V}_\mathcal{R}} \mu_i$. Since $\{f_i\}_{v_i\in\mathcal{V}_\mathcal{R}}$ are strongly convex functions, let $\boldsymbol{x}_i^* \in \mathbb{R}^d$ be the minimizer of the function f_i, i.e., $f_i(\boldsymbol{x}_i^*) = \min_{\boldsymbol{x}\in\mathbb{R}^d} f_i(\boldsymbol{x})$. Moreover, let $\boldsymbol{c}^* \in \mathbb{R}^d$ and $r^* \in \mathbb{R}_{\geq 0}$ be such that $\boldsymbol{x}_i^* \in \mathcal{B}(\boldsymbol{c}^*, r^*)$ for all $v_i \in \mathcal{V}_\mathcal{R}$. Let $\boldsymbol{x}^* \in \mathbb{R}^d$ be the minimizer of the function $f(\boldsymbol{x})$, i.e., the solution of Problem 5.3. In other words, $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x})$. For convenience, we also denote $f^* := f(\boldsymbol{x}^*)$ and $\boldsymbol{g}_i[k] := \nabla f_i(\tilde{\boldsymbol{x}}_i[k])$.

Assumption 5.6.2. Given a positive integer $F \in \mathbb{Z}_+$, the Byzantine agents form a F-local set.

Assumption 5.6.3. There exists a positive number $\omega \in \mathbb{R}_{>0}$ such that for all $k \in \mathbb{N}$ and $\ell \in [d]$, the non-zero weights $w_{ij}[k]$ in `full_average` and `safe_point`, and $w_{ij}^{(\ell)}[k]$ in `cw_average` (all defined in Section 5.5.2) are lower bounded by ω.

Now, we introduce certain properties of the filtering step of our algorithmic framework. These definitions will be important ingredients in proving the convergence result in Section 5.6.2 and consensus result in Section 5.6.3.

Definition 5.6.1. *For a vector $\boldsymbol{x}_c \in \mathbb{R}^d$, constant $\gamma \in \mathbb{R}_{\geq 0}$, and sequence $\{c[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$, a resilient distributed optimization algorithm A in RedGraf is said to satisfy the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property if it holds that $\lim_{k\to\infty} c[k] = 0$ and for all $k \in \mathbb{N}$ and $v_i \in \mathcal{V}_\mathcal{R}$,*

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\| \leq \sqrt{\gamma} \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| + c[k]. \tag{5.8}$$

In the above definition, the vector $\boldsymbol{x}_c$ is called the *contraction center* and the constant γ is called the *contraction factor*. In general, we want the contraction factor γ to be small so that the intermediate state $\tilde{\boldsymbol{x}}_i[k]$ remains close to the center $\boldsymbol{x}_c$. The sequence $\{c[k]\}_{k\in\mathbb{N}}$ captures a perturbation to the contraction term in each time-step. If an algorithm A in RedGraf satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property, we define

$$r_c := \max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_c - \boldsymbol{x}_i^*\|. \tag{5.9}$$

Definition 5.6.2. *Suppose Assumption 5.6.1 holds, and a resilient distributed optimization algorithm A in RedGraf satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property (for some $\boldsymbol{x}_c \in \mathbb{R}^d$, $\gamma \in \mathbb{R}_{\geq 0}$, and $\{c[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$). Then, algorithm A is said to satisfy the (γ, α)-reduction property if $\gamma \in [0, 1)$ and $\alpha_k = \alpha \in \left(0, \frac{1}{\tilde{L}}\right]$, or $\gamma \in \left[1, \frac{1}{1-\frac{\tilde{\mu}}{\tilde{L}}}\right)$ and $\alpha_k = \alpha \in \left(\frac{1}{\tilde{\mu}}\left(1 - \frac{1}{\gamma}\right), \frac{1}{\tilde{L}}\right]$.*

Let $\mathcal{V}_\mathcal{R} = \{v_{i_1}, v_{i_2}, \ldots, v_{i_{|\mathcal{V}_\mathcal{R}|}}\}$ denote the set of all regular agents. For a set of vectors $\{\boldsymbol{u}_i\}_{i\in\mathcal{V}} \subset \mathbb{R}^d$ and $\ell \in [d]$, we denote $\boldsymbol{u}^{(\ell)} = \left[u_{i_1}^{(\ell)}, u_{i_2}^{(\ell)}, \ldots, u_{i_{|\mathcal{V}_\mathcal{R}|}}^{(\ell)}\right]^T \in \mathbb{R}^{|\mathcal{V}_\mathcal{R}|}$, the vector containing ℓ-th dimension of each vector $\boldsymbol{u}_i$ corresponding to the regular agents' indices. The

following definition characterizes the dynamics of all the regular agents in the network which will be a crucial ingredient in showing the approximate consensus result in Section 5.6.3.

Definition 5.6.3. *For a set of sequences of matrices* $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\,\ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ *and constant* $G \in \mathbb{R}_{\geq 0}$*, a resilient distributed optimization algorithm* A *in* RedGraf *is said to possess* $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$*-mixing dynamics if the state dynamics can be written as*

$$\boldsymbol{x}^{(\ell)}[k+1] = \boldsymbol{W}^{(\ell)}[k]\boldsymbol{x}^{(\ell)}[k] - \alpha_k \boldsymbol{g}^{(\ell)}[k] \tag{5.10}$$

for all $k \in \mathbb{N}$ *and* $\ell \in [d]$*, the sequences of graphs* $\{\mathbb{G}(\boldsymbol{W}^{(\ell)}[k])\}_{k\in\mathbb{N}}$ *are repeatedly jointly rooted for all* $\ell \in [d]$*, and* $\limsup_k \|\boldsymbol{g}_i[k]\|_\infty \leq G$ *for all* $v_i \in \mathcal{V}_\mathcal{R}$*.*

The matrix $\boldsymbol{W}^{(\ell)}[k]$ is called a *mixing matrix* which directly affects the ability of the nodes to reach consensus [89] while the constant G quantifies an upper bound on the perturbation (i.e., the scaled gradient $\alpha_k \boldsymbol{g}^{(\ell)}[k]$) to the consensus process.

5.6.2 The Region To Which The States Converge

In this subsection, we will derive a convergence result for some particular algorithms in RedGraf (Theorem 5.6.4), and show that the minimizer $\boldsymbol{x}^*$ which is the solution of Problem 5.3 is, in fact, in the convergence region (Theorem 5.6.5). For convenience, if Assumption 5.6.1 holds and the step-size $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, we define

$$\beta := \sqrt{1 - \alpha\tilde{\mu}}. \tag{5.11}$$

We now come to one of the main results of this chapter, showing that the states of all the regular agents will converge to a ball for all algorithms in RedGraf satisfying the reduction (Definition 5.6.2) and the states contraction (Definition 5.6.1) properties. The proof is provided in Appendix C.2.3.

Theorem 5.6.4 (Convergence)**.** *Suppose Assumption 5.6.1 holds. If an algorithm A satisfies the (γ, α)-reduction property (for some $\gamma \in \mathbb{R}_{\geq 0}$ and $\alpha \in \mathbb{R}_{>0}$), then for all $v_i \in \mathcal{V}_{\mathcal{R}}$, it holds that*

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| \leq \frac{r_c\sqrt{\alpha\tilde{L}}}{1-\beta\sqrt{\gamma}} := R^*, \tag{5.12}$$

where $\boldsymbol{x}_c$, r_c and β are defined in Definition 5.6.1, (5.9) and (5.11), respectively. Furthermore, if $c[k] = \mathcal{O}(\xi^k)$ and $\xi \in (0,1) \setminus \{\beta\sqrt{\gamma}\}$, then for all $v_i \in \mathcal{V}_{\mathcal{R}}$,

$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| \leq R^* + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big). \tag{5.13}$$

We refer to R^* in (5.12) as the *convergence radius*, and $\mathcal{B}(\boldsymbol{x}_c, R^*)$ as the *convergence region*. In particular, the convergence region is the ball which has the center at $\boldsymbol{x}_c$ and the radius R^* depending on the functions' parameters μ_i and L_i, the contraction factor γ, the constant step-size α, and the constant capturing the position of the contraction center r_c (defined in (5.9)). We emphasize that the convergence region does not depend on the contraction perturbation sequence $\{c[k]\}_{k\in\mathbb{N}}$ as long as the sequence converges to 0.

Considering the expression of the convergence radius R^* in (5.12), it should be noted that R^* is strictly decreasing with respect to γ. In addition, applying Lemma C.1.4 in Appendix C.1.3 (by setting $\frac{\tilde{L}}{\tilde{\mu}} = \sigma$ and $\alpha = \frac{s}{\tilde{\mu}}$ in R^*) and noting that R^* is a continuous function with respect to α, we can conclude as follows.

- For $\gamma \in [0, 1)$, a small constant step-size α would yield a small convergence radius R^* (since $R^*\big|_{\alpha=0} = 0$ and $R^*\big|_{\alpha=\frac{1}{\tilde{\mu}}} = r_c\sqrt{\frac{\tilde{L}}{\tilde{\mu}}}$). Furthermore, we have $R^* \leq R^*\big|_{\alpha=\frac{1}{\tilde{\mu}}(1-\gamma)} = \frac{r_c}{\sqrt{1-\gamma}}\sqrt{\frac{\tilde{L}}{\tilde{\mu}}}$ for all valid values of α.
- For $\gamma \in [1, \infty)$, the optimal convergence radius is obtained by choosing $\alpha = \frac{1}{\tilde{L}}$ (due to the condition on α in Theorem 5.6.4) and the corresponding radius is $R^*\big|_{\alpha=\frac{1}{\tilde{L}}} = \frac{r_c}{1-\sqrt{\gamma}\sqrt{1-\frac{\tilde{\mu}}{\tilde{L}}}}$.

From (5.13), we can conclude that the states of the regular agents converge **geometrically** to the convergence region $\mathcal{B}(\boldsymbol{x}_c, R^*)$. Furthermore, recall the definition of β from

(5.11). For $\beta\sqrt{\gamma} > \xi$, the inequality (5.13) suggests that the convergence rate increases as the constant step-size α increases.

Next, recall that $\boldsymbol{x}^* \in \mathbb{R}^d$ is the minimizer of the function $\frac{1}{|\mathcal{V}_\mathcal{R}|}\sum_{v_i \in \mathcal{V}_\mathcal{R}} f_i(\boldsymbol{x})$ which is our objective function (Problem 5.3). The following theorem shows that, in fact, the true minimizer $\boldsymbol{x}^*$ is in the convergence region given that a certain condition on γ and α holds (proved in Appendix C.2.4).

Theorem 5.6.5. *Suppose Assumption 5.6.1 holds. Then,* $\boldsymbol{x}^* \in \mathcal{B}\Big(\boldsymbol{x}_c, \frac{\tilde{L}}{\tilde{\mu}} r_c\Big)$. *Furthermore, if* $\gamma \in \left[1, \frac{1}{1-\frac{\tilde{\mu}}{\tilde{L}}}\right)$ *and* $\alpha \in \left(\frac{1}{\tilde{\mu}}\left(1-\frac{1}{\gamma}\right), \frac{1}{\tilde{L}}\right]$, *then* $\boldsymbol{x}^* \in \mathcal{B}(\boldsymbol{x}_c, R^*)$.

It is important to emphasize that Theorems 5.6.4 and 5.6.5 together implies that the regular agents' states converge geometrically to a neighborhood of the true minimizer $\boldsymbol{x}^*$, which is impossible to determine exactly in the presence of Byzantine agents.

Our result from Theorem 5.6.4 offers a different approach to convergence proofs than those typically found in the literature, which are often designed for specific algorithms. By focusing on proving the state contraction property (Definition 5.6.1), rather than the details of the functions involved, one can save a considerable amount of time and effort. However, it is worth noting that this approach only provides a sufficient condition for convergence. There may be resilient algorithms that do not satisfy the property but still converge geometrically. In fact, finding general necessary conditions for convergence in resilient distributed optimization remains an open question in the literature.

Remark 8. The work [68] introduces a contraction property which seems to be similar to Definition 5.6.1. However, there is a subtle difference in that their contraction center is time-varying (since it is a function of neighbors' states) while it is a constant (but depends on algorithms) in our case. However, it is unclear whether their notion of contraction allows for the proof of geometric convergence, as demonstrated in Theorem 5.6.4.

Having established convergence of all regular agent's values to a ball that contains the true minimizer, we now turn our attention to characterizing the distance between the regular agents' values within that ball. Given the states contraction property in Definition 5.6.1, using (5.12), we can simply obtain a bound on the distance between the values held by different nodes as $\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq 2R^*$ for all $v_i, v_j \in \mathcal{V}_\mathcal{R}$. However, this bound

is not meaningful since the quantity on the RHS can be large. As we will show in the next subsection, the mixing dynamics (Definition 5.6.3) and a constant step-size are sufficient to obtain a better bound on the approximate consensus.

5.6.3 Convergence to Approximate Consensus of States

The following theorem characterizes the approximate consensus among the regular agents in the network under the mixing dynamics (Definition 5.6.3) and a constant step-size (proved in Appendix C.3.2).

Theorem 5.6.6 (Consensus)**.** *If an algorithm A in* RedGraf *satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property (for some $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\,\ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ and $G \in \mathbb{R}_{\geq 0}$) and $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, then there exist $\rho \in \mathbb{R}_{\geq 0}$ and $\lambda \in (0,1)$ such that for all $v_i, v_j \in \mathcal{V}_\mathcal{R}$, it holds that*

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \frac{\alpha\rho G\sqrt{d}}{1-\lambda}. \tag{5.14}$$

From the consensus theorem above, we note that $\max_{v_i,v_j\in\mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| = \mathcal{O}(\alpha\sqrt{d})$ if G does not depend on the constant step-size α and the dimension d.

Remark 9. According to [89], the quantity $\lambda \in (0,1)$ depends only on the network topology (for each time-step) induced by the sequence of graphs $\{\mathbb{G}(\boldsymbol{W}^{(\ell)}[k])\}$ while the quantity $\rho \in \mathbb{R}_{\geq 0}$ depends on the number of regular agents $|\mathcal{V}_\mathcal{R}|$ and the quantity λ.

In fact, the states contraction property (Definition 5.6.3) implies a bound on the gradient $\|\boldsymbol{g}_i[k]\|_\infty$ (the formal statement is provided in Appendix C.3.1) which is one of the requirements of the mixing dynamics. Thus, we can achieve a similar approximate consensus result as Theorem 5.6.6 given that an algorithm satisfies the states contraction property and the associated sequence of graphs for each dimension is repeatedly jointly rooted as shown in the following corollary whose proof is provided in Appendix C.3.3.

Corollary 5.6.7. *Suppose Assumption 5.6.1 holds. If an algorithm A satisfies the (γ, α)-reduction property (for some $\gamma \in \mathbb{R}_{\geq 0}$ and $\alpha \in \mathbb{R}_{>0}$), and the dynamics of the regular states*

can be written as (5.10) *where* $\{\mathbb{G}(\boldsymbol{W}^{(\ell)}[k])\}_{k\in\mathbb{N}}$ *is repeatedly jointly rooted for all* $\ell \in [d]$*, then there exists* $\rho \in \mathbb{R}_{\geq 0}$ *and* $\lambda \in (0,1)$ *such that for all* $v_i, v_j \in \mathcal{V}_{\mathcal{R}}$*,*

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \frac{\alpha\rho r_c \tilde{L}\sqrt{d}}{1-\lambda}\left(1 + \frac{\sqrt{\alpha\gamma\tilde{L}}}{1-\beta\sqrt{\gamma}}\right) := D^*, \tag{5.15}$$

where r_c *and* β *are defined in* (5.9) *and* (5.11)*, respectively. Furthermore, if* $c[k] = \mathcal{O}(\xi^k)$ *where* $\xi \in (0,1)\backslash\{\beta\sqrt{\gamma}\}$*, then there exists* $\rho \in \mathbb{R}_{\geq 0}$ *and* $\lambda \in (0,1)$ *such that for all* $v_i, v_j \in \mathcal{V}_{\mathcal{R}}$*,*

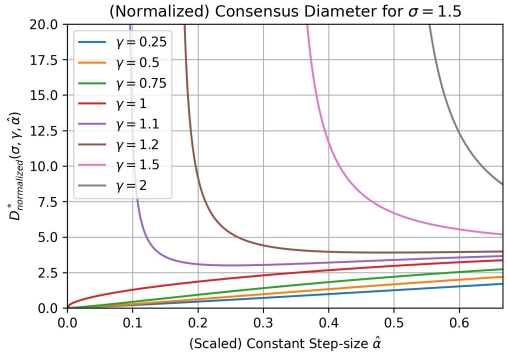
$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq D^* + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi, \lambda\})^k\Big). \tag{5.16}$$

We refer to D^* in (5.15) as the *approximate consensus diameter*. To analyze the expression of D^*, we simplify the expression as follows. Let $\mathrm{dom}(D^*_{\mathrm{normalized}}) = \Big\{(\sigma,\gamma,\hat{\alpha}) \in \mathbb{R}^3 : \sigma \in [1,\infty), \gamma \in [0,1)$ and $\hat{\alpha} \in \Big(0, \frac{1}{\sigma}\Big]$, or $\sigma \in [1,\infty), \gamma \in \Big[1, \frac{\sigma}{\sigma-1}\Big)$ and $\hat{\alpha} \in \Big(1-\frac{1}{\gamma}, \frac{1}{\sigma}\Big]\Big\}$. Using changes of variables $\sigma = \frac{\tilde{L}}{\tilde{\mu}}$ and $\hat{\alpha} = \alpha\tilde{\mu}$ and then normalizing the expression by $\frac{\rho r_c\sqrt{d}}{1-\lambda}$, we have the (normalized) approximate consensus diameter $D^*_{\mathrm{normalized}} : \mathrm{dom}(D^*_{\mathrm{normalized}}) \to \mathbb{R}_{\geq 0}$ defined as

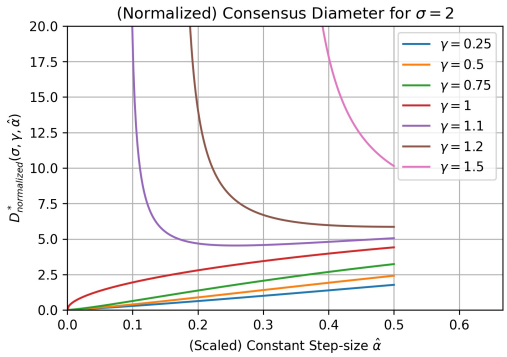
$$D^*_{\mathrm{normalized}}(\sigma,\gamma,\hat{\alpha}) = \sigma\hat{\alpha}\left(1 + \frac{\sqrt{\sigma\gamma\hat{\alpha}}}{1-\sqrt{\gamma}\cdot\sqrt{1-\hat{\alpha}}}\right).$$

Note that the valid values of each variable are $\sigma \in [1,\infty)$, $\gamma \in \Big[1, \frac{\sigma}{\sigma-1}\Big)$ and $\hat{\alpha} \in \Big(1-\frac{1}{\gamma}, \frac{1}{\sigma}\Big]$. It can be noted that $D^*_{\mathrm{normalized}}$ is strictly increasing with both σ and γ. However, $D^*_{\mathrm{normalized}}$ is neither an increasing nor decreasing function with respect to $\hat{\alpha}$. The plots between the (normalized) approximate consensus diameter $D^*_{\mathrm{normalized}}$ and the (scaled) constant step-size $\hat{\alpha}$ for some values of σ and γ are given in Figures 5.1a, 5.1b, and 5.1c. The plots suggest that for $\gamma \leq 1$, small constant step-sizes α provide small approximate consensus diameters D^* while large constant step-sizes α may be preferable in the case that $\gamma > 1$.

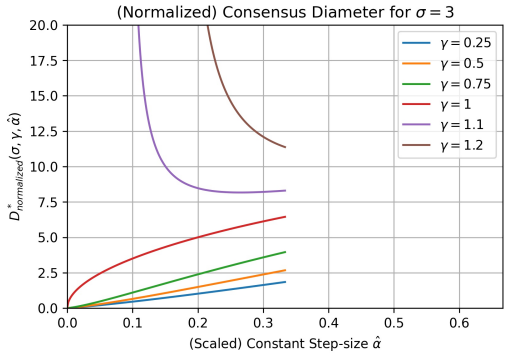
From (5.16), we can conclude that the distance between any two regular agents' states converge **geometrically** to the approximate consensus diameter D^*. Furthermore, as suggested by (5.16), for $\beta\sqrt{\gamma} > \max\{\xi,\lambda\}$, the convergence rate increases as the constant step-size α increases.



(a) $\sigma = 1.5$



(b) $\sigma = 2$



(c) $\sigma = 3$

Figure 5.1. The (normalized) approximate consensus diameter $D^*_{\text{normalized}}$ for different values of the contraction factor γ and for legitimate values of the (scaled) constant step-size $\hat{\alpha}$.

5.6.4 Implications for Existing Resilient Distributed Optimization Algorithms

We now describe the implication of our above results (for our general framework) for the specific existing algorithms discussed in Section 5.5.3: SDMMFD, SDFD [93], CWTM [30], [33], [45], [50]–[52], and RVO [74], [77]. In particular, we show that the algorithms satisfy the states contraction (Definition 5.6.1) and the mixing dynamics (Definition 5.6.3) properties with different quantities which are determined in the following theorem whose proof is provided in Appendix C.4.1.

Before stating the theorem, recall that d is the number of dimensions of the optimization variable $\boldsymbol{x}$ in (5.3) and F is the parameter in the F-local model. Since the step in RVO depends on a specific implemented algorithm, we assume that there exists a function $p : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{Z}_+$ such that if the graph $\mathcal{G}$ is $p(d, F)$-robust then the step in RVO returns a non-empty state.

Theorem 5.6.8. *Suppose Assumptions 5.6.1-5.6.3 hold, $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, and r_c and β are defined in (5.9) and (5.11), respectively. Let $\{0[k]\} = \{0\}_{k\in\mathbb{N}}$.*

- *If $\mathcal{G}$ is $((2d+1)F+1)$-robust then there exists $c_1, c_2 \in \mathbb{R}_{\geq 0}$ such that the SDMMFD from [93] satisfies the $(\boldsymbol{y}[\infty], 1, \{2c_1e^{-c_2k}\})$-states contraction property and there exists $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ such that the algorithm satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property with $G = r_c\tilde{L}\left(1 + \frac{\sqrt{\alpha\tilde{L}}}{1-\beta}\right)$.*

- *If $\mathcal{G}$ is $(2F+1)$-robust then there exists $c_1, c_2 \in \mathbb{R}_{\geq 0}$ such that the SDFD from [93] satisfies the $(\boldsymbol{y}[\infty], 1, \{2c_1e^{-c_2k}\})$-states contraction property.*

- *If $\mathcal{G}$ is $(2F+1)$-robust then the CWTM from [30], [33], [45], [50]–[52] satisfies the $(\boldsymbol{c}^*, d, \{0[k]\})$-states contraction property and there exists $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ such that the algorithm satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property with $G = r_c\tilde{L}\left(1 + \frac{\sqrt{\alpha d\tilde{L}}}{1-\beta\sqrt{d}}\right)$.*

- *If $\mathcal{G}$ is $p(d, F)$-robust then the RVO from [74], [77] satisfies the $(\boldsymbol{c}^*, 1, \{0[k]\})$-states contraction property and there exists $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ such that the algorithm satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property with $G = r_c\tilde{L}\left(1 + \frac{\sqrt{\alpha\tilde{L}}}{1-\beta}\right)$.*

Recall the definition of r_c from (5.9). It is worth noting that the constant r_c appears in two important quantities: the convergence radius R^* and approximate consensus diameter D^* defined in (5.12) and (5.15), respectively. In fact, r_c can be upper bounded by a quantity depending on the diameter of the minimizers of the regular agents' functions r^* defined in Section 5.6.1. The formal statement is provided below and its proof is deferred to Appendix C.4.2.

Lemma 5.6.9. *Suppose Assumption 5.6.1 holds and for the initialization step of SDMMFD and SDFD, there exists $\epsilon^* \in \mathbb{R}_{\geq 0}$ such that $\|\hat{\boldsymbol{x}}_i^* - \boldsymbol{x}_i^*\|_\infty \leq \epsilon^*$ for all $v_i \in \mathcal{V}_\mathcal{R}$.*

- *For SDMMFD and SDFD, we have $r_c \leq \sqrt{d}(r^* + \epsilon^*) + r^*$.*
- *For CWTM and SCC, we have $r_c \leq r^*$.*

Applying the lemma to (5.12), we can conclude that the convergence radius R^* is $\mathcal{O}(\sqrt{d}r^*)$ for SDMMFD and SDFD, and $\mathcal{O}(r^*)$ for CWTM and RVO. In fact, even in the simple case of univariate functions, the convergence radius of $\mathcal{O}(r^*)$ is typical in the literature [30], [33], and an additional assumption, e.g., i.i.d. training samples [44], [45] or redundancy among the local functions [47] is needed to obtain $R^* = o(r^*)$. Still, the question regarding a tight lower bound on the convergence radius for the general case (whether it is $\Omega(r^*)$) remains an open problem.

Remark 10. It is worth noting that the algorithms proposed in [86] and [85] do not satisfy the states contraction property (Definition 5.6.1). However, in fact, they satisfy inequality (5.8) with the perturbation term being bounded by a constant, and thus it is not difficult to use our techniques to show that they geometrically converge to a region with the contraction center $\boldsymbol{x}_c$ but the region has the radius greater than R^* given in (5.12). On the other hand, the algorithms in [49], [87] do not satisfy the contraction property and may require other techniques to establish convergence (if possible).

Having proved the states contraction and mixing dynamics properties of the algorithms from [30], [33], [45], [50]–[52], [74], [77], [93], from Theorem 5.6.4, we can deduce that under certain conditions on the graph robustness and step-size α_k, the states of the regular

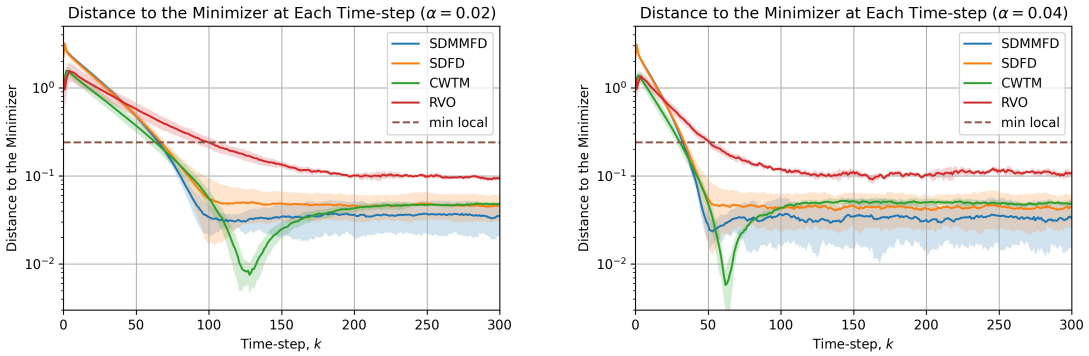
agents geometrically converge to the convergence region with $\boldsymbol{x}_c$ and γ determined by Theorem 5.6.8. In addition, by Theorem 5.6.5, the convergence region $\mathcal{B}(\boldsymbol{x}_c, R^*)$ (which depends on the implemented algorithm) contains the true minimizer $\boldsymbol{x}^*$. On the other hand, from Theorem 5.6.6, we can deduce that the states of the regular agents geometrically converge together at least until the diameter reaches the approximate consensus diameter D^* (where the parameters r_c and γ depend on the implemented algorithm).

To the best of our knowledge, our work is the first to show the geometric convergence results and characterize the convergence region for the resilient algorithms mentioned above. Thus, our framework, defined properties, and proof techniques provide a general approach for analyzing the convergence region and rate for a wide class of resilient optimization algorithms.

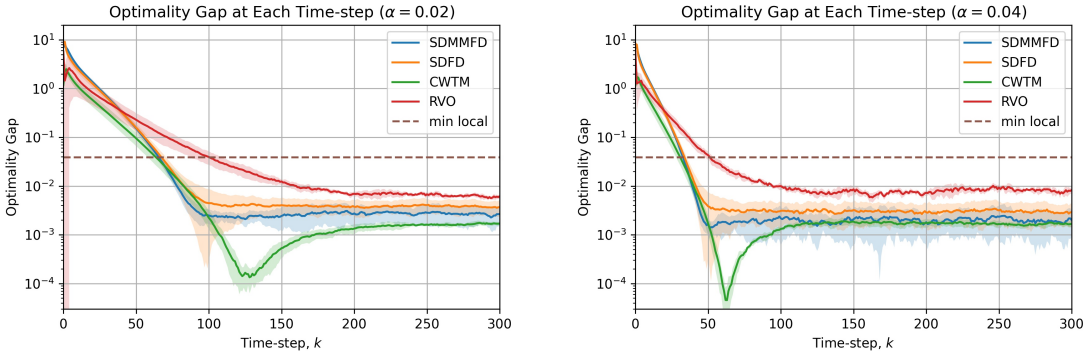
5.7 Numerical Experiment

We now provide a numerical experiment to illustrate the behavior of the algorithms discussed in Section 5.5, SDMMFD, SDFD, CWTM and RVO. In the experiment, we consider quadratic functions with two independent variables as the local cost functions. We choose the number of agents in the network to be 40 and construct an 11-robust graph. We consider the F-local adversary model with $F = 2$. Each Byzantine agent sends to each regular neighboring agent, a random vector which is close to the other vectors received by the regular agent. For all the algorithms, we set the constant step-size to be $\alpha = 0.02$ or 0.04. Fixing the network, local functions and Byzantine agents, we re-run the experiment for each algorithm 5 times due to the randomness of the states initialization and adversarial behavior, and report the average and standard deviation over all runs of the metrics described below.

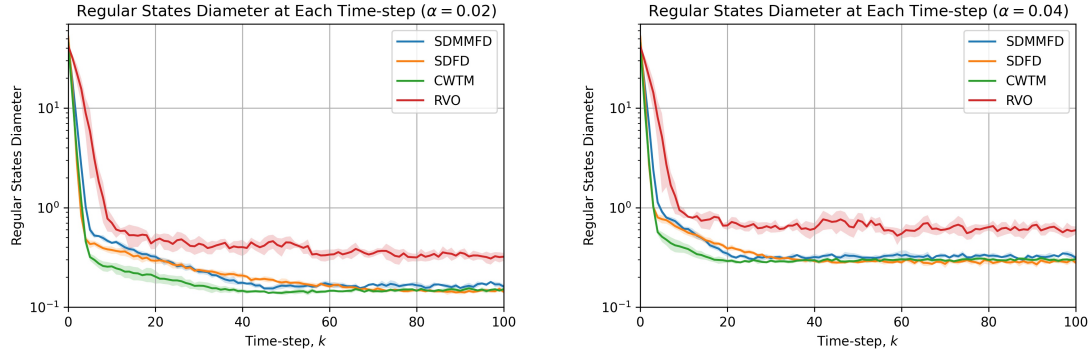
In Figure 5.2a, each solid curve corresponds to the Euclidean distance from the average of the regular agents' states to the true minimizer, i.e., $\|\bar{\boldsymbol{x}}[k] - \boldsymbol{x}^*\|$, and the dashed line labeled as "min_local" corresponds to the minimum over all regular agents of the Euclidean distance from the minimizers of the local functions to the true minimizer, i.e., $\min_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i^* - \boldsymbol{x}^*\|$. In Figure 5.2b, each solid curve corresponds to the optimality gap computed at the average of the regular agents' states, i.e., $f(\bar{\boldsymbol{x}}[k]) - f^*$ where $\bar{\boldsymbol{x}}[k] = \frac{1}{|\mathcal{V}_\mathcal{R}|} \sum_{v_i \in \mathcal{V}_\mathcal{R}} \boldsymbol{x}_i[k]$, and the



(a) The Euclidean distance from the average of the regular agents' states to the true minimizer $\|\bar{\boldsymbol{x}} - \boldsymbol{x}^*\|$ for each algorithm.



(b) The optimality gap evaluated at the average of the regular agents' states $f(\bar{\boldsymbol{x}}) - f^*$ for each algorithm.



(c) The maximum Euclidean distance between two regular agents' states (regular states' diameter) $\max_{v_i, v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ for each algorithm.

Figure 5.2. The plots show the results obtained from SDMMFD (blue), SDFD (orange), CWTM (green), and RVO (red) for given constant step-sizes $\alpha = 0.02$ (left), and $\alpha = 0.04$ (right).

dashed line labeled as “min_local" corresponds to the minimum over all regular agents of the optimality gap computed at the minimizers of the local functions, i.e., $\min_{v_i \in \mathcal{V}_\mathcal{R}} f(\boldsymbol{x}_i^*) - f^*$. In Figure 5.2c, each solid curve corresponds to the maximum Euclidean distance between two regular agents' states (or regular agents' diameter), i.e., $\max_{v_i, v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\|$. In Figures 5.2a, 5.2b and 5.2c, the solid curves are the means over all the experiment rounds and the shaded regions represent ± 1 standard deviation from the means.

As we can see from Figure 5.2a, for both cases, $\alpha = 0.02$ and $\alpha = 0.04$, the distances to the true minimizer drop at geometric rates in the early time-steps. However, in the later time-steps, the distances to the true minimizer hardly change (but oscillate with small magnitudes) as the regular nodes' states have entered the convergence region. In addition, the convergence rates of all algorithms increase as the constant step-size α changes from 0.02 to 0.04 (as predicted by Theorem 5.6.4). Still, for each algorithm, the distances to the minimizer at later time-steps are comparable for both constant step-sizes. The optimality gaps shown in Figure 5.2b exhibit similar behaviors as the distance to the minimizer in Figure 5.2a. For Figure 5.2c, note that the plots show a shorter horizon for the time-step k. We can observe that the diameters drop at geometric rates in the early time-steps and then hardly change after that. In addition, as the constant step-size α changes from 0.02 to 0.04, the convergence rates of SDMMFD, SDFD and CWTM increase (as shown in Corollary 5.6.7) and the diameters of all the algorithms at the later time-steps increase (as discussed in Section 5.6.3). Note that even though in general, from Theorem 5.6.8, the contraction factor of CWTM is $\gamma = d$ (and in this case $d = 2$), for most of the time-steps, in this experiment, the states contraction property (5.8) for CWTM holds with $\gamma = 1$.

From Figures 5.2a and 5.2b, in fact, we observe that the results from the algorithms are usually better than the best value achieved by the local minimizers (i.e., comparing the solid curves to the dashed line) even though all the local minimizers are inside the convergence region $\mathcal{B}(\boldsymbol{x}_c, R^*)$. Besides, comparing the algorithms, SDMMFD, SDFD and CWTM achieve comparable mean performance for both convergence rate and final value for all the metrics but RVO performs worse than the other three algorithms in this case. Also, SDMMFD and SDFD are more sensitive to the adversarial behavior than CWTM and RVO as we can see from their standard deviations in Figures 5.2a and 5.2b.

5.8 Conclusions

In this chapter, we considered the (peer-to-peer) distributed optimization problem in the presence of Byzantine agents. We introduced a general resilient (peer-to-peer) distributed gradient descent framework called RedGraf which includes some state-of-the-art resilient algorithms such as SDMMFD, SDFD (i.e., Algorithms 2 and 3) [93], CWTM [30], [33], [45], [50]–[52], and RVO [74], [77] as special cases. We analyzed the convergence of algorithms captured by our framework, assuming they satisfy a certain states contraction property. In particular, we derived a geometric rate of convergence of all regular agents to the convergence region under the strong convexity of the local functions and constant step-size regime. As we have shown, the convergence region, in fact, contains the true minimizer (the minimizer of the sum of the regular agents' functions). In addition, given a mixing dynamics property, we also derived a geometric rate of convergence of all regular agents to approximate consensus with a certain diameter under similar conditions. Considering each resilient algorithm, we analyzed the states contraction and mixing dynamics properties which in turn, dictate the convergence rates, the size of the convergence region and the approximate consensus diameter.

Future work includes developing resilient algorithms satisfying both the states contraction and mixing dynamics properties which give fast rates of convergences as well as a small convergence region and small approximate consensus diameter, identifying other properties for resilient algorithms to achieve good performance, analyzing the convergence property of other existing algorithms from the literature, considering non-convex functions with certain properties, and establishing a tight lower bound for the convergence region.

6. SUMMARY AND FUTURE WORK

6.1 Summary

In this dissertation, we delved into the challenges of resilient distributed optimization, where a network of benign agents collaboratively optimizes the sum of local functions while contending with Byzantine agents that can behave arbitrarily and disrupt the optimization process. To address this issue, we adopted the notion of unknown functions with shared properties for Byzantine agents. In Chapters 2 and 3, we focused on characterizing the potential solution region, which represents the feasible region for the minimizer of the sum. In Chapter 2, we specifically considered the sum of two strongly convex functions and obtained the closed-form equation of the boundary of the potential solution region (Theorem 2.6.1). In Chapter 3, we explored the case where only one function is unknown, and we established necessary conditions (Theorems 3.4.1 and 3.5.1) for a point to be a minimizer, along with an algorithm (Algorithm 1) to determine the region satisfying these conditions.

Shifting our focus to resilient optimization algorithms in Chapters 4 and 5, one fundamental limitation is that it is not possible to determine the optimal solution under Byzantine behavior. Therefore, we adopted alternative criteria to measure algorithm performance, aiming for solutions reasonably close to the optimal solution. In Chapter 4, we proposed two resilient optimization algorithms, Algorithms 2 and 3, based on distance-based and extreme value filtering techniques, applicable to multi-variate functions. Under convexity and bounded gradient assumptions, for Algorithm 2, we derived asymptotic convergence and consensus theorems (Theorems 4.4.9 and 4.4.7, respectively), while Algorithm 3 provides asymptotic convergence (Theorem 4.4.9) with significantly less redundancy, thus enabling more scalability.

In Chapter 5, we introduced an algorithmic framework (Algorithmic Framework 1) that encompasses several state-of-the-art resilient algorithms, including Algorithms 2 and 3. We analyzed the convergence rate and identified the convergence region under strong convexity and smoothness assumptions. Essentially, leveraging the novel concept of states contraction, we established geometric convergence of several resilient algorithms towards the

convergence region (Theorem 5.6.4), which guarantees to encompass the optimal solution (Theorem 5.6.5).

In conclusion, this dissertation makes significant contributions to the field of resilient distributed optimization by providing insights into the potential solution region, proposing resilient optimization algorithms with theoretical guarantees, and analyzing their convergence properties. These findings contribute to advancing the understanding and development of robust and scalable optimization algorithms in the presence of Byzantine agents.

6.2 Future Work

In this section, we list some potential research directions. Some of them may be jointly considered.

(i) **Analysis of the minimizer of the sum of multiple strongly convex functions:** In the first part of this work (Chapter 2), we considered the region containing the minimizer of the sum of two strongly convex functions. In the case of two functions, we obtained the explicit characterization of the region that contains the minimizer by considering the necessary conditions (Section 2.5) and the sufficient conditions (Section 2.4) separately, and noticed the tightness of inner and outer approximations. However, in the case of multiple functions, it might not be possible to use the same techniques as in Chapter 2 and the resulted regions might be more under-estimated or over-estimated.

(ii) **Analysis of the minimizer of the sum of two strongly convex functions with Lipschitz continuous gradient:** In the first part of this work (Chapter 2), we have considered the region containing the minimizer of the sum of two strongly convex functions where the gradient norm of both functions at the potential minimizer is bounded by a specified constant. This condition, in particular, is not a common characterization of a strongly convex function and instead, a more popular characterization is *Lipschitz continuous gradient* (or *smoothness*) [81]. It is of interest to characterize the region containing the minimizer of the sum of two strongly convex functions given this condition.

(iii) **Byzantine-resilient distributed optimization algorithm with previous states exploitation:** In the third part of this work (Chapter 4), we have provided two algorithms: Simultaneous Distance-MinMax Filtering Dynamics (Algorithm 2) and Simultaneous Distance Filtering Dynamics (Algorithm 3). In the proposed algorithms, in each time-step, each regular agent considers only its (main and auxiliary) states and its neighboring states at that time-step. It might be possible to design an algorithm that can leverage the knowledge of the states of the previous time-steps to obtain better performance or even better theoretical guarantees.

(iv) **Byzantine-resilient distributed optimization algorithm with gradient tracking:** In the third part of this work (Chapter 4), the two provided algorithms are constructed based on *Distributed Gradient Descent (DGD)* algorithm [70] which is one of the most simplest distributed algorithm with a convergence guarantee. In the distributed optimization literature, one of the most popular approaches to improve the convergence rate in the deterministic setting is to use the technique called *gradient tracking* [80]. It might be of interest to modify an algorithm with gradient tracking to be resilient to Byzantine agents while maintaining the linear-rate convergence guarantee.

(v) **Resilient distributed stochastic optimization:** We have considered resilient distributed (deterministic) optimization problem in Chapters 4 and 5. However, popular applications such as machine learning problems are framed as stochastic optimization problems. Furthermore, safety training of a machine learning model is one of the most important requirements for real-world applications [84]. Hence, it is of interest to study the distributed stochastic optimization in the presence of adversarial agents and provide a resilient algorithm that is invulnerable to the attack.

(vi) **Lower bound on the convergence radius:** In both Chapters 4 and 5, we have provided the convergence radii R^* for resilient optimization algorithms, and showed that the convergence regions contains the true minimizer. In particular, we conclude that the convergence radius $R^* = \mathcal{O}(r^*)$ where r^* is the radius of a ball containing

the local minimizers of the regular agents, is typical in the literature [30], [33] without extra assumptions. However, a tight lower bound on the convergence radius for the general case (whether it is $\Omega(r^*)$) remains an open problem.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[2] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice Hall Upper Saddle River, NJ, 2000, vol. 1.

[3] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, "Constrained model predictive control: Stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[4] A. E. Bryson, *Applied optimal control: optimization, estimation and control*. Routledge, 2018.

[5] G. C. Calafiore and L. Fagiano, "Robust model predictive control via scenario optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 219–224, 2013.

[6] M. Zhu and S. Martínez, *Distributed optimization-based control of multi-agent networks in complex environments*. Springer, 2015.

[7] E. Montijano and A. Mosteo, "Efficient multi-robot formations using distributed optimization," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 6167–6172.

[8] K. Shin and N. McKay, "Minimum-time control of robotic manipulators with geometric path constraints," *IEEE Transactions on Automatic Control*, vol. 30, no. 6, pp. 531–541, 1985.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[10] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[12] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed ADMM via dual averaging," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 904–909.

[13] J. N. Tsitsiklis, "Problems in decentralized decision making and computation.," Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems, Tech. Rep., 1984.

[14] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice hall Englewood Cliffs, NJ, 1989, vol. 23.

[15] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[16] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, 2014.

[17] S. Hosseinalipour, C. G. Brinton, V. Aggarwal, H. Dai, and M. Chiang, "From federated to fog learning: Distributed machine learning over heterogeneous wireless networks," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 41–47, 2020.

[18] D. K. Molzahn, F. Dörfler, H. Sandberg, *et al.*, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017.

[19] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *IEEE Power and Energy Society General Meeting*, 2011, pp. 1–8.

[20] J. Cai, D. Kim, R. Jaramillo, J. E. Braun, and J. Hu, "A general multi-agent control approach for building energy system optimization," *Energy and Buildings*, vol. 127, pp. 337–351, 2016.

[21] M. M. Zavlanos, A. Ribeiro, and G. J. Pappas, "Network integrity in mobile robotic networks," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 3–18, 2012.

[22] R. Tron, J. Thomas, G. Loianno, K. Daniilidis, and V. Kumar, "A distributed optimization framework for localization and formation control: Applications to vision-based measurements," *IEEE Control Systems Magazine*, vol. 36, no. 4, pp. 22–44, 2016.

[23] A. Nedi, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[24] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157–1170, 2010.

[25] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2011.

[26] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2013.

[27] A. Nedi and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2014.

[28] J. Xu, Y. Tian, Y. Sun, and G. Scutari, "Accelerated primal-dual algorithms for distributed smooth convex optimization over networks," in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2020, pp. 2381–2391.

[29] T. Yang, X. Yi, J. Wu, *et al.*, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[30] S. Sundaram and B. Gharesifard, "Distributed optimization under adversarial nodes," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1063–1076, 2018.

[31] N. Ravi, A. Scaglione, and A. Nedi, "A case of distributed optimization in adversarial environment," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 5252–5256.

[32] S. X. Wu, H.-T. Wai, A. Scaglione, A. Nedi, and A. Leshem, "Data injection attack on decentralized optimization," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2018, pp. 3644–3648.

[33] L. Su and N. Vaidya, *Byzantine multi-agent optimization: Part i*, 2015. arXiv: 1506.04681 [cs.DC].

[34] L. Su and N. H. Vaidya, "Fault-tolerant multi-agent optimization: Optimal iterative distributed algorithms," in *Proceedings of the 2016 ACM symposium on principles of distributed computing*, 2016, pp. 425–434.

[35] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of Operations Research*, vol. 23, no. 4, pp. 769–805, 1998.

[36] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM Review*, vol. 53, no. 3, pp. 464–501, 2011.

[37] H. Jiang and U. V. Shanbhag, "On the solution of stochastic optimization and variational problems in imperfect information regimes," *SIAM Journal on Optimization*, vol. 26, no. 4, pp. 2394–2429, 2016.

[38] J. Konený, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, *Federated learning: Strategies for improving communication efficiency*, 2017. arXiv: 1610.05492 [cs.LG].

[39] P. Kairouz, H. B. McMahan, B. Avent, *et al.*, "Advances and open problems in federated learning," *Foundations and Trendső in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[40] K. Kuwaranancharoen, L. Xin, and S. Sundaram, "Byzantine-resilient distributed optimization of multi-dimensional functions," in *2020 American Control Conference (ACC)*, IEEE, 2020, pp. 4399–4404.

[41] C. Zhao, J. He, and Q.-G. Wang, "Resilient distributed optimization algorithm against adversary attacks," in *IEEE International Conference on Control & Automation (ICCA)*, 2017, pp. 473–478.

[42] N. Gupta and N. H. Vaidya, *Byzantine fault tolerant distributed linear regression*, 2019. arXiv: 1903.08752 [cs.LG].

[43] P. Blanchard, R. Guerraoui, J. Stainer, *et al.*, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Advances in Neural Information Processing Systems*, 2017, pp. 119–129.

[44] Z. Yang and W. U. Bajwa, "Byrdie: Byzantine-resilient distributed coordinate descent for decentralized learning," *IEEE Transactions on Signal and Information Processing over Networks*, 2019.

[45] C. Fang, Z. Yang, and W. U. Bajwa, "Bridge: Byzantine-resilient decentralized gradient descent," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 610–626, 2022.

[46] S. Guo, T. Zhang, X. Xie, L. Ma, T. Xiang, and Y. Liu, "Towards byzantine-resilient learning in decentralized systems," *arXiv preprint arXiv:2002.08569*, 2020.

[47] N. Gupta, T. T. Doan, and N. H. Vaidya, "Byzantine fault-tolerance in decentralized optimization under 2f-redundancy," in *2021 American Control Conference (ACC)*, IEEE, 2021, pp. 3632–3637.

[48] N. Ravi and A. Scaglione, "Detection and isolation of adversaries in decentralized optimization for non-strongly convex objectives," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 381–386, 2019.

[49] S. Guo, T. Zhang, H. Yu, *et al.*, "Byzantine-resilient decentralized stochastic gradient descent," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[50] L. Su and N. H. Vaidya, "Byzantine-resilient multiagent optimization," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2227–2233, 2020.

[51] W. Fu, Q. Ma, J. Qin, and Y. Kang, "Resilient consensus-based distributed optimization under deception attacks," *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 1803–1816, 2021.

[52] C. Zhao, J. He, and Q.-G. Wang, "Resilient distributed optimization algorithm against adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4308–4315, 2019.

[53] S. Sundaram and B. Gharesifard, "Secure local filtering algorithms for distributed optimization," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, IEEE, 2016, pp. 1871–1876.

[54] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, 2015, pp. 1310–1321.

[55] K. Kuwaranancharoen and S. Sundaram, "On the location of the minimizer of the sum of two strongly convex functions," in *2018 IEEE Conference on Decision and Control (CDC)*, IEEE, 2018, pp. 1769–1774.

[56] K. Kuwarananchaoen and S. Sundaram, "On the set of possible minimizers of a sum of known and unknown functions," in *2020 American Control Conference (ACC)*, IEEE, 2020, pp. 106–111.

[57] D. Castano, V. E. Paksoy, and F. Zhang, "Angles, triangle inequalities, correlation matrices and metric-preserving and subadditive functions," *Linear Algebra and its Applications*, vol. 491, pp. 15–29, 2016.

[58] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[59] B. Biggio, B. Nelson, and P. Laskov, *Poisoning attacks against support vector machines*, 2013. arXiv: 1206.6389 [cs.LG].

[60] M. Mozaffari-Kermani, S. Sur-Kolay, A. Raghunathan, and N. K. Jha, "Systematic poisoning attacks on and defenses for machine learning in healthcare," *IEEE journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1893–1905, 2014.

[61] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE transactions on automatic control*, vol. 31, no. 9, pp. 803–812, 1986.

[62] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of optimization theory and applications*, vol. 129, no. 3, pp. 469–488, 2006.

[63] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *2011 50th IEEE conference on decision and control and European control conference*, IEEE, 2011, pp. 3800–3805.

[64] M. Eisen, A. Mokhtari, and A. Ribeiro, "Decentralized quasi-newton methods," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2613–2628, 2017.

[65] R. Xin, C. Xi, and U. A. Khan, "Frostfast row-stochastic optimization with uncoordinated step-sizes," *EURASIP Journal on Advances in Signal Processing*, vol. 2019, no. 1, p. 1, 2019.

[66] J. Zeng and W. Yin, "On nonconvex decentralized gradient descent," *IEEE Transactions on signal processing*, vol. 66, no. 11, pp. 2834–2848, 2018.

[67] K. Pillutla, S. M. Kakade, and Z. Harchaoui, "Robust aggregation for federated learning," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1142–1154, 2022.

[68] Z. Wu, T. Chen, and Q. Ling, *Byzantine-resilient decentralized stochastic optimization with robust aggregation rules*, 2023. arXiv: 2206.04568 [cs.DC].

[69] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 766–781, 2013.

[70] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[71] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Transactions on Automatic control*, vol. 57, no. 3, pp. 592–606, 2011.

[72] D. Jakoveti, J. Xavier, and J. M. Moura, "Fast distributed gradient methods," *IEEE Transactions on Automatic Control*, vol. 59, no. 5, pp. 1131–1146, 2014.

[73] M. Pirani, A. Mitra, and S. Sundaram, *A survey of graph-theoretic approaches for analyzing the resilience of networked control systems*, 2022. arXiv: 2205.12498 [eess.SY].

[74] W. Abbas, M. Shabbir, J. Li, and X. Koutsoukos, "Resilient distributed vector consensus using centerpoint," *Automatica*, vol. 136, p. 110 046, 2022.

[75] L. Guerrero-Bonilla, A. Prorok, and V. Kumar, "Formations for resilient robot teams," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 841–848, 2017.

[76] D. Saldana, A. Prorok, S. Sundaram, M. F. Campos, and V. Kumar, "Resilient consensus for time-varying networks of dynamic agents," in *2017 American control conference (ACC)*, IEEE, 2017, pp. 252–258.

[77] H. Park and S. A. Hutchinson, "Fault-tolerant rendezvous of multirobot systems," *IEEE transactions on robotics*, vol. 33, no. 3, pp. 565–582, 2017.

[78] J. Yan, Y. Mo, X. Li, L. Xing, and C. Wen, "Resilient vector consensus: An event-based approach," in *2020 IEEE 16th International Conference on Control & Automation (ICCA)*, IEEE, 2020, pp. 889–894.

[79] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[80] A. Nedic, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2597–2633, 2017.

[81] S. Pu, W. Shi, J. Xu, and A. Nedi, "Push–pull gradient methods for distributed optimization in networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 1, pp. 1–16, 2020.

[82] A. Nedi, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.

[83] R. Xin, S. Pu, A. Nedi, and U. A. Khan, "A general framework for decentralized optimization with first-order methods," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 1869–1889, 2020.

[84] Z. Yang, A. Gang, and W. U. Bajwa, "Adversary-resilient distributed and decentralized statistical inference and machine learning: An overview of recent advances under the byzantine threat model," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 146–159, 2020.

[85] J. Peng, W. Li, and Q. Ling, "Byzantine-robust decentralized stochastic optimization over static and time-varying networks," *Signal Processing*, vol. 183, p. 108 020, 2021.

[86] W. Ben-Ameur, P. Bianchi, and J. Jakubowicz, "Robust distributed consensus using total variation," *IEEE Transactions on Automatic Control*, vol. 61, no. 6, pp. 1550–1564, 2015.

[87] A. R. Elkordy, S. Prakash, and S. Avestimehr, "Basil: A fast and Byzantine-resilient approach for decentralized training," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2694–2716, 2022.

[88] L. He, S. P. Karimireddy, and M. Jaggi, *Byzantine-robust decentralized learning via clippedgossip*, 2023. arXiv: 2202.01545 [cs.LG].

[89] M. Cao, A. S. Morse, and B. D. Anderson, "Reaching a consensus in a dynamically changing environment: A graphical approach," *SIAM Journal on Control and Optimization*, vol. 47, no. 2, pp. 575–600, 2008.

[90] N. A. Lynch, *Distributed algorithms*. Elsevier, 1996.

[91] D. Yin, Y. Chen, R. Kannan, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *International Conference on Machine Learning*, PMLR, 2018, pp. 5650–5659.

[92] S. Farhadkhani, R. Guerraoui, N. Gupta, R. Pinot, and J. Stephan, "Byzantine machine learning made easy by resilient averaging of momentums," in *International Conference on Machine Learning*, PMLR, 2022, pp. 6246–6283.

[93] K. Kuwarananchaeroen, L. Xin, and S. Sundaram, *Scalable distributed optimization of multi-dimensional functions despite byzantine adversaries*, 2022. arXiv: 2003.09038 [math.OC].

[94] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM Journal on Optimization*, vol. 26, no. 3, pp. 1835–1854, 2016.

[95] H. Hendrikx, F. Bach, and L. Massoulié, "Accelerated decentralized optimization with local updates for smooth and strongly convex objectives," in *The 22nd International Conference on Artificial Intelligence and Statistics*, PMLR, 2019, pp. 897–906.

[96] D. Kovalev, A. Koloskova, M. Jaggi, P. Richtarik, and S. Stich, "A linearly convergent algorithm for decentralized optimization: Sending less bits for free!" In *International Conference on Artificial Intelligence and Statistics*, PMLR, 2021, pp. 4087–4095.

[97] B. S. Mordukhovich and N. M. Nam, "An easy path to convex analysis and applications," *Synthesis Lectures on Mathematics and Statistics*, vol. 6, no. 2, pp. 1–218, 2013.

[98] X. Zhou, *On the fenchel duality between strong convexity and lipschitz continuous gradient*, 2018. arXiv: 1803.06573 [math.OC].

A. SUPPLEMENTARY MATERIALS FOR CHAPTER 2

A.1 Proofs of Theoretical Results for Outer Approximation

A.1.1 Proof of Lemma 2.4.1

Proof of Lemma 2.4.1. Consider part (i) of the lemma. Suppose $\boldsymbol{x}_\epsilon = \boldsymbol{x}_1^* + \epsilon \boldsymbol{e}_1$ for $\epsilon \in \mathbb{R}$. We want to show that

(a) If $r \in \left(0,\ \frac{L}{2\sigma_2}\right]$ then for all $\epsilon \in \left(0,\ \min\left\{\frac{L}{\sigma_1}, 2r\right\}\right)$, we have $\tilde{\phi}_1(\boldsymbol{x}_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}_\epsilon) > \psi(\boldsymbol{x}_\epsilon)$.

(b) If $r \in \left(0,\ \frac{L}{2\sigma_2}\right)$ then for all $\epsilon \in \left(-\min\left\{\frac{L}{\sigma_1}, \frac{L}{\sigma_2} - 2r\right\},\ 0\right)$, we have $\tilde{\phi}_1(\boldsymbol{x}_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}_\epsilon) < \psi(\boldsymbol{x}_\epsilon)$.

(c) If $r = \frac{L}{2\sigma_2}$ then for all $\epsilon \in (-\infty, 0)$, we have $\boldsymbol{x}_\epsilon \notin (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$.

First, consider case (a). Suppose $\epsilon \in \left(0,\ \min\left\{\frac{L}{\sigma_1}, 2r\right\}\right)$. Let $x_{\epsilon,1}$ be the first component of the point $\boldsymbol{x}_\epsilon$. We have $x_{\epsilon,1} \in \left(-r,\ \min\left\{-r + \frac{L}{\sigma_1},\ r\right\}\right)$. Since $r \in \left(0,\ \frac{L}{2\sigma_2}\right]$, we have $-r \in \left[r - \frac{L}{\sigma_2}, r\right)$. Therefore, $\boldsymbol{x}_\epsilon \in (\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. By the location of $\boldsymbol{x}_\epsilon$, from (2.10), we get $\alpha_1(\boldsymbol{x}_\epsilon) = 0$ and $\alpha_2(\boldsymbol{x}_\epsilon) = \boldsymbol{\pi}$. Consequently, from (2.11), we obtain $\psi(\boldsymbol{x}_\epsilon) = 0$. Since $\boldsymbol{x}_\epsilon \notin \partial\mathcal{B}_1 \cup \partial\mathcal{B}_2$, from (2.9), we have that $\tilde{\phi}_i(\boldsymbol{x}_\epsilon) \in \left(0, \frac{\pi}{2}\right]$ for $i \in \{1, 2\}$. This means that $\tilde{\phi}_1(\boldsymbol{x}_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}_\epsilon) > \psi(\boldsymbol{x}_\epsilon)$.

Second, consider case (b). Suppose $\epsilon \in \left(-\min\left\{\frac{L}{\sigma_1}, \frac{L}{\sigma_2} - 2r\right\},\ 0\right)$. We have $x_{\epsilon,1} \in \left(\max\left\{-r - \frac{L}{\sigma_1},\ r - \frac{L}{\sigma_2}\right\},\ -r\right)$ which implies that $\boldsymbol{x}_\epsilon \in (\mathcal{B}_1 \cap \mathcal{B}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. By the location of $\boldsymbol{x}_\epsilon$, from (2.10), we get $\alpha_1(\boldsymbol{x}_\epsilon) = \boldsymbol{\pi}$ and $\alpha_2(\boldsymbol{x}_\epsilon) = \boldsymbol{\pi}$. Consequently, from (2.11), we obtain $\psi(\boldsymbol{x}_\epsilon) = \boldsymbol{\pi}$. Since $\boldsymbol{x}_\epsilon \notin \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, from (2.9), we have that $\tilde{\phi}_i(\boldsymbol{x}_\epsilon) \in \left[0, \frac{\pi}{2}\right)$ for $i \in \{1, 2\}$. This means that $\tilde{\phi}_1(\boldsymbol{x}_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}_\epsilon) < \psi(\boldsymbol{x}_\epsilon)$.

Third, consider case (c). Suppose $\epsilon \in (-\infty, 0)$. If $\boldsymbol{x} \in \overline{\mathcal{B}}_2$ then $x_1 \in \left[r - \frac{L}{\sigma_2}, r + \frac{L}{\sigma_2}\right] = [-r, 3r]$. However, we have $x_{\epsilon,1} \in (-\infty, -r)$. This means that $\boldsymbol{x}_\epsilon \notin \overline{\mathcal{B}}_2$ which implies that $\boldsymbol{x}_\epsilon \notin (\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2) \setminus \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. We complete the proof of our claim.

Recall the definition of $\mathcal{M}^\uparrow$ and $\mathcal{M}_\downarrow$ in (2.13) and (2.14), respectively. From our claim, we can see that if $r \in \left(0,\ \frac{L}{2\sigma_2}\right]$ then for all $\delta \in \mathbb{R}_{>0}$, there exist $\boldsymbol{x}_{\text{in}}, \boldsymbol{x}_{\text{out}} \in \mathcal{B}(\boldsymbol{x}_1^*, \delta)$ such that $\boldsymbol{x}_{\text{in}} \in \mathcal{M}^\uparrow$, $\boldsymbol{x}_{\text{in}} \in \mathcal{M}_\downarrow$, $\boldsymbol{x}_{\text{out}} \notin \mathcal{M}^\uparrow$ and $\boldsymbol{x}_{\text{out}} \notin \mathcal{M}_\downarrow$. This implies that $\boldsymbol{x}_1^* \in \partial\mathcal{M}^\uparrow$ and $\boldsymbol{x}_1^* \in \partial\mathcal{M}_\downarrow$. The analysis for part (ii) of the lemma follows in an identical manner. □

A.1.2 Proof of Lemma 2.4.2

Proof of Lemma 2.4.2. Recall from (2.10) that $\alpha_i(\boldsymbol{x}) = \angle(\boldsymbol{x} - \boldsymbol{x}_i^*,\ \boldsymbol{x}_2^* - \boldsymbol{x}_1^*)$ for $i \in \{1, 2\}$. Suppose $\boldsymbol{y} \notin \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$. Since $-r \le x_1 = y_1 \le r$ and $\|\tilde{\mathbf{x}}\| > \|\tilde{\mathbf{y}}\|$, we have $\alpha_1(\boldsymbol{x}) \ge \alpha_1(\boldsymbol{y})$ and $\alpha_2(\boldsymbol{x}) \le \alpha_2(\boldsymbol{y})$, which implies that

$$\psi(\boldsymbol{x}) = \boldsymbol{\pi} - (\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})) \ge \boldsymbol{\pi} - (\alpha_2(\boldsymbol{y}) - \alpha_1(\boldsymbol{y})) = \psi(\boldsymbol{y}). \tag{A.1}$$

On the other hand, since $x_1 = y_1$ and $\|\tilde{\mathbf{x}}\| > \|\tilde{\mathbf{y}}\|$, we get

$$\|\boldsymbol{x} - \boldsymbol{x}_1^*\| > \|\boldsymbol{y} - \boldsymbol{x}_1^*\| > 0 \quad \text{and} \quad \|\boldsymbol{x} - \boldsymbol{x}_2^*\| > \|\boldsymbol{y} - \boldsymbol{x}_2^*\| > 0.$$

Using the above inequalities and the definition of (2.9), we get that $\tilde{\phi}_i(\boldsymbol{x}) < \tilde{\phi}_i(\boldsymbol{y})$ for $i \in \{1, 2\}$. Applying $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \ge \psi(\boldsymbol{x})$ and inequality (A.1), we can write

$$\psi(\boldsymbol{y}) \le \psi(\boldsymbol{x}) \le \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) < \tilde{\phi}_1(\boldsymbol{y}) + \tilde{\phi}_2(\boldsymbol{y}),$$

which completes the proof. □

A.1.3 Proof of Lemma 2.4.3

Proof of Lemma 2.4.3. Consider part (i) of the lemma. Since $\boldsymbol{x} \in \partial\mathcal{B}_1$, we get $\|\boldsymbol{x} - \boldsymbol{x}_1^*\| = \frac{L}{\sigma_1}$ and thus $\tilde{\phi}_1(\boldsymbol{x}) = 0$ from (2.9). Consider the inequality $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) \lesseqgtr \psi(\boldsymbol{x})$. Substitute $\tilde{\phi}_1(\boldsymbol{x}) = 0$ and $\psi(\boldsymbol{x}) = \boldsymbol{\pi} - (\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x}))$, and take cosine of both sides of the inequality (and use $\cos\tilde{\phi}_2(\boldsymbol{x}) = \frac{\sigma_2}{L} d_2(\boldsymbol{x})$ from (2.9)) to get

$$\frac{\sigma_2}{L} d_2(\boldsymbol{x}) \gtreqless -\cos\Big(\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\Big).$$

Expand the cosine and substitute the equations (2.23) and (2.24) to obtain

$$\frac{\sigma_2}{L} d_2(\boldsymbol{x}) \gtreqless -\frac{x_1^2 + \|\tilde{\mathbf{x}}\|^2 - r^2}{d_1(\boldsymbol{x}) \cdot d_2(\boldsymbol{x})}. \tag{A.2}$$

Since $\boldsymbol{x} \in \partial\mathcal{B}_1$, we have $d_1(\boldsymbol{x}) = \frac{L}{\sigma_1}$ and $\|\tilde{\mathbf{x}}\|^2 = \frac{L^2}{\sigma_1^2} - (x_1 + r)^2$ from (2.25). Also, from (2.21), we have

$$d_2^2(\boldsymbol{x}) = (x_1 - r)^2 + \|\tilde{\mathbf{x}}\|^2 = (x_1 - r)^2 + \frac{L^2}{\sigma_1^2} - (x_1 + r)^2 = -4rx_1 + \frac{L^2}{\sigma_1^2}. \tag{A.3}$$

Multiply inequality (A.2) by $d_1(\boldsymbol{x}) \cdot d_2(\boldsymbol{x})$ and then substitute $d_1(\boldsymbol{x})$, $\|\tilde{\mathbf{x}}\|^2$, and $d_2^2(\boldsymbol{x})$ to get

$$\begin{aligned}
&\frac{\sigma_2}{\sigma_1}\Big(- 4rx_1 + \frac{L^2}{\sigma_1^2}\Big) \gtrless 2r^2 + 2rx_1 - \frac{L^2}{\sigma_1^2}, \\
\Leftrightarrow \quad & x_1\Big(2r + 4r\frac{\sigma_2}{\sigma_1}\Big) \lessgtr \frac{\sigma_2}{\sigma_1} \cdot \frac{L^2}{\sigma_1^2} + \frac{L^2}{\sigma_1^2} - 2r^2, \\
\Leftrightarrow \quad & x_1 \lessgtr \Big(\frac{1+\beta}{1+2\beta}\Big)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} = \lambda_1,
\end{aligned}$$

where γ_1 and β are defined in (2.26). The proof of the second part is similar to the first part. □

A.1.4 Proof of Lemma 2.4.4

Proof of Lemma 2.4.4. We will prove the case that $i = 1$; however, the proof of the case that $i = 2$ can be obtained using the same approach. Suppose $\boldsymbol{x} \in \partial\mathcal{B}_1$. Substituting $\|\tilde{\mathbf{x}}\|^2 = \frac{L^2}{\sigma_1^2} - (x_1 + r)^2$, $d_1(\boldsymbol{x}) = \frac{L}{\sigma_1}$ and (A.3) into the expression of $\mathcal{T}$ in (2.22), we get

$$\frac{x_1^2 + \|\tilde{\mathbf{x}}\|^2 - r^2}{d_1^2(\boldsymbol{x}) \cdot d_2^2(\boldsymbol{x})} + \frac{\sigma_1\sigma_2}{L^2} = \frac{\frac{L^2}{\sigma_1^2} - 2rx_1 - 2r^2}{\frac{L^2}{\sigma_1^2}\Big(- 4rx_1 + \frac{L^2}{\sigma_1^2}\Big)} + \frac{\sigma_1\sigma_2}{L^2} = 0. \tag{A.4}$$

Multiply both sides of the above equation by $\frac{L^2}{\sigma_1^2}\Big(- 4rx_1 + \frac{L^2}{\sigma_1^2}\Big)$ and then use the definition of γ_1 and β in (2.26) to get

$$4\beta r x_1 - \beta\gamma_1 = \gamma_1 - 2rx_1 - 2r^2, \quad \Leftrightarrow \quad x_1 = \Big(\frac{1+\beta}{1+2\beta}\Big)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} = \lambda_1.$$

Then, substituting $x_1 = \left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta}$ back into the expression of $\partial\mathcal{B}_1$ in (2.25), we get

$$\begin{aligned}\|\tilde{\mathbf{x}}\|^2 &= \gamma_1 - \left[\left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} + r\right]^2 \\ &= -\frac{1}{4r^2(1+2\beta)^2}\Big[(\gamma_1 - 4r^2)(\gamma_1(1+\beta)^2 - 4\beta^2 r^2)\Big], \\ \Rightarrow \quad \|\tilde{\mathbf{x}}\| &= \frac{r}{2(1+2\beta)}\sqrt{-\left(\frac{\gamma_1}{r^2} - 4\right)\left(\frac{\gamma_1}{r^2}(1+\beta)^2 - 4\beta^2\right)} = \nu_1.\end{aligned}$$

This equation is valid only when the term in the square root is a non-negative real number. Equivalently,

$$\frac{4\beta^2}{(1+\beta)^2} \le \frac{\gamma_1}{r^2} \le 4, \quad \Leftrightarrow \quad \frac{L}{2\sigma_1} \le r \le \frac{L}{2\sigma_1}\left(1 + \frac{1}{\beta}\right) = \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right).$$

Since we multiplied (A.4) by $d_1^2(\boldsymbol{x}) \cdot d_2^2(\boldsymbol{x}) = \frac{L^2}{\sigma_1^2}\big(- 4rx_1 + \frac{L^2}{\sigma_1^2}\big)$ to obtain the result, $\boldsymbol{x}_1^*$ or $\boldsymbol{x}_2^*$ might appear in the intersection $\mathcal{T} \cap \partial\mathcal{B}_1$ even if $\{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\} \not\subseteq \mathcal{T}$. Therefore, we need to verify that the intersection points are not $\boldsymbol{x}_1^*$ or $\boldsymbol{x}_2^*$. Considering the solution of the equation $\|\tilde{\mathbf{x}}\| = \nu_1 = 0$ (i.e., $\boldsymbol{x}$ is on the x_1-axis), we see that $r = \frac{L}{2\sigma_1}$ and $r = \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$ are the candidates that we need to check. Substituting $r = \frac{L}{2\sigma_1}$ into the equation $x_1 = \lambda_1$, we get $x_1 = \frac{L}{2\sigma_1} = r$ which is $\boldsymbol{x}_2^* = (r, \mathbf{0})$ and therefore we conclude that there are no intersection points when $r = \frac{L}{2\sigma_1}$. Next, substituting $r = \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$ into the equation $x_1 = \lambda_1$, we get $x_1 = \frac{L}{2}\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right)$ which is a legitimate value of an intersection point. □

A.1.5 Proof of Lemma 2.4.5

Proof of Lemma 2.4.5. To simplify notations in the proof, we define a new variable

$$\chi_1 := \frac{L}{\sigma_1 r} = \frac{\sqrt{\gamma_1}}{r}. \tag{A.5}$$

First, consider part (ii) of the lemma. By the definition of λ_1 in (2.27), we can rewrite the inequality $\lambda_1 < \frac{L}{\sigma_1} - r$ as $\left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r^2} - \frac{1}{1+2\beta} < \frac{L}{\sigma_1 r} - 1$. Using χ_1 defined in (A.5), the inequality becomes

$$\frac{1}{2}\left(\frac{1+\beta}{1+2\beta}\right)\chi_1^2 - \frac{1}{1+2\beta} < \chi_1 - 1. \tag{A.6}$$

Multiplying both sides by $1+2\beta$ and then rearranging the resulting inequality, we get

$$(\chi_1 - 2)\Big(\frac{1}{2}(1+\beta)\chi_1 - \beta\Big) < 0, \quad \Leftrightarrow \quad \frac{2\beta}{1+\beta} < \chi_1 < 2.$$

The last equivalence holds since $\frac{2\beta}{1+\beta} < 2$. Substituting $\chi_1 = \frac{L}{\sigma_1 r}$ and $\beta = \frac{\sigma_2}{\sigma_1}$, the inequality $\frac{2\beta}{1+\beta} < \chi_1$ becomes $r < \frac{L}{2}\big(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\big)$ and the inequality $\chi_1 < 2$ becomes $r > \frac{L}{2\sigma_1}$, which completes the proof of part (ii).

Next, we can see that part (ii) of the lemma implies that if $r \in \big(0,\ \frac{L}{2\sigma_1}\big]$ then $\lambda_1 \geq \frac{L}{\sigma_1} - r$. By proceeding the algebraic simplification similar to (A.6) except using equality instead, we can conclude that given $r \in \big(0,\ \frac{L}{2\sigma_1}\big]$, we have $\lambda_1 = \frac{L}{\sigma_1} - r$ only when $r = \frac{L}{2\sigma_1}$. This completes the proof of part (i).

The proof of part (iii) and part (iv) can be carried out using similar steps as the proof given above. □

A.1.6 Proof of Lemma 2.4.6

Proof of Lemma 2.4.6. Let $\boldsymbol{x} = (x_1, \tilde{\mathbf{x}}) \in \mathcal{T}$, and $\boldsymbol{e}_2'(\boldsymbol{x}) = \frac{\boldsymbol{v}(\boldsymbol{x})}{\|\boldsymbol{v}(\boldsymbol{x})\|}$ where $\boldsymbol{v}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}_1^*) - \langle \boldsymbol{x} - \boldsymbol{x}_1^*, \boldsymbol{e}_1\rangle\, \boldsymbol{e}_1$. Suppose $\theta(\boldsymbol{x}) = \frac{1}{2}(\alpha_1(\boldsymbol{x}) + \alpha_2(\boldsymbol{x}))$ where $\alpha_i(\boldsymbol{x})$ for $i \in \{1,2\}$ are defined in (2.10) and

$$\boldsymbol{x}_\epsilon' = \boldsymbol{x} + \epsilon\Big(\cos(\theta(\boldsymbol{x}))\ \boldsymbol{e}_1 + \sin(\theta(\boldsymbol{x}))\ \boldsymbol{e}_2'(\boldsymbol{x})\Big) \quad \text{for} \quad \epsilon \in \mathbb{R}. \tag{A.7}$$

Our claim is that if $r \in \big(0,\ \frac{L}{2}\big(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\big)\big)$ then

(i) for all

$$\epsilon \in \left(\max\left\{\max_{i=1,2} -2d_i(\boldsymbol{x}) \sin\frac{1}{2}\Big(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})\Big),\ -\|\tilde{\mathbf{x}}\|\csc(\theta(\boldsymbol{x}))\right\},\ 0\right),$$

it holds that $\tilde{\phi}_1(\boldsymbol{x}_\epsilon') + \tilde{\phi}_2(\boldsymbol{x}_\epsilon') > \psi(\boldsymbol{x}_\epsilon')$ (i.e., $\boldsymbol{x}_\epsilon' \in \mathcal{M}^\uparrow$ and $\boldsymbol{x}_\epsilon' \in \mathcal{M}_\downarrow$), and

(ii) for all $\epsilon \in (0, \infty)$, either $\tilde{\phi}_1(\boldsymbol{x}_\epsilon') + \tilde{\phi}_2(\boldsymbol{x}_\epsilon') < \psi(\boldsymbol{x}_\epsilon')$, or $\tilde{\phi}_1(\boldsymbol{x}_\epsilon')$ or $\tilde{\phi}_2(\boldsymbol{x}_\epsilon')$ is not well-defined (i.e., $\boldsymbol{x}_\epsilon' \in (\mathcal{M}^\uparrow)^c$ and $\boldsymbol{x}_\epsilon' \in (\mathcal{M}_\downarrow)^c$).

Recall from Proposition 2.4.2 that $\mathcal{T} = \Big\{\boldsymbol{z} \in \mathbb{R}^n : \tilde{\phi}_1(\boldsymbol{z}) + \tilde{\phi}_2(\boldsymbol{z}) = \psi(\boldsymbol{z})\Big\}$. First, we will show that $\boldsymbol{e}_2'(\boldsymbol{x})$ is well-defined, i.e., $\boldsymbol{v}(\boldsymbol{x}) \neq \mathbf{0}$ if $\boldsymbol{x} \in \mathcal{T}$. Note that $\boldsymbol{v}(\boldsymbol{x}) = \mathbf{0}$ if and only if

$\boldsymbol{x} = \boldsymbol{x}_1^* + k\boldsymbol{e}_1$ for some $k \in \mathbb{R}$. In fact, this is equivalent to $\boldsymbol{x} = (k', \mathbf{0})$ for some $k' \in \mathbb{R}$ since $\boldsymbol{x}_1^* = (-r, \mathbf{0})$. Suppose $\boldsymbol{x} = (x_1, \mathbf{0})$. Consider the following cases.

- Consider $x_1 \in (-\infty, -r) \cup (r, \infty)$. Then, we have $\psi(\boldsymbol{x}) = \boldsymbol{\pi}$ (since $\alpha_1(\boldsymbol{x}) = \alpha_2(\boldsymbol{x}) = \boldsymbol{\pi}$ when $x_1 \in (-\infty, r)$, and $\alpha_1(\boldsymbol{x}) = \alpha_2(\boldsymbol{x}) = 0$ when $x_1 \in (r, \infty)$). This implies that $\tilde{\phi}_1(\boldsymbol{x}) = \tilde{\phi}_2(\boldsymbol{x}) = \frac{\pi}{2}$ since $\tilde{\phi}_i \in \left[0, \frac{\pi}{2}\right]$ for $i \in \{1, 2\}$. Using the definition of $\tilde{\phi}_i$ for $i \in \{1, 2\}$ in (2.9), we obtain that $\boldsymbol{x} = \boldsymbol{x}_1^*$ and $\boldsymbol{x} = \boldsymbol{x}_2^*$ which contradicts the fact that $r = \frac{1}{2}\|\boldsymbol{x}_2^* - \boldsymbol{x}_1^*\| > 0$.

- Consider $x_1 \in (-r, r)$. Then, we have $\psi(\boldsymbol{x}) = 0$ (since $\alpha_1(\boldsymbol{x}) = 0$ and $\alpha_1(\boldsymbol{x}) = \boldsymbol{\pi}$). This implies that $\tilde{\phi}_1(\boldsymbol{x}) = \tilde{\phi}_2(\boldsymbol{x}) = 0$ since $\tilde{\phi}_i \in \left[0, \frac{\pi}{2}\right]$ for $i \in \{1, 2\}$. Using the definition of $\tilde{\phi}_i$ for $\in \{1, 2\}$ in (2.9), we obtain that $|x_1 + r| = \frac{L}{\sigma_1}$ and $|x_1 - r| = \frac{L}{\sigma_2}$. Since $x_1 \in (-r, r)$, we have $x_1 = -r + \frac{L}{\sigma_1}$ and $x_1 = r - \frac{L}{\sigma_2}$. This implies that $r = \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$ which contradict our assumption.

- Consider $x_1 = -r$ or $x_1 = r$. We have that $\boldsymbol{x} = \boldsymbol{x}_1^*$ or $\boldsymbol{x} = \boldsymbol{x}_2^*$. However, $\boldsymbol{x}_1^* \notin \mathcal{T}$ and $\boldsymbol{x}_2^* \notin \mathcal{T}$ by the definition of $\mathcal{T}$.

Since $\boldsymbol{x} \neq (k, \mathbf{0})$ for all $k \in \mathbb{R}$, we conclude that $\boldsymbol{e}_2'(\boldsymbol{x})$ is well-defined.

Next, we will verify that for $\boldsymbol{x} = (x_1, \tilde{\mathbf{x}}) \in \mathcal{T}$ and $i \in \{1, 2\}$, the statements $-2d_i(\boldsymbol{x}) \cdot \sin\frac{1}{2}(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})) < 0$ and $-\|\tilde{\mathbf{x}}\| \cdot \csc(\theta(\boldsymbol{x})) < 0$, which are used to specify a range of ϵ's value in our claim, hold. Since $\boldsymbol{x} \notin \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we have $d_i(\boldsymbol{x}) > 0$ for $i \in \{1, 2\}$. In addition, from the above analysis, since $\boldsymbol{x} \notin \text{span}\{\boldsymbol{e}_1\}$, we have $\tilde{\phi}_i(\boldsymbol{x}) \in \left(0, \frac{\pi}{2}\right)$ for $i \in \{1, 2\}$. Therefore, it holds that $\frac{1}{2}(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})) \in \left(0, \frac{\pi}{2}\right)$ and then combining the two facts, we verify the first statement. Since $\boldsymbol{x} \notin \text{span}\{\boldsymbol{e}_1\}$ from the analysis above, we have $\|\tilde{\mathbf{x}}\| > 0$, and $\alpha_i(\boldsymbol{x}) \in (0, \boldsymbol{\pi})$ for $i \in \{1, 2\}$ (from the definition of α_i in (2.10)). Therefore, we have $\csc(\theta(\boldsymbol{x})) = \csc\left(\frac{1}{2}(\alpha_1(\boldsymbol{x}) + \alpha_2(\boldsymbol{x}))\right) \in [1, \infty)$. Combining the two facts, we verify the second statement.

Then, given a point $\boldsymbol{x} \in \mathcal{T}$, we consider the new set of bases $\mathcal{J}(\boldsymbol{x}) = \{\boldsymbol{e}_1'(\boldsymbol{x}), \boldsymbol{e}_2'(\boldsymbol{x}), \ldots, \boldsymbol{e}_n'(\boldsymbol{x})\}$ where $\boldsymbol{e}_1'(\boldsymbol{x}) = \boldsymbol{e}_1$, $\boldsymbol{e}_2'(\boldsymbol{x}) = \frac{\boldsymbol{v}(\boldsymbol{x})}{\|\boldsymbol{v}(\boldsymbol{x})\|}$ (as discussed at the beginning of the proof), and $\boldsymbol{e}_3'(\boldsymbol{x}), \boldsymbol{e}_4'(\boldsymbol{x}), \ldots, \boldsymbol{e}_n'(\boldsymbol{x})$ can be obtained by applying Gram-Schmidt procedure to the set of standard bases. Note that $\langle \boldsymbol{e}_1, \boldsymbol{e}_2'(\boldsymbol{x})\rangle = 0$ by our construction. We denote $(\cdot, \cdot, \ldots, \cdot)_{\mathcal{J}(\boldsymbol{x})}$

as the coordinate (or vector's component) with respect to the set of new bases. Let $\hat{x}_2 = \|\tilde{\mathbf{x}}\| \in \mathbb{R}_{>0}$. Then, by our construction, we have $\boldsymbol{x} = (x_1, \hat{x}_2, \mathbf{0})_{\mathcal{J}(\boldsymbol{x})}$. Let the point

$$\boldsymbol{y} = \Big(x_1 - \hat{x}_2 \cot\theta(\boldsymbol{x}),\ \mathbf{0}\Big)_{\mathcal{J}(\boldsymbol{x})}.$$

Recall that $\boldsymbol{x}'_\epsilon$ is defined as given in (A.7). We can write $\boldsymbol{y} = \boldsymbol{x}'_\epsilon$ with $\epsilon = -\hat{x}_2 \csc\theta(\boldsymbol{x})$. We will show that $y_1 = x_1 - \hat{x}_2 \cot\theta(\boldsymbol{x}) \in (-r, r)$. Since $\alpha_2(\boldsymbol{x}) > \alpha_1(\boldsymbol{x})$ (by the fact that $\boldsymbol{x} \notin \operatorname{span}\{\boldsymbol{e}_1\}$), we have

$$\cot\alpha_1(\boldsymbol{x}) > \cot\theta(\boldsymbol{x}), \quad \Leftrightarrow \quad \frac{x_1 + r}{\hat{x}_2} > \cot\theta(\boldsymbol{x}), \quad \Leftrightarrow \quad x_1 - \hat{x}_2 \cot\theta(\boldsymbol{x}) > -r.$$

The last inequality is due to $\hat{x}_2 > 0$. Similarly, since $\alpha_2(\boldsymbol{x}) > \alpha_1(\boldsymbol{x})$ (by the fact that $\boldsymbol{x} \notin \operatorname{span}\{\boldsymbol{e}_1\}$), we have

$$\cot\alpha_2(\boldsymbol{x}) < \cot\theta(\boldsymbol{x}), \quad \Leftrightarrow \quad \frac{x_1 - r}{\hat{x}_2} < \cot\theta(\boldsymbol{x}), \quad \Leftrightarrow \quad x_1 - \hat{x}_2 \cot\theta(\boldsymbol{x}) < r.$$

For convenience, let

$$a(\boldsymbol{x}) = \max\Big\{ \max_{i\in\{1,2\}} -2d_i(\boldsymbol{x}) \sin\frac{1}{2}\Big(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})\Big),\ -\|\tilde{\mathbf{x}}\| \csc\theta(\boldsymbol{x})\Big\}. \tag{A.8}$$

Next, we will show that for all $\epsilon \in (a(\boldsymbol{x}),\ 0)$, we have $\alpha_1(\boldsymbol{x}'_\epsilon) < \alpha_1(\boldsymbol{x})$ and $\alpha_2(\boldsymbol{x}'_\epsilon) > \alpha_2(\boldsymbol{x})$, and for all $\epsilon \in (0, \infty)$, we have $\alpha_1(\boldsymbol{x}'_\epsilon) > \alpha_1(\boldsymbol{x})$ and $\alpha_2(\boldsymbol{x}'_\epsilon) < \alpha_2(\boldsymbol{x})$ which will lead to proving parts (i) and (ii), respectively, of our claim. Consider the triangle formed by the points $\boldsymbol{x}_1^*$, $\boldsymbol{x}$, and $\boldsymbol{y}$. Suppose $\epsilon \in (a(\boldsymbol{x}),\ 0)$. Then, the point $\boldsymbol{x}'_\epsilon$ is on the line segment connecting $\boldsymbol{y}$ and $\boldsymbol{x}$. Since for a given $\boldsymbol{z} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_1^*\}$, $\alpha_1(\boldsymbol{z}) = \angle(\boldsymbol{z} - \boldsymbol{x}_1^*,\ \boldsymbol{e}_1)$ (from the definition of α_1 in (2.10)) and $y_1 > -r$, we have $\alpha_1(\boldsymbol{x}'_\epsilon) < \alpha_1(\boldsymbol{x})$. Next, suppose $\epsilon \in (0, \infty)$. Then, the point $\boldsymbol{x}'_\epsilon$ is on the ray from $\boldsymbol{y}$ to $\boldsymbol{x}$ but not on the line segment connecting $\boldsymbol{y}$ and $\boldsymbol{x}$ which implies that $\alpha_1(\boldsymbol{x}'_\epsilon) > \alpha_1(\boldsymbol{x})$. Now, consider the triangle formed by the points $\boldsymbol{x}_2^*$, $\boldsymbol{x}$, and $\boldsymbol{y}$. By using similar argument as before (with the fact that $\alpha_2(\boldsymbol{z}) = \angle(\boldsymbol{z} - \boldsymbol{x}_2^*,\ \boldsymbol{e}_1)$ for a given $\boldsymbol{z} \in \mathbb{R}^n \setminus \{\boldsymbol{x}_2^*\}$, and $y_1 < r$), we can verify that $\alpha_2(\boldsymbol{x}'_\epsilon) < \alpha_2(\boldsymbol{x})$.

We will show that for all $\epsilon \in (a(\boldsymbol{x}),\ 0)$ and $i \in \{1,2\}$, we have $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| < \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$, and for all $\epsilon \in (0,\infty)$ for $i \in \{1,2\}$, we have $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| > \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$. Suppose $\epsilon \in (a(\boldsymbol{x}),\ 0)$. Then, we have $\epsilon > -2d_1(\boldsymbol{x}) \sin \frac{1}{2}\big(\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x})\big)$. Using $\tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \boldsymbol{\pi} - (\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x}))$ and $\sin\left(\frac{\pi}{2} - z\right) = \cos z$ for all $z \in \mathbb{R}$, we can write

$$\epsilon > -2d_1(\boldsymbol{x}) \cos \frac{1}{2}\big(\alpha_2(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\big), \quad \Leftrightarrow \quad \epsilon + 2d_1(\boldsymbol{x}) \cos\Big(\theta(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\Big) > 0.$$

Expanding the cosine and using $d_1(\boldsymbol{x}) \cos \alpha_1(\boldsymbol{x}) = x_1 + r$ and $d_1(\boldsymbol{x}) \sin \alpha_1(\boldsymbol{x}) = \hat{x}_2$ from (2.23) and (2.24), respectively, we obtain $\epsilon + 2(x_1 + r) \cos \theta(\boldsymbol{x}) + 2\hat{x}_2 \sin \theta(\boldsymbol{x}) > 0$. Multiplying the inequality by ϵ (which is negative) and then adding $(x_1 + r)^2 + \hat{x}_2^2$ to both sides, we get

$$(x_1 + \epsilon \cos \theta(\boldsymbol{x}) + r)^2 + (\hat{x}_2 + \epsilon \sin \theta(\boldsymbol{x}))^2 < (x_1 + r)^2 + \hat{x}_2^2,$$

which is $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_1\| < \|\boldsymbol{x} - \boldsymbol{x}^*_1\|$. Next, suppose $\epsilon \in (0,\infty)$. Recall the definition of $d_i(\boldsymbol{x})$ for $i \in \{1,2\}$ in (2.21). Since $0 < \alpha_1(\boldsymbol{x}) \leq \alpha_2(\boldsymbol{x}) < \boldsymbol{\pi}$ (by the fact that $\boldsymbol{x} \notin \text{span}\{\boldsymbol{e}_1\}$), we have

$$\cos\Big(\theta(\boldsymbol{x}) - \alpha_1(\boldsymbol{x})\Big) > 0, \quad \Leftrightarrow \quad d_1(\boldsymbol{x}) \cos \alpha_1(\boldsymbol{x}) \cos \theta(\boldsymbol{x}) + d_1(\boldsymbol{x}) \sin \alpha_1(\boldsymbol{x}) \sin \theta(\boldsymbol{x}) > 0.$$

Since $d_1(\boldsymbol{x}) \cos \alpha(\boldsymbol{x}) = x_1 + r$ and $d_1(\boldsymbol{x}) \sin \alpha(\boldsymbol{x}) = \hat{x}_2$ from (2.23) and (2.24), respectively, we obtain $(x_1 + r) \cos \theta(\boldsymbol{x}) + \hat{x}_2 \sin \theta(\boldsymbol{x}) > 0$. Using the assumption that $\epsilon > 0$, we can write

$$\frac{\epsilon}{2} + (x_1 + r) \cos \theta(\boldsymbol{x}) + \hat{x}_2 \sin \theta(\boldsymbol{x}) > 0.$$

Multiplying the above inequality by 2ϵ (which is positive), then adding $(x_1 + r)^2 + \hat{x}_2^2$ to both sides, and rearranging, we get $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_1\| > \|\boldsymbol{x} - \boldsymbol{x}^*_1\|$. By using similar steps as above, we can also show that $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_2\| < \|\boldsymbol{x} - \boldsymbol{x}^*_2\|$ for $\epsilon \in (a(\boldsymbol{x}),\ 0)$ and $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_2\| > \|\boldsymbol{x} - \boldsymbol{x}^*_2\|$ for $\epsilon \in (0,\infty)$.

Here, we will prove part (i) of our claim. Suppose $\epsilon \in (a(\boldsymbol{x}),\ 0)$. From the above analysis, we have $\alpha_1(\boldsymbol{x}'_\epsilon) < \alpha_1(\boldsymbol{x})$, $\alpha_2(\boldsymbol{x}'_\epsilon) > \alpha_2(\boldsymbol{x})$ and $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| < \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$ for $i \in \{1,2\}$. Since $\alpha_1(\boldsymbol{x}'_\epsilon) < \alpha_1(\boldsymbol{x})$ and $\alpha_2(\boldsymbol{x}'_\epsilon) > \alpha_2(\boldsymbol{x})$, we obtain that $\psi(\boldsymbol{x}'_\epsilon) < \psi(\boldsymbol{x})$ by the definition of ψ in

(2.11). For $i \in \{1,2\}$, since $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| < \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$, we obtain that $\tilde{\phi}_i(\boldsymbol{x}'_\epsilon) > \tilde{\phi}_i(\boldsymbol{x})$ by the definition of $\tilde{\phi}_i$ in (2.9). Therefore, it holds that

$$\tilde{\phi}_1(\boldsymbol{x}'_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}'_\epsilon) > \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x}) > \psi(\boldsymbol{x}'_\epsilon).$$

Here, we will prove part (ii) of our claim. Suppose $\epsilon \in (0, \infty)$. From the above analysis, we have $\alpha_1(\boldsymbol{x}'_\epsilon) > \alpha_1(\boldsymbol{x})$, $\alpha_2(\boldsymbol{x}'_\epsilon) < \alpha_2(\boldsymbol{x})$ and $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| > \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$ for $i \in \{1,2\}$. By using similar argument as the proof of part (i), we get $\psi(\boldsymbol{x}'_\epsilon) > \psi(\boldsymbol{x})$. However, for $i \in \{1,2\}$, $\|\boldsymbol{x}'_\epsilon - \boldsymbol{x}^*_i\| > \|\boldsymbol{x} - \boldsymbol{x}^*_i\|$ implies that either $\tilde{\phi}_i(\boldsymbol{x}'_\epsilon) < \tilde{\phi}_i(\boldsymbol{x})$, or $\tilde{\phi}_i(\boldsymbol{x}'_\epsilon)$ is not well-defined. In the case that $\tilde{\phi}_i(\boldsymbol{x}'_\epsilon) < \tilde{\phi}_i(\boldsymbol{x})$ for $i \in \{1,2\}$, it holds that

$$\tilde{\phi}_1(\boldsymbol{x}'_\epsilon) + \tilde{\phi}_2(\boldsymbol{x}'_\epsilon) < \tilde{\phi}_1(\boldsymbol{x}) + \tilde{\phi}_2(\boldsymbol{x}) = \psi(\boldsymbol{x}) < \psi(\boldsymbol{x}'_\epsilon).$$

Finally, suppose $\boldsymbol{x} \in \mathcal{T}$. Recall the quantity $a(\boldsymbol{x})$ from (A.8). For $\delta \in \mathbb{R}_{>0}$, let $\boldsymbol{x}_{\text{in}} = \boldsymbol{x}'_\epsilon$ with $\epsilon = \max\left\{a(\boldsymbol{x}), -\frac{\delta}{2}\right\}$ and $\boldsymbol{x}_{\text{out}} = \boldsymbol{x}'_\epsilon$ with $\epsilon = \frac{\delta}{2}$. From our claim, we have that for all $\delta \in \mathbb{R}_{>0}$, it holds that $\boldsymbol{x}_{\text{in}}, \boldsymbol{x}_{\text{out}} \in \mathcal{B}(\boldsymbol{x}, \delta)$, $\boldsymbol{x}_{\text{in}} \in \mathcal{M}^{\uparrow}$, $\boldsymbol{x}_{\text{in}} \in \mathcal{M}_{\downarrow}$, $\boldsymbol{x}_{\text{out}} \in (\mathcal{M}^{\uparrow})^c$, and $\boldsymbol{x}_{\text{out}} \in (\mathcal{M}_{\downarrow})^c$. Thus, we conclude that $\boldsymbol{x} \in \partial\mathcal{M}^{\uparrow}$ and $\boldsymbol{x} \in \partial\mathcal{M}_{\downarrow}$. □

A.1.7 Proof of Lemma 2.4.7

Proof of Lemma 2.4.7. Consider part (i) of the lemma. From Lemma 2.4.5 part (ii), we have $\lambda_1 < \frac{L}{\sigma_1} - r$. So, the interval $\left[\lambda_1,\ -r + \frac{L}{\sigma_1}\right]$ is well-defined. From the definition of β and γ_1 in (2.26), we have $(1+\beta)\gamma_1 \geq 0$ and $-4\beta r^2 < 0$. Combining the two inequalities, we get $(1+\beta)\gamma_1 > -4\beta r^2$. By subtracting $2r^2$ from both sides and then rearranging the inequality, we obtain

$$\lambda_1 = \left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} > -r.$$

On the other hand, since $r > \frac{L}{2\sigma_1}$, we get $-r + \frac{L}{\sigma_1} < r$. Combining the two inequalities yields $\left[\lambda_1,\ -r + \frac{L}{\sigma_1}\right] \subseteq (-r, r)$.

For $\boldsymbol{x} \in \partial\mathcal{B}_1 \cap \mathcal{H}_1^+$, we have $x_1 \in \left[\lambda_1,\ -r + \frac{L}{\sigma_1}\right]$ (from (2.25)) which from above, implies that $\boldsymbol{x} \in \overline{\mathcal{B}}_1 \setminus \{\boldsymbol{x}^*_1, \boldsymbol{x}^*_2\}$. It remains to show that $\boldsymbol{x} \in \overline{\mathcal{B}}_2$. By examining the equation describing the

set $\partial\mathcal{B}_1$ in (2.25) and the inequality describing the set $\overline{\mathcal{B}}_2 = \{\boldsymbol{z} \in \mathbb{R}^n : (z_1 + r)^2 + \|\tilde{\mathbf{z}}\|^2 \leq \gamma_2\}$ where γ_2 is defined in (2.26), one can verify that

$$\partial\mathcal{B}_1 \cap \overline{\mathcal{B}}_2 = \left\{\boldsymbol{z} \in \partial\mathcal{B}_1 : z_1 \in \left[\frac{1}{4r}(\gamma_1 - \gamma_2),\ -r + \frac{L}{\sigma_1}\right]\right\}. \tag{A.9}$$

On the other hand, by performing some algebraic manipulation and using the definition of β and γ_i for $i \in \{1, 2\}$ in (2.26), one can verify that

$$r \in \left(0,\ \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)\right), \quad \Leftrightarrow \quad \lambda_1 = \left(\frac{1+\beta}{1+2\beta}\right)\frac{\gamma_1}{2r} - \frac{r}{1+2\beta} > \frac{1}{4r}(\gamma_1 - \gamma_2).$$

From the definition of $\mathcal{H}_1^+$ in (2.29) and equation (A.9), this means that $\partial\mathcal{B}_1 \cap \mathcal{H}_1^+ \subset \partial\mathcal{B}_1 \cap \overline{\mathcal{B}}_2 \subseteq \overline{\mathcal{B}}_2$. For part (ii) of the lemma, we can proceed in a similar manner as the above analysis. □

A.2 Proofs of Theoretical Results for Inner Approximation

A.2.1 Proof of Proposition 2.5.1

Before proving Proposition 2.5.1, we consider an equivalent condition for the existence of a quadratic function with two independent variables satisfying certain properties.

Lemma A.2.1. *Let $\mathcal{Q}$ be defined as in (2.4). Suppose we are given $\boldsymbol{x}^* \in \mathbb{R}^2$, $\boldsymbol{x}_0 \in \mathbb{R}^2$ such that $\boldsymbol{x}_0 \neq \boldsymbol{x}^*$, $\boldsymbol{g} \in \mathbb{R}^2$, and $\sigma \in \mathbb{R}_{>0}$. Then, there exists a function $f \in \mathcal{Q}^{(2)}(\boldsymbol{x}^*, \sigma)$ with a gradient $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$ if and only if*

(i) $\boldsymbol{x}_0 \in \overline{\mathcal{B}}\left(\boldsymbol{x}^*, \frac{\|\boldsymbol{g}\|}{\sigma}\right)$ *and*

(ii) $\angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) \in \{0\} \cup \left[0,\ \arccos\left(\frac{\sigma}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\right)\right)$.

Note that if $\sigma\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| = \|\boldsymbol{g}\|$, then $\left[0,\ \arccos(\frac{\sigma}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|)\right) = \emptyset$.

Proof. For the forward direction, suppose that $f \in \mathcal{Q}^{(2)}(\boldsymbol{x}^*, \sigma)$ and $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$. Since $\mathcal{Q}(\boldsymbol{x}^*, \sigma) \subset \mathcal{S}(\boldsymbol{x}^*, \sigma)$, from (2.3), we have

$$\|\boldsymbol{g}\|\ \|\boldsymbol{x}_0 - \boldsymbol{x}^*\| \geq \langle \nabla f(\boldsymbol{x}_0),\ \boldsymbol{x}_0 - \boldsymbol{x}^* \rangle \geq \sigma\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|^2. \tag{A.10}$$

We then have $\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| \leq \frac{\|\boldsymbol{g}\|}{\sigma}$ (i.e., $\boldsymbol{x}_0 \in \overline{\mathcal{B}}\Big(\boldsymbol{x}^*, \frac{\|\boldsymbol{g}\|}{\sigma}\Big)$) which corresponds to part (i). On the other hand, we can rewrite the second inequality of (A.10) as

$$\|\boldsymbol{g}\| \cos \angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) \geq \sigma \|\boldsymbol{x}_0 - \boldsymbol{x}^*\| > 0,$$

which implies that $\|\boldsymbol{g}\| \in \mathbb{R}_{>0}$ and $\angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) \in \Big[0, \frac{\pi}{2}\Big)$.

Since $f \in \mathcal{Q}^{(2)}(\boldsymbol{x}^*, \sigma)$, we can write $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{P}\boldsymbol{x} + \boldsymbol{b}^\intercal \boldsymbol{x} + c$ where $\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} \in \mathsf{S}^2$, $\boldsymbol{b} \in \mathbb{R}^2$, and $c \in \mathbb{R}$. The gradient of f is $\nabla f(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{b}$. Since $\boldsymbol{x}^*$ is the minimizer of the quadratic function, by substituting $\boldsymbol{x}^*$ into the gradient equation and using $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, we get $\boldsymbol{b} = -\boldsymbol{P}\boldsymbol{x}^*$, and we can rewrite the gradient as $\boldsymbol{g} = \nabla f(\boldsymbol{x}_0) = \boldsymbol{P}(\boldsymbol{x}_0 - \boldsymbol{x}^*)$. Let $\boldsymbol{v} = \boldsymbol{x}_0 - \boldsymbol{x}^*$ and then rewrite the gradient equation into two equations as follows:

$$\begin{cases} g_1 = p_{11}v_1 + p_{12}v_2, \\ g_2 = p_{12}v_1 + p_{22}v_2, \end{cases} \quad \Leftrightarrow \quad \begin{cases} p_{12}v_2 = -p_{11}v_1 + g_1, \\ p_{22}v_2^2 = v_1(p_{11}v_1 - g_1) + g_2 v_2. \end{cases} \tag{A.11}$$

Since σ is an eigenvalue of matrix $\boldsymbol{P}$, by expanding the equation $\det(\sigma \boldsymbol{I} - \boldsymbol{P}) = 0$ and rearranging the resulting equation, we get

$$p_{12}^2 = \sigma^2 - (p_{11} + p_{22})\sigma + p_{11}p_{22}. \tag{A.12}$$

Multiply (A.12) by v_2^2, then substitute $p_{12}v_2$ and $p_{22}v_2^2$ from (A.11) into the resulting equation, and rearrange it to get

$$(\sigma v_1^2 - g_1 v_1 + \sigma v_2^2 - g_2 v_2)p_{11} = \sigma g_1 v_1 + \sigma^2 v_2^2 - \sigma g_2 v_2 - g_1^2. \tag{A.13}$$

Let the function $\boldsymbol{R} : (-\boldsymbol{\pi}, \boldsymbol{\pi}] \to \mathbb{R}^{2\times 2}$ be such that $\boldsymbol{R}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ (i.e., a rotation matrix), $d = \|\boldsymbol{v}\| = \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|$, and $\phi = \measuredangle(\boldsymbol{g},\ \boldsymbol{v})$ where $\measuredangle(\cdot,\cdot)$ is defined in (2.1). Since

$\angle(\boldsymbol{g}, \boldsymbol{v}) \in \left[0, \frac{\pi}{2}\right)$ from the above analysis, we have that $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and we can decompose gradient $\boldsymbol{g}$ as

$$\boldsymbol{g} = \|\boldsymbol{g}\| \, \boldsymbol{R}(\phi) \left(\frac{\boldsymbol{x}_0 - \boldsymbol{x}^*}{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|}\right) = \frac{\|\boldsymbol{g}\|}{d} \, \boldsymbol{R}(\phi) \, \boldsymbol{v}. \tag{A.14}$$

For simplicity of notations, we let $\hat{L} = \frac{\|\boldsymbol{g}\|}{d}$ which can be viewed as the norm of the gradient at $\boldsymbol{x}_0$ (i.e., $\|\nabla f(\boldsymbol{x}_0)\|$) normalized by distance from the minimizer (i.e., $\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|$). Note that since $\|\boldsymbol{g}\| > 0$ and $d > 0$, we have $\hat{L} > 0$. Then, rewrite the expression (A.14) into two equations as follows:

$$\begin{cases} g_1 = \hat{L}(v_1 \cos\phi - v_2 \sin\phi), \\ g_2 = \hat{L}(v_1 \sin\phi + v_2 \cos\phi). \end{cases} \tag{A.15}$$

We then substitute (A.15) into (A.13) to get

$$d^2(\hat{L}\cos\phi - \sigma)p_{11} = \hat{L}(\hat{L}\cos\phi - \sigma)(v_1^2 \cos\phi - 2v_1 v_2 \sin\phi - v_2^2 \cos\phi) + (\hat{L}^2 - \sigma^2)v_2^2. \tag{A.16}$$

Multiply $p_{22}v_2^2$'s equation in (A.11) by $d^2(\hat{L}\cos\phi - \sigma)$, then substitute (A.15) and (A.16) into the resulting equation, and rearrange it yields

$$d^2(\hat{L}\cos\phi - \sigma)p_{22} = (\hat{L}\cos\phi - \sigma)(\sigma v_1^2 + 2\hat{L}v_1 v_2 \sin\phi + \hat{L}v_2^2 \cos\phi) + (\hat{L}v_1 \sin\phi)^2. \tag{A.17}$$

We then add (A.16) to (A.17) and simplify the expression to obtain

$$(\hat{L}\cos\phi - \sigma)(p_{11} + p_{22}) = \hat{L}^2 - \sigma^2. \tag{A.18}$$

Let $\mu \in \mathbb{R}$ be the other eigenvalue of matrix $\boldsymbol{P}$. We want to compute μ in terms of σ, $\hat{L}$, and ϕ. Since $\mathrm{Tr}(\boldsymbol{P}) = p_{11} + p_{22} = \sigma + \mu$, using equation (A.18), we have

$$(\hat{L}\cos\phi - \sigma)\mu = \hat{L}(\hat{L} - \sigma\cos\phi). \tag{A.19}$$

Consider the following cases on the term $\hat{L}\cos\phi - \sigma$.

- Suppose $\sigma = \hat{L}\cos\phi$. Then, (A.19) becomes $0 = \hat{L}^2(1 - \cos^2\phi)$ which implies that $\phi = 0$.

- Suppose $\sigma \neq \hat{L}\cos\phi$. Then, we can rewrite equation (A.19) as

$$\mu = \frac{\hat{L}(\hat{L} - \sigma\cos\phi)}{\hat{L}\cos\phi - \sigma}. \tag{A.20}$$

Since $\lambda_{\min}(\boldsymbol{P}) = \sigma$, we have that $\mu \geq \sigma$ where μ is given in (A.20). Specifically, we will show that

given μ as expressed in (A.20) with $\hat{L} > 0$ and $\sigma > 0$,

it holds that $\mu \geq \sigma$ if and only if $\hat{L}\cos\phi - \sigma > 0$. (A.21)

For the forward direction, suppose $\mu \geq \sigma$ and $\hat{L}\cos\phi - \sigma < 0$. We have

$$\mu = \frac{\hat{L}(\hat{L} - \sigma\cos\phi)}{\hat{L}\cos\phi - \sigma} \geq \sigma, \quad \Leftrightarrow \quad \cos\phi \geq \frac{1}{2}\Big(\frac{\hat{L}}{\sigma} + \frac{\sigma}{\hat{L}}\Big).$$

However, since $\hat{L} > 0$ and $\sigma > 0$, we get $\frac{1}{2}\Big(\frac{\hat{L}}{\sigma} + \frac{\sigma}{\hat{L}}\Big) \geq 1$. Therefore, we obtain that $\cos\phi = \frac{1}{2}\Big(\frac{\hat{L}}{\sigma} + \frac{\sigma}{\hat{L}}\Big) = 1$ which implies that $\phi = 0$ and $\hat{L} = \sigma$. However, we get $\hat{L}\cos\phi - \sigma = 0$ which makes μ undefined. For the converse, suppose $\mu < \sigma$ and $\hat{L}\cos\phi - \sigma > 0$, we have

$$\mu = \frac{\hat{L}(\hat{L} - \sigma\cos\phi)}{\hat{L}\cos\phi - \sigma} < \sigma, \quad \Leftrightarrow \quad \cos\phi > \frac{1}{2}\Big(\frac{\hat{L}}{\sigma} + \frac{\sigma}{\hat{L}}\Big).$$

However, this is not possible since $\frac{1}{2}\Big(\frac{\hat{L}}{\sigma} + \frac{\sigma}{\hat{L}}\Big) \geq 1$ for $\frac{\sigma}{\hat{L}} > 0$ and we have proved the claim.

Since $\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| \in \Big(0, \frac{\|\boldsymbol{g}\|}{\sigma}\Big]$ (or equivalently $\frac{\sigma d}{\|\boldsymbol{g}\|} \in (0, 1]$), the expression $\arccos\Big(\frac{\sigma}{\hat{L}}\Big)$ is well-defined. Combining the two cases ($\sigma = \hat{L}\cos\phi$ and $\sigma \neq \hat{L}\cos\phi$), we have shown that if there

exists $f \in \mathcal{Q}(\boldsymbol{x}^*, \sigma)$ with the gradient $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$ then from the definition of $\measuredangle(\cdot,\cdot)$ and $\angle(\cdot,\cdot)$ in (2.1), we have

$$\begin{aligned} & \measuredangle(\boldsymbol{g}, \boldsymbol{v}) \in \{0\} \cup \Big(-\arccos\Big(\frac{\sigma}{\hat{L}}\Big),\ \arccos\Big(\frac{\sigma}{\hat{L}}\Big)\Big), \\ \Leftrightarrow \quad & \angle(\boldsymbol{g}, \boldsymbol{v}) \in \{0\} \cup \Big[0,\ \arccos\Big(\frac{\sigma}{\hat{L}}\Big)\Big), \end{aligned}$$

which corresponds to part (ii).

For the converse, let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{P}\boldsymbol{x} - (\boldsymbol{x}^*)^\intercal \boldsymbol{P}\boldsymbol{x}$ and each entry of $\boldsymbol{P}$ will be specified below. First, suppose that $\sigma < \hat{L}$ (i.e., $\boldsymbol{x}_0 \in \mathcal{B}\big(\boldsymbol{x}^*, \frac{\|\boldsymbol{g}\|}{\sigma}\big)$) and $\phi = \angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) \in \Big[0,\ \arccos\Big(\frac{\sigma}{\hat{L}}\Big)\Big)$. This implies that $\hat{L}\cos\phi - \sigma > 0$. If $v_2 \neq 0$, let

$$\begin{aligned} p_{12} = -\frac{v_1}{d^2 v_2}(\hat{L}v_1^2\cos\phi - 2\hat{L}v_1v_2\sin\phi - \hat{L}v_2^2\cos\phi) \\ - \frac{(\hat{L}^2 - \sigma^2)v_1v_2}{d^2(\hat{L}\cos\phi - \sigma)} + \hat{L}\Big(\frac{v_1}{v_2}\cos\phi - \sin\phi\Big). \end{aligned} \tag{A.22}$$

We choose p_{11}, p_{12}, and p_{22} as given in (A.16), (A.22), and (A.17), respectively. Note that p_{11}, p_{12}, and p_{22} are well-defined since $\hat{L}\cos\phi - \sigma \neq 0$. If $v_2 = 0$, we choose

$$p_{11} = \hat{L}\cos\phi, \quad p_{12} = \hat{L}\sin\phi, \quad \text{and} \quad p_{22} = \sigma + \frac{(\hat{L}\sin\phi)^2}{\hat{L}\cos\phi - \sigma}.$$

In both cases ($v_2 \neq 0$ and $v_2 = 0$), one can easily verify that $\nabla f(\boldsymbol{x}_0) = \boldsymbol{P}\boldsymbol{v} = \boldsymbol{g}$ and $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$. In order to show that $\lambda_{\min}(\boldsymbol{P}) = \sigma$, first, we check that p_{11}, p_{12}, and p_{22} satisfy (A.12) which means that σ is an eigenvalue of $\boldsymbol{P}$. Next, using the fact that the other eigenvalue $\mu = \mathrm{Tr}(\boldsymbol{P}) - \sigma = (p_{11} + p_{22}) - \sigma$, we obtain μ as expressed in (A.20) for both cases. Since $\hat{L}\cos\phi - \sigma > 0$, from the statement (A.21), we have $\mu \geq \sigma$.

Now suppose that $\sigma = \hat{L}$ (i.e., $\boldsymbol{x}_0 \in \partial\mathcal{B}\big(\boldsymbol{x}^*, \frac{\|\boldsymbol{g}\|}{\sigma}\big)$) and $\phi = \angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) = 0$. Let the orthogonal matrix $\boldsymbol{T} = \begin{bmatrix} \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} & \frac{\boldsymbol{v}_\perp}{\|\boldsymbol{v}_\perp\|} \end{bmatrix} \in \mathbb{R}^{2\times 2}$ where $\boldsymbol{v}_\perp \neq \boldsymbol{0}$ is a vector perpendicular to vector $\boldsymbol{v}$ and $\boldsymbol{P} = \sigma\boldsymbol{T}\boldsymbol{T}^\intercal = \sigma\boldsymbol{I}$. We have $\nabla f(\boldsymbol{x}_0) = \boldsymbol{P}\boldsymbol{v} = \sigma\boldsymbol{v} = \|\boldsymbol{g}\| \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$. However, since $\angle(\boldsymbol{g},\ \boldsymbol{v}) = 0$, we have $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$. We also have $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ and $\lambda_{\min}(\boldsymbol{P}) = \sigma$ by our construction of f. Thus, for both cases, we obtain that $f \in \mathcal{Q}^{(2)}(\boldsymbol{x}^*, \sigma)$ and $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$. $\square$

Proof of Proposition 2.5.1. For given points $\boldsymbol{x}^* \in \mathbb{R}^n$, $\boldsymbol{x}_0 \in \mathbb{R}^n$ such that $\boldsymbol{x}_0 \neq \boldsymbol{x}^*$ and vector $\boldsymbol{g} \in \mathbb{R}^n$, let

$$\boldsymbol{e}_1' = \frac{\boldsymbol{x}_0 - \boldsymbol{x}^*}{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|}, \quad \boldsymbol{e}_2' = \begin{cases} \frac{\boldsymbol{g} - \langle \boldsymbol{g}, \boldsymbol{e}_1' \rangle \boldsymbol{e}_1'}{\|\boldsymbol{g} - \langle \boldsymbol{g}, \boldsymbol{e}_1' \rangle \boldsymbol{e}_1'\|} & \text{if} \quad \boldsymbol{g} \notin \operatorname{span}(\{\boldsymbol{e}_1'\}), \\ \boldsymbol{u} \text{ where } \boldsymbol{u} \in \mathcal{N}((\boldsymbol{e}_1')^\intercal) & \text{otherwise,} \end{cases} \tag{A.23}$$

and $\boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1' & \boldsymbol{e}_2' \end{bmatrix} \in \mathbb{R}^{n \times 2}$. We let $\boldsymbol{x}_{\boldsymbol{g}} = \boldsymbol{x}^* + \boldsymbol{g}$ to be the point generated by vector $\boldsymbol{g}$ and

$$\mathcal{P}_2 = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{s} \quad \text{for some } \boldsymbol{s} \in \mathbb{R}^2\} = \boldsymbol{x}^* + \mathcal{R}(\boldsymbol{E})$$

to be a 2-D plane in $\mathbb{R}^n$. One can verify that $\boldsymbol{e}_1'$ and $\boldsymbol{e}_2'$ are orthonormal, and $\{\boldsymbol{x}^*, \boldsymbol{x}_0, \boldsymbol{x}_{\boldsymbol{g}}\} \subset \mathcal{P}_2$. We let $\boldsymbol{T} : \mathbb{R}^2 \to \mathcal{P}_2$ to be such that $\boldsymbol{T}(\boldsymbol{s}) = \boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{s}$. Since function $\boldsymbol{T}$ is bijective, we then have $\boldsymbol{T}^{-1} : \mathcal{P}_2 \to \mathbb{R}^2$ such that $\boldsymbol{T}^{-1}(\boldsymbol{x}) = \boldsymbol{E}^\intercal(\boldsymbol{x} - \boldsymbol{x}^*)$. Next, consider a property of the norm and angle function which we will use in the subsequent analysis. Using $\boldsymbol{E}^\intercal\boldsymbol{E} = \boldsymbol{I}$ and that $\boldsymbol{T}$ is bijective, we have that for all $i \in \{1, 2, 3, 4\}$, $\boldsymbol{s}_i \in \mathbb{R}^2$ and $\boldsymbol{x}_i \in \mathcal{P}_2$,

- *distance invariance:* $\|\boldsymbol{T}(\boldsymbol{s}_1) - \boldsymbol{T}(\boldsymbol{s}_2)\| = \|\boldsymbol{s}_1 - \boldsymbol{s}_2\|$ and $\|\boldsymbol{T}^{-1}(\boldsymbol{x}_1) - \boldsymbol{T}^{-1}(\boldsymbol{x}_2)\| = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$, and
- *angle invariance:* $\angle(\boldsymbol{T}(\boldsymbol{s}_1) - \boldsymbol{T}(\boldsymbol{s}_2),\ \boldsymbol{T}(\boldsymbol{s}_3) - \boldsymbol{T}(\boldsymbol{s}_4)) = \angle(\boldsymbol{s}_1 - \boldsymbol{s}_2,\ \boldsymbol{s}_3 - \boldsymbol{s}_4)$ and $\angle(\boldsymbol{T}^{-1}(\boldsymbol{x}_1) - \boldsymbol{T}^{-1}(\boldsymbol{x}_2),\ \boldsymbol{T}^{-1}(\boldsymbol{x}_3) - \boldsymbol{T}^{-1}(\boldsymbol{x}_4)) = \angle(\boldsymbol{x}_1 - \boldsymbol{x}_2,\ \boldsymbol{x}_3 - \boldsymbol{x}_4)$.

For the forward direction, suppose $f \in \mathcal{Q}^{(n)}(\boldsymbol{x}^*, \sigma)$ and $\nabla f(\boldsymbol{x}_0) = \boldsymbol{g}$. Let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\intercal\boldsymbol{P}\boldsymbol{x} + \boldsymbol{b}^\intercal\boldsymbol{x} + c$ for some $\boldsymbol{P} \in \mathfrak{S}^n$, $\boldsymbol{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$ that satisfies the conditions. Let $\tilde{f}(\boldsymbol{s}) = f(\boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{s})$. We then have

$$\tilde{f}(\boldsymbol{s}) = \frac{1}{2}\boldsymbol{s}^\intercal\boldsymbol{E}^\intercal\boldsymbol{P}\boldsymbol{E}\boldsymbol{s} + (\boldsymbol{P}\boldsymbol{x}^* + \boldsymbol{b})^\intercal\boldsymbol{E}\boldsymbol{s} + \left(\frac{1}{2}(\boldsymbol{x}^*)^\intercal\boldsymbol{P}\boldsymbol{x}^* + \boldsymbol{b}^\intercal\boldsymbol{x}^* + c\right), \tag{A.24}$$

and

$$\nabla\tilde{f}(\boldsymbol{T}^{-1}(\boldsymbol{x}_0)) = \boldsymbol{E}^\intercal\Big(\boldsymbol{P}\boldsymbol{E}\boldsymbol{T}^{-1}(\boldsymbol{x}_0) + \boldsymbol{P}\boldsymbol{x}^* + \boldsymbol{b}\Big) = \boldsymbol{E}^\intercal\Big(\boldsymbol{P}\boldsymbol{T}(\boldsymbol{T}^{-1}(\boldsymbol{x}_0)) + \boldsymbol{b}\Big) = \boldsymbol{E}^\intercal\boldsymbol{g},$$

where the second and third equalities are obtained by using the transformation $\boldsymbol{T}(\boldsymbol{s}) = \boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{s}$ at $\boldsymbol{s} = \boldsymbol{T}^{-1}(\boldsymbol{x}_0)$ and $\nabla f(\boldsymbol{x}_0) = \boldsymbol{P}\boldsymbol{x}_0 + \boldsymbol{b} = \boldsymbol{g}$, respectively. Similarly, we have $\nabla \tilde{f}(\boldsymbol{T}^{-1}(\boldsymbol{x}^*)) = \boldsymbol{0}$. Let $\tilde{\sigma}$ be the strong convexity parameter of $\tilde{f}$. We conclude that $\tilde{f} \in \mathcal{Q}^{(2)}(\boldsymbol{T}^{-1}(\boldsymbol{x}^*),\ \tilde{\sigma})$ and $\nabla \tilde{f}(\boldsymbol{T}^{-1}(\boldsymbol{x}_0)) = \boldsymbol{E}^\intercal \boldsymbol{g}$. However, from (A.24), we have that

$$\tilde{\sigma} = \lambda_{\min}(\boldsymbol{E}^\intercal \boldsymbol{P}\boldsymbol{E}) = \min_{\boldsymbol{s}\in\mathbb{R}^2} \frac{\boldsymbol{s}^\intercal(\boldsymbol{E}^\intercal \boldsymbol{P}\boldsymbol{E})\boldsymbol{s}}{\boldsymbol{s}^\intercal \boldsymbol{s}} = \min_{\boldsymbol{x}\in\mathcal{R}(\boldsymbol{E})} \frac{\boldsymbol{x}^\intercal \boldsymbol{P}\boldsymbol{x}}{\boldsymbol{x}^\intercal \boldsymbol{x}} \geq \min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{\boldsymbol{x}^\intercal \boldsymbol{P}\boldsymbol{x}}{\boldsymbol{x}^\intercal \boldsymbol{x}} = \sigma. \tag{A.25}$$

Using Lemma A.2.1, the distance invariance of $\boldsymbol{T}^{-1}$ and (A.25), we have

$$\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| = \|\boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\| \leq \frac{\|\boldsymbol{E}^\intercal \boldsymbol{g}\|}{\tilde{\sigma}} = \frac{\|\boldsymbol{T}^{-1}(\boldsymbol{x}_g) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\|}{\tilde{\sigma}} \leq \frac{\|\boldsymbol{g}\|}{\sigma}.$$

In addition, using Lemma A.2.1, the angle invariance of $\boldsymbol{T}^{-1}$ and (A.25), we have

$$\begin{aligned}
\angle(\boldsymbol{g},\ \boldsymbol{x}_0 - \boldsymbol{x}^*) &= \angle\Big(\boldsymbol{T}^{-1}(\boldsymbol{x}_g) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*),\ \boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\Big) \\
&= \angle\Big(\boldsymbol{E}^\intercal \boldsymbol{g},\ \boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\Big) \\
&\in \{0\} \cup \left[0,\ \arccos\left(\frac{\tilde{\sigma}}{\|\boldsymbol{E}^\intercal \boldsymbol{g}\|}\|\boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\|\right)\right) \\
&\subseteq \{0\} \cup \left[0,\ \arccos\left(\frac{\sigma}{\|\boldsymbol{g}\|}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\right)\right),
\end{aligned}$$

which complete the proof of the forward direction of the proposition.

For the converse, suppose $\|\boldsymbol{x}_0 - \boldsymbol{x}^*\| \leq \frac{\|\boldsymbol{g}\|}{\sigma}$ and $\angle(\boldsymbol{g}, \boldsymbol{x}_0 - \boldsymbol{x}^*) \in \left[0, \arccos\left(\frac{\sigma}{\|\boldsymbol{g}\|} \cdot \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|\right)\right) \cup \{0\}$. Using similar techniques as above, we can write

- $\|\boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\| \leq \frac{\|\boldsymbol{E}^\intercal \boldsymbol{g}\|}{\sigma}$ and
- $\angle(\boldsymbol{E}^\intercal \boldsymbol{g},\ \boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)) \in \left[0,\ \arccos\left(\frac{\sigma}{\|\boldsymbol{E}^\intercal \boldsymbol{g}\|}\|\boldsymbol{T}^{-1}(\boldsymbol{x}_0) - \boldsymbol{T}^{-1}(\boldsymbol{x}^*)\|\right)\right) \cup \{0\}$.

Using Lemma A.2.1, we have that there exists a function $f \in \mathcal{Q}^{(2)}(\boldsymbol{T}^{-1}(\boldsymbol{x}^*),\ \sigma)$ with the gradient $\nabla f(\boldsymbol{T}^{-1}(\boldsymbol{x}_0)) = \boldsymbol{E}^\intercal \boldsymbol{g}$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be such that $f(\boldsymbol{s}) = \frac{1}{2}\boldsymbol{s}^\intercal \boldsymbol{P}\boldsymbol{s} + \boldsymbol{b}^\intercal \boldsymbol{s} + c$ where $\boldsymbol{P} \in \mathfrak{S}^2$, $\boldsymbol{b} \in \mathbb{R}^2$ and $c \in \mathbb{R}$. To satisfy the conditions of f, we have $\boldsymbol{0} = \nabla f(\boldsymbol{T}^{-1}(\boldsymbol{x}^*)) = \nabla f(\boldsymbol{0}) = \boldsymbol{b}$,

$$\boldsymbol{E}^\intercal \boldsymbol{g} = \nabla f(\boldsymbol{T}^{-1}(\boldsymbol{x}_0)) = \nabla f(\boldsymbol{E}^\intercal(\boldsymbol{x}_0 - \boldsymbol{x}^*)) = \boldsymbol{P}\boldsymbol{E}^\intercal(\boldsymbol{x}_0 - \boldsymbol{x}^*), \tag{A.26}$$

and $\lambda_{\min}(\boldsymbol{P}) = \sigma$.

To construct a quadratic function $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ satisfying $\tilde{f} \in \mathcal{Q}^{(n)}(\boldsymbol{x}^*, \sigma)$ and $\nabla \tilde{f}(\boldsymbol{x}_0) = \boldsymbol{g}$, recall the expression of $\boldsymbol{e}_1'$ and $\boldsymbol{e}_2'$ from (A.23). Let $\{\tilde{\boldsymbol{v}}_1, \tilde{\boldsymbol{v}}_2, \ldots, \tilde{\boldsymbol{v}}_{n-2}\}$ be a set of unit vectors in $\mathbb{R}^n$ for which $\begin{bmatrix} \boldsymbol{e}_1' & \boldsymbol{e}_2' & \tilde{\boldsymbol{v}}_1 & \tilde{\boldsymbol{v}}_2 & \cdots & \tilde{\boldsymbol{v}}_{n-2} \end{bmatrix} \in \mathbb{R}^{n \times n}$ is orthogonal. Let $\widetilde{\boldsymbol{V}} = \begin{bmatrix} \tilde{\boldsymbol{v}}_1 & \tilde{\boldsymbol{v}}_2 & \cdots & \tilde{\boldsymbol{v}}_{n-2} \end{bmatrix} \in \mathbb{R}^{n \times (n-2)}$, $\widetilde{\boldsymbol{\Lambda}} = \operatorname{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_{n-2}) \in \mathbb{R}^{(n-2)\times(n-2)}$ such that $\{\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_{n-2}\}$ is chosen to satisfy $\widetilde{\boldsymbol{\Lambda}} \succeq \sigma \boldsymbol{I}$, and $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ be such that

$$\tilde{f}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\intercal \boldsymbol{Q}\boldsymbol{x} + (-\boldsymbol{Q}\boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{b})^\intercal \boldsymbol{x} + \left(\frac{1}{2}(\boldsymbol{x}^*)^\intercal \boldsymbol{Q}\boldsymbol{x}^* - \boldsymbol{b}^\intercal \boldsymbol{E}^\intercal \boldsymbol{x}^* + c\right),$$

where $\boldsymbol{Q} = \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^\intercal + \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{V}}^\intercal \in \mathbb{R}^{n\times n}$. Note that $\boldsymbol{E}^\intercal \boldsymbol{E} = \boldsymbol{I}$, $\widetilde{\boldsymbol{V}}^\intercal \widetilde{\boldsymbol{V}} = \boldsymbol{I}$ and $\widetilde{\boldsymbol{V}}^\intercal \boldsymbol{E} = \boldsymbol{0}$. Since $\boldsymbol{b} = \boldsymbol{0}$, we also have $\nabla \tilde{f}(\boldsymbol{x}^*) = \boldsymbol{Q}\boldsymbol{x}^* + (-\boldsymbol{Q}\boldsymbol{x}^* + \boldsymbol{E}\boldsymbol{b}) = \boldsymbol{0}$ and $\nabla \tilde{f}(\boldsymbol{x}_0) = \boldsymbol{Q}(\boldsymbol{x}_0 - \boldsymbol{x}^*)$. Using

$$\boldsymbol{E}\boldsymbol{E}^\intercal \boldsymbol{g} = \boldsymbol{E}\Big(\boldsymbol{E}^\intercal(\boldsymbol{x}_{\boldsymbol{g}} - \boldsymbol{x}^*)\Big) = \boldsymbol{E}\boldsymbol{T}^{-1}(\boldsymbol{x}_{\boldsymbol{g}}) = \boldsymbol{T}(\boldsymbol{T}^{-1}(\boldsymbol{x}_{\boldsymbol{g}})) - \boldsymbol{x}^* = \boldsymbol{g},$$

$\widetilde{\boldsymbol{V}}^\intercal(\boldsymbol{x}_0 - \boldsymbol{x}^*) = \boldsymbol{0}$ (since $\boldsymbol{x}_0 - \boldsymbol{x}^* \in \operatorname{span}(\{\boldsymbol{e}_1'\})$) and (A.26), we can write

$$\nabla \tilde{f}(\boldsymbol{x}_0) = \boldsymbol{Q}(\boldsymbol{x}_0 - \boldsymbol{x}^*) = \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^\intercal(\boldsymbol{x}_0 - \boldsymbol{x}^*) + \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{V}}^\intercal(\boldsymbol{x}_0 - \boldsymbol{x}^*) = \boldsymbol{E}\boldsymbol{E}^\intercal \boldsymbol{g} = \boldsymbol{g}.$$

It remains to show that $\lambda_{\min}(\boldsymbol{Q}) = \sigma$.

Suppose $\boldsymbol{P} \in \mathfrak{S}^2$ can be decomposed as $\boldsymbol{P} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\intercal$ where $\boldsymbol{V} \in \mathbb{R}^{2\times 2}$ is a matrix whose i-th column is the (unit) eigenvector $\boldsymbol{v}_i$ of $\boldsymbol{P}$, and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. We can rewrite $\boldsymbol{Q} = \boldsymbol{E}\boldsymbol{P}\boldsymbol{E}^\intercal + \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{V}}^\intercal$ as

$$\boldsymbol{Q} = \boldsymbol{E}(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\intercal)\boldsymbol{E}^\intercal + \widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{V}}^\intercal = \begin{bmatrix} \boldsymbol{E}\boldsymbol{V} & \widetilde{\boldsymbol{V}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{\Lambda}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}^\intercal \boldsymbol{E}^\intercal \\ \widetilde{\boldsymbol{V}}^\intercal \end{bmatrix}. \tag{A.27}$$

Since $\begin{bmatrix} \boldsymbol{E}\boldsymbol{V} & \widetilde{\boldsymbol{V}} \end{bmatrix}$ is an orthogonal matrix (using the fact that $(\boldsymbol{E}\boldsymbol{V})^\intercal \widetilde{\boldsymbol{V}} = \boldsymbol{0}$ and $\|\boldsymbol{E}\boldsymbol{v}_i\| = 1$ for $i \in \{1, 2\}$), we have that equation (A.27) is the eigendecomposition of $\boldsymbol{Q}$. By our construction of $\widetilde{\boldsymbol{\Lambda}}$ and $\lambda_{\min}(\boldsymbol{P}) = \sigma$, we have $\lambda_{\min}(\boldsymbol{Q}) = \sigma$. Thus, we conclude that $\tilde{f} \in \mathcal{Q}^{(n)}(\boldsymbol{x}^*, \sigma)$ with $\nabla \tilde{f}(\boldsymbol{x}_0) = \boldsymbol{g}$. □

A.2.2 Proof of Theorem 2.5.2

Proof of Theorem 2.5.2. **Part (i):** $r \in \left(0,\ \frac{L}{2\sigma_1}\right]$. The proof of the characterization of the boundary $\partial\mathcal{M}_\downarrow$ can be carried out in the same manner as in the proof of part (i) of Theorem 2.4.8. We are left to show the property related to the set $\mathcal{M}_\downarrow$ and the characterization of $(\mathcal{M}_\downarrow)^\circ$. By the definition of $\mathcal{M}_\downarrow$ in (2.14) and $\partial\mathcal{M}_\downarrow = \mathcal{T} \sqcup \{\boldsymbol{x}_1^*, \boldsymbol{x}_2^*\}$, we have that $\partial\mathcal{M}_\downarrow \subset (\mathcal{M}_\downarrow)^c$. Therefore, we can conclude that $\mathcal{M}_\downarrow$ is open. Since $\mathcal{M}_\downarrow$ is open, $(\mathcal{M}_\downarrow)^\circ = \mathcal{M}_\downarrow = \widetilde{\mathcal{T}}$, where $\widetilde{\mathcal{T}}$ is defined in (2.30).

Part (ii): $r \in \left(\frac{L}{2\sigma_1}, \frac{L}{2\sigma_2}\right]$. Consider points in the set $(\mathcal{H}_1^-)^c$. By proceeding the same steps as in the proof of (2.36), we obtain that

$$\mathcal{T} \cap (\mathcal{H}_1^-)^c = \emptyset. \tag{A.28}$$

By using the same reasoning as in the corresponding part of the proof of Theorem 2.4.8 part (ii), we obtain the results, which are similar to (2.37) and (2.34), that

$$\mathcal{M}_\downarrow \cap (\mathcal{H}_1^-)^c = \overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c \quad \text{and} \quad \partial\mathcal{M}_\downarrow \cap (\mathcal{H}_1^-)^c = \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c. \tag{A.29}$$

Then, recall the definition of λ_1 and ν_1 from (2.27) and (2.28), respectively. Consider points in the set $\mathcal{H}_1^+ \cap \mathcal{H}_1^-$, i.e., $\{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1\}$. Considering three disjoint regions in $\mathcal{H}_1^+ \cap \mathcal{H}_1^-$: $\|\tilde{\mathbf{z}}\| > \nu_1$, $\|\tilde{\mathbf{z}}\| < \nu_1$, and $\|\tilde{\mathbf{z}}\| = \nu_1$, we obtain similar results as in the corresponding part of the proof of Theorem 2.4.8 part (ii) summarized as follows:

$$\begin{cases} \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| > \nu_1\} \subseteq ((\mathcal{M}_\downarrow)^c)^\circ \cap \mathcal{T}^c, \\ \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| < \nu_1\} \subseteq (\mathcal{M}_\downarrow)^\circ \cap \mathcal{T}^c, \\ \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| = \nu_1\} = \mathcal{C}_1 \subseteq \partial\mathcal{M}_\downarrow \cap \mathcal{T}. \end{cases} \tag{A.30}$$

Combining these results, we get a similar result as in (2.38), i.e.,

$$\partial\mathcal{M}_\downarrow \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) = \mathcal{C}_1 = \mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-). \tag{A.31}$$

From the first equation in (A.29), (A.30), and $\mathcal{C}_1 \subseteq (\mathcal{M}_\downarrow)^c$ (since $\mathcal{C}_1 \subseteq \mathcal{T}$), we have

$$\begin{aligned} \mathcal{M}_\downarrow \cap \mathcal{H}_1^+ &= \left[\mathcal{M}_\downarrow \cap (\mathcal{H}_1^-)^c\right] \cup \left[\mathcal{M}_\downarrow \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\right] \\ &= \left[\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c\right] \cup \{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| < \nu_1\}. \end{aligned} \tag{A.32}$$

Next, consider points in the set $(\mathcal{H}_1^+)^c$. Recall the definition of φ from (2.31). By using the same reasoning as in the proof of (2.39) and (2.41), we obtain that

$$\partial(\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c \subseteq \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) < 0\} \cup \{\boldsymbol{x}_1^*\}, \tag{A.33}$$

and

$$\partial\mathcal{M}_\downarrow \cap (\mathcal{H}_1^+)^c = \left[\mathcal{T} \cap (\mathcal{H}_1^+)^c\right] \sqcup \{\boldsymbol{x}_1^*\}, \tag{A.34}$$

respectively. Using (A.28), we can write $\mathcal{T} = \left[\mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\right] \sqcup \left[\mathcal{T} \cap (\mathcal{H}_1^+)^c\right]$, and combining the second equation in (A.29), equation (A.31) and equation (A.34) together yields the characterization of $\partial\mathcal{M}_\downarrow$:

$$\partial\mathcal{M}_\downarrow = \left[\partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c\right] \sqcup \overbrace{\left[\mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\right] \sqcup \left[\mathcal{T} \cap (\mathcal{H}_1^+)^c\right]}^{=\mathcal{T}} \sqcup \{\boldsymbol{x}_1^*\}.$$

For the characterization of $(\mathcal{M}_\downarrow)^\circ$, we can also use the same technique as in the proof of Theorem 2.4.8 part (ii).

For the property related to the set $\mathcal{M}_\downarrow$, recall the definition of $\mathcal{C}_1$ from (2.45). Using equation (A.32) and $\{\boldsymbol{z} \in \mathbb{R}^n : z_1 = \lambda_1,\ \|\tilde{\mathbf{z}}\| \leq \nu_1\} = \overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)$, we have

$$(\mathcal{M}_\downarrow \cup \mathcal{C}_1) \cap \mathcal{H}_1^+ = (\mathcal{M}_\downarrow \cap \mathcal{H}_1^+) \cup \mathcal{C}_1 = \left[\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^-)^c\right] \cup \left[\overline{\mathcal{B}}_1 \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-)\right] = \overline{\mathcal{B}}_1 \cap \mathcal{H}_1^+, \tag{A.35}$$

which is closed. Next, from (A.33), we can write

$$\mathcal{M}_\downarrow \subseteq \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) \geq 0\} \setminus \{\boldsymbol{x}_1^*\} \subseteq \Big(\partial(\mathcal{B}_1 \cap \mathcal{B}_2)\Big)^c \cup (\mathcal{H}_1^+)^c.$$

Using the fact that $\mathcal{M}_{\downarrow} \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, the above inclusion implies that $\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+)^c \subseteq (\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c$. Suppose $\boldsymbol{x} \in \mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+)^c$. However, from the definition of $\mathcal{M}_{\downarrow}$ in (2.14), we can write

$$\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+)^c = \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) > 0\} \cap (\mathcal{H}_1^+)^c = \{\boldsymbol{z} \in \mathcal{B}_1 \cap \mathcal{B}_2 : \varphi(\boldsymbol{z}) > 0\} \cap (\mathcal{H}_1^+)^c. \quad \text{(A.36)}$$

Since φ is continuous, and $\mathcal{B}_1 \cap \mathcal{B}_2$ and $(\mathcal{H}_1^+)^c$ are open, there exists $\epsilon \in \mathbb{R}_{>0}$ such that for all $\boldsymbol{x}_0 \in \mathcal{B}(\boldsymbol{x}, \epsilon)$ such that $\boldsymbol{x}_0 \in \{\boldsymbol{z} \in \mathbb{R}^n : \varphi(\boldsymbol{z}) > 0\} \cap (\mathcal{B}_1 \cap \mathcal{B}_2) \cap (\mathcal{H}_1^+)^c = \mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+)^c$. Thus, we conclude that $\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+)^c$ is open.

Part (iii): $r \in \left(\frac{L}{2\sigma_2}, \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)\right)$. In this case, we can use similar argument as in the proof of part (ii) to show the following statements.

$$\begin{cases} \partial\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^-)^c = \partial\mathcal{B}_1 \cap (\mathcal{H}_1^-)^c, & \text{(similar to proving (A.29))} \\ \partial\mathcal{M}_{\downarrow} \cap (\mathcal{H}_2^+)^c = \partial\mathcal{B}_2 \cap (\mathcal{H}_2^+)^c, & \text{(similar to proving (A.29))} \\ \partial\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-) = \mathcal{T} \cap (\mathcal{H}_1^+ \cap \mathcal{H}_1^-), & \text{(similar to proving (A.31))} \\ \partial\mathcal{M}_{\downarrow} \cap (\mathcal{H}_2^+ \cap \mathcal{H}_2^-) = \mathcal{T} \cap (\mathcal{H}_2^+ \cap \mathcal{H}_2^-), & \text{(similar to proving (A.31))} \\ \partial\mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c = \mathcal{T} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c. & \text{(similar to proving (A.34))} \end{cases}$$

Combining these equations, we obtain the characterization of $\partial\mathcal{M}_{\downarrow}$. For the characterization of $(\mathcal{M}^{\uparrow})^{\circ}$, we can use the same technique as shown in the proof of Theorem 2.4.8 part (ii). For the property related to the set $\mathcal{M}_{\downarrow}$, recall the definition of $\mathcal{C}_i$ from (2.45). Using a similar approach to part (ii), we can show that

$$\begin{cases} (\mathcal{M}_{\downarrow} \cup \mathcal{C}_1) \cap \mathcal{H}_1^+ = \overline{\mathcal{B}}_1 \cap \mathcal{H}_1^+ \text{ which is closed}, & \text{(similar to proving (A.35))} \\ (\mathcal{M}_{\downarrow} \cup \mathcal{C}_2) \cap \mathcal{H}_2^- = \overline{\mathcal{B}}_2 \cap \mathcal{H}_2^- \text{ which is closed}, & \text{(similar to proving (A.35))} \\ \mathcal{M}_{\downarrow} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c = \{\boldsymbol{z} \in \mathcal{B}_1 \cap \mathcal{B}_2 : \varphi(\boldsymbol{z}) > 0\} \cap (\mathcal{H}_1^+ \cup \mathcal{H}_2^-)^c \\ \qquad\qquad\qquad\qquad \text{which is open.} \quad \text{(similar to proving (A.36))} \end{cases}$$

Part (iv): $r = \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$. In this case, we obtain that $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2 = \left\{\left(\frac{L}{2}\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right),\ \mathbf{0}\right)\right\}$. Suppose $\boldsymbol{x} = \left(\frac{L}{2}\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right),\ \mathbf{0}\right)$. Since $\mathcal{M}_{\downarrow} \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, we only need to check the point $\boldsymbol{x}$. However, the point $\boldsymbol{x} \in \mathcal{X}$ where $\mathcal{X}$ is defined in (2.12), and we obtain the result due to the definition of $\mathcal{M}_{\downarrow}$.

Part (v): $r \in \left(\frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right),\ \infty\right)$. Since $r > \frac{L}{2}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right)$, we have $\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2 = \emptyset$. Since $\mathcal{M}_{\downarrow} \subseteq \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2$, we conclude that $\mathcal{M}_{\downarrow} = \emptyset$. □

B. SUPPLEMENTARY MATERIALS FOR CHAPTER 4

B.1 Additional Lemma

We provide a lemma which is utilized in the proof of Theorem 4.4.10 and Lemma 4.4.12.

Lemma B.1.1. *For given $\hat{\boldsymbol{x}} \in \mathbb{R}^d$ and $R \geq 0$, if $\boldsymbol{x} \notin \mathcal{B}(\hat{\boldsymbol{x}}, R)$ then*

$$\max_{\boldsymbol{y} \in \mathcal{B}(\hat{\boldsymbol{x}}, R)} \angle(\boldsymbol{x} - \boldsymbol{y},\ \boldsymbol{x} - \hat{\boldsymbol{x}}) = \arcsin\left(\frac{R}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}\right).$$

Proof. Since the angle is measured with respect to the vector $\boldsymbol{x} - \hat{\boldsymbol{x}}$, consider any 2-D planes passed through the center $\hat{\boldsymbol{x}}$ and the point $\boldsymbol{x}$. Since the planes pass through $\hat{\boldsymbol{x}}$, the intersections between of the ball $\mathcal{B}(\hat{\boldsymbol{x}}, R)$ and the planes are great circles of radius R. Thus, all of the intersections generated from each plane are identical and we can consider the angle using a great circle instead of the ball. From geometry, the maximum angle $\phi = \angle(\boldsymbol{x} - \boldsymbol{y}^*,\ \boldsymbol{x} - \hat{\boldsymbol{x}})$ only occurs when the ray starting from the point $\boldsymbol{x}$ touches the circle at point $\boldsymbol{y}^*$. Therefore, $\angle(\hat{\boldsymbol{x}} - \boldsymbol{y}^*, \boldsymbol{x} - \boldsymbol{y}^*) = \frac{\pi}{2}$ and $\|\hat{\boldsymbol{x}} - \boldsymbol{y}^*\| = R$. We have

$$\sin\phi = \frac{\|\hat{\boldsymbol{x}} - \boldsymbol{y}^*\|}{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|} = \frac{R}{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|}$$

and the result follows. □

B.2 Proof of the Auxiliary States Proposition

Proof of Proposition 4.4.1. For any $\mathcal{S} \subseteq \mathcal{V}$, $\zeta \in \mathbb{R}$ and $\bar{k}, k \in \mathbb{N}$ with $\bar{k} \geq k$, define the sets

$$\mathcal{J}_M^{(\ell)}(\mathcal{S}, \bar{k}, k, \zeta) := \{v_i \in \mathcal{S} : y_i^{(\ell)}[\bar{k}] > M^{(\ell)}[k] - \zeta\},$$
$$\mathcal{J}_m^{(\ell)}(\mathcal{S}, \bar{k}, k, \zeta) := \{v_i \in \mathcal{S} : y_i^{(\ell)}[\bar{k}] < m^{(\ell)}[k] + \zeta\}.$$

Consider a fixed $\ell \in \{1, 2, \ldots, d\}$ and any time-step $k \in \mathbb{N}$. Define $\zeta_0^{(\ell)} = \frac{1}{2}D^{(\ell)}[k]$. Note that the set $\mathcal{J}_M^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)}) \cap \mathcal{J}_m^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)}) = \emptyset$.

By the definition of these sets, when $D^{(\ell)}[k] > 0$, the sets $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)}) \neq \emptyset$ and $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)}) \neq \emptyset$. Since the graph is $(2F+1)$-robust, at least one of $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$

or $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$ is $(2F+1)$-reachable which means that at least one of them contains a vertex that has at least $2F+1$ in-neighbors from outside it.

If such a node v_i is in $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$, we claim that in the update, v_i cannot use the values strictly greater than $M^{(\ell)}[k]$ and it uses at least one value from $\mathcal{V} \setminus \mathcal{J}_M^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)})$. To show the first claim, note that the nodes that possess the value (in ℓ-component) greater than $M^{(\ell)}[k]$ must be Byzantine agents by the definition of $M^{(\ell)}[k]$. Since the regular node v_i discards up to F-highest values and there are at most F Byzantine in-neighbors, the Byzantine agents that hold the value greater than $M^{(\ell)}[k]$ must be discarded. To show the second claim, let $\mathcal{S}_1^{(\ell)}[k] = \mathcal{J}_M^{(\ell)}(\mathcal{A}, k, k, \zeta_0^{(\ell)})$ and $\mathcal{S}_2^{(\ell)}[k] = \mathcal{V} \setminus \mathcal{J}_M^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)})$ to simplify the notation. We have

- $\mathcal{V} \setminus \mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)}) = \mathcal{S}_1^{(\ell)}[k] \cup \mathcal{S}_2^{(\ell)}[k]$, and
- $\mathcal{S}_1^{(\ell)}[k] \cap \mathcal{S}_2^{(\ell)}[k] = \emptyset$.

From Assumption 4.4.4, we have $|\mathcal{S}_1^{(\ell)}[k] \cap \mathcal{N}_i^{\text{in}}| \leq F$. Applying $(2F+1)$-reachable property of v_i and two above properties, we obtain that $|\mathcal{S}_2^{(\ell)}[k] \cap \mathcal{N}_i^{\text{in}}| \geq F+1$. Let $\bar{\mathcal{V}}_i^{(\ell)}[k] \subseteq \mathcal{N}_i^{\text{in}}$ be the set of nodes that $v_i \in \mathcal{R}$ discards their values in dimension ℓ at time-step k. From the fact that $v_i \in \mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$, and **Line 11** of Algorithm 2, we know that $|\mathcal{S}_2^{(\ell)}[k] \cap \bar{\mathcal{V}}_i^{(\ell)}[k]| \leq F$. Combining this with the former statement, we can conclude that v_i uses at least one value from $\mathcal{S}_2^{(\ell)}[k]$ in its update, i.e., $(\mathcal{S}_2^{(\ell)}[k] \cap \mathcal{N}_i^{\text{in}}) \setminus \bar{\mathcal{V}}_i^{(\ell)}[k] \neq \emptyset$.

Consider the auxiliary point update rule (4.7) (in **Line 12** of Algorithm 2). We can rewrite the update as

$$y_i^{(\ell)}[k+1] = \sum_{v_j \in (\mathcal{S}_2^{(\ell)}[k] \cap \mathcal{N}_i^{\text{in}}) \setminus \bar{\mathcal{V}}_i^{(\ell)}[k]} w_{y,ij}^{(\ell)}[k]\; y_j^{(\ell)}[k] + \sum_{v_j \in \left(v_i \cup (\mathcal{J}_M(\mathcal{V}, k, k, \zeta_0^{(\ell)}) \cap \mathcal{N}_i^{\text{in}})\right) \setminus \bar{\mathcal{V}}_i^{(\ell)}[k]} w_{y,ij}^{(\ell)}[k]\; y_j^{(\ell)}[k].$$

Since $y_j^{(\ell)}[k]$ on the first and second terms on the RHS are upper bounded by $M^{(\ell)}[k]-\zeta_0^{(\ell)}$ and $M^{(\ell)}[k]$, respectively, and the non-zero weights $w_{y,ij}^{(\ell)}[k]$ are lower bounded by the constant ω (Assumption 4.4.5), the value of this node at the next time-step is upper bounded as

$$y_i^{(\ell)}[k+1] \leq \omega(M^{(\ell)}[k] - \zeta_0^{(\ell)}) + (1-\omega)M^{(\ell)}[k] = M^{(\ell)}[k] - \omega\zeta_0^{(\ell)}.$$

Note that the above bound is applicable to any node that is in $\mathcal{R}\backslash\mathcal{J}_M^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)})$, since such a node will use its own value in its update. Similarly, if there is a node $v_j \in \mathcal{J}_m^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$ that uses the value of a node outside that set, then $y_j^{(\ell)}[k+1] \geq m^{(\ell)}[k] + \omega\zeta_0^{(\ell)}$. This bound is also applicable to any node that is in $\mathcal{R} \setminus \mathcal{J}_m^{(\ell)}(\mathcal{V}, k, k, \zeta_0^{(\ell)})$.

Now, define the quantity $\zeta_1^{(\ell)} = \omega\zeta_0^{(\ell)}$. We have that the set $\mathcal{J}_M^{(\ell)}(\mathcal{V}, k+1, k, \zeta_1^{(\ell)}) \cap \mathcal{J}_m^{(\ell)}(\mathcal{V}, k+1, k, \zeta_1^{(\ell)}) = \emptyset$. Furthermore, by the bounds provided above, we see that at least one of the following must be true:

$$|\mathcal{J}_M^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)})| < |\mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})|, \quad \text{or}$$
$$|\mathcal{J}_m^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)})| < |\mathcal{J}_m^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})|.$$

If $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)}) \neq \emptyset$ and $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)}) \neq \emptyset$, then again by the fact that the graph is $(2F+1)$-robust, there is at least one node in one of these sets that has at least $2F+1$ in-neighbors outside from the set. Suppose $v_i \in \mathcal{J}_M^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)})$ is such a node. Then, v_i cannot use the values strictly greater than $M^{(\ell)}[k+1]$ and it uses at least one value from $\mathcal{V} \setminus \mathcal{J}_M^{(\ell)}(\mathcal{V}, k+1, k, \zeta_1^{(\ell)})$. Since at time-step k, all regular nodes cannot use values that are strictly greater than $M^{(\ell)}[k]$ in the update, we have that $M^{(\ell)}[k+1] \leq M^{(\ell)}[k]$. Therefore, the value of node v_i at the next time-step is upper bounded as

$$y_i^{(\ell)}[k+2] \leq \omega(M^{(\ell)}[k] - \zeta_1^{(\ell)}) + (1-\omega)M^{(\ell)}[k+1] \leq M^{(\ell)}[k] - \omega^2\zeta_0^{(\ell)}.$$

Again, this upper bound also holds for any regular node that is in $\mathcal{R} \setminus \mathcal{J}_M^{(\ell)}(\mathcal{V}, k+1, k, \zeta_1^{(\ell)})$. Similarly, if there is a node $v_j \in \mathcal{J}_m^{(\ell)}(\mathcal{R}, k+1, k, \zeta_1^{(\ell)})$ that has $2F+1$ in-neighbors from

outside that set then $y_j^{(\ell)}[k+2] \geq m^{(\ell)}[k] + \omega^2 \zeta_0^{(\ell)}$. This bound also holds for any regular node that is not in the set $\mathcal{R} \setminus \mathcal{J}_m^{(\ell)}(\mathcal{V}, k+1, k, \zeta_1^{(\ell)})$.

We continue in this manner by defining $\zeta_s^{(\ell)} = \omega^s \zeta_0^{(\ell)}$ for $s \in \mathbb{N}$. At each time step $k+s$, if both $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k+s, k, \zeta_s^{(\ell)}) \neq \emptyset$ and $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k+s, k, \zeta_s^{(\ell)}) \neq \emptyset$ then at least one of these sets will shrink in the next time-step. If either of the sets is empty, then it will stay empty at the next time-step, since every regular node outside that set will have its value upper bounded by $M^{(\ell)}[k] - \zeta_s^{(\ell)}$ or lower bounded by $m^{(\ell)}[k] + \zeta_s^{(\ell)}$. After $|\mathcal{R}| - 1$ time-steps, at least one of the sets $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k+|\mathcal{R}|-1, k, \zeta_{|\mathcal{R}|-1}^{(\ell)})$ or $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k+|\mathcal{R}|-1, k, \zeta_{|\mathcal{R}|-1}^{(\ell)})$ must be empty since the sets $\mathcal{J}_M^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$ and $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k, k, \zeta_0^{(\ell)})$ can contain at most $\mathcal{R} - 1$ regular nodes. Suppose the former set is empty; this means that

$$M^{(\ell)}[k+|\mathcal{R}|-1] \leq M^{(\ell)}[k] - \zeta_{|\mathcal{R}|-1}^{(\ell)}.$$

Since $m^{(\ell)}[k+|\mathcal{R}|-1] \geq m^{(\ell)}[k]$, we obtain

$$D^{(\ell)}[k+|\mathcal{R}|-1] \leq D^{(\ell)}[k] - \zeta_{|\mathcal{R}|-1}^{(\ell)} = \left(1 - \frac{\omega^{|\mathcal{R}|-1}}{2}\right) D^{(\ell)}[k] = \gamma D^{(\ell)}[k]. \tag{B.1}$$

The first equality comes from the fact that $\zeta_s^{(\ell)} = \omega^s \zeta_0^{(\ell)}$ and $\zeta_0^{(\ell)} = \frac{1}{2} D^{(\ell)}[k]$. The same expression as (B.1) arises if the set $\mathcal{J}_m^{(\ell)}(\mathcal{R}, k+|\mathcal{R}|-1, k, \zeta_{|\mathcal{R}|-1}^{(\ell)}) = \emptyset$.

Using the fact that

$$\left[m^{(\ell)}[k+1],\ M^{(\ell)}[k+1]\right] \subseteq \left[m^{(\ell)}[k],\ M^{(\ell)}[k]\right] \tag{B.2}$$

for all $k \in \mathbb{N}$ and the inequality (B.1), we can conclude that for all $v_i \in \mathcal{R}$, $\lim_{k\to\infty} y_i^{(\ell)}[k] = y^{(\ell)}[\infty]$ exists and for all k, we have

$$y^{(\ell)}[\infty] \in \left[m^{(\ell)}[k],\ M^{(\ell)}[k]\right]. \tag{B.3}$$

This completes the first part of the proof.

For the second part, let consider the quantity $D^{(\ell)}[k]$ as follows. For all $k \in \mathbb{N}$, we can write

$$D^{(\ell)}[k] \leq D^{(\ell)}\left[\left\lfloor \frac{k}{|\mathcal{R}|-1} \right\rfloor (|\mathcal{R}|-1)\right] \leq \gamma^{\left\lfloor \frac{k}{|\mathcal{R}|-1} \right\rfloor} D^{(\ell)}[0] < \gamma^{\frac{k}{|\mathcal{R}|-1}-1} D^{(\ell)}[0]. \tag{B.4}$$

The first inequality is obtained by using $x \geq \lfloor x \rfloor$ and (B.2). To obtain the second inequality, we apply the inequality (B.1) $\left\lfloor \frac{k}{|\mathcal{R}|-1} \right\rfloor$ times. The last inequality comes from the fact that $\gamma < 1$ and $\lfloor x \rfloor > x - 1$ implies $\gamma^{\lfloor x \rfloor} < \gamma^{x-1}$. From (B.3) and (B.4), for all $v_i \in \mathcal{R}$, we have

$$|y_i^{(\ell)}[k] - y^{(\ell)}[\infty]| \leq D^{(\ell)}[k] < \gamma^{\frac{k}{|\mathcal{R}|-1}-1} D^{(\ell)}[0]. \tag{B.5}$$

Since the inequality (B.5) holds for all $\ell \in \{1, 2, \ldots, d\}$, we have

$$\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|^2 = \sum_{\ell=1}^{d} |y_i^{(\ell)}[k] - y^{(\ell)}[\infty]|^2 < \gamma^{2\left(\frac{k}{|\mathcal{R}|-1}-1\right)} \sum_{\ell=1}^{d} \left(D^{(\ell)}[0]\right)^2.$$

Taking square root of both sides yields

$$\|\boldsymbol{y}_i[k] - \boldsymbol{y}_i[\infty]\| < \gamma^{\frac{k}{|\mathcal{R}|-1}-1} \|\boldsymbol{D}[0]\| = \frac{1}{\gamma} \|\boldsymbol{D}[0]\| \, e^{-\frac{1}{|\mathcal{R}|-1} \log(\frac{1}{\gamma}) \, k},$$

which completes the proof. □

B.3 Proof of Proposition 4.4.2

Proof of Proposition 4.4.2. Consider a regular agent $v_i \in \mathcal{R}$. From Assumption 4.4.1, for all $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^d$, we have $f_i(\boldsymbol{y}) \geq f_i(\boldsymbol{x}) + \langle \tilde{\boldsymbol{g}}_i(\boldsymbol{x}), \ \boldsymbol{y} - \boldsymbol{x} \rangle$, where $\tilde{\boldsymbol{g}}_i(\boldsymbol{x}) \in \partial f_i(\boldsymbol{x})$. Substitute a minimizer $\boldsymbol{x}_i^*$ of the function f_i into the variable $\boldsymbol{y}$ to get

$$-\langle \tilde{\boldsymbol{g}}_i(\boldsymbol{x}), \ \boldsymbol{x}_i^* - \boldsymbol{x} \rangle \geq f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*). \tag{B.6}$$

Let $\hat{\theta}_i(\boldsymbol{x}) = \angle(\tilde{\boldsymbol{g}}_i(\boldsymbol{x}), \ \boldsymbol{x} - \boldsymbol{x}_i^*)$. The inequality (B.6) becomes

$$\|\tilde{\boldsymbol{g}}_i(\boldsymbol{x})\| \, \|\boldsymbol{x}_i^* - \boldsymbol{x}\| \cos \hat{\theta}_i(\boldsymbol{x}) \geq f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*).$$

Fix $\epsilon \in \mathbb{R}_{>0}$, and suppose that $\boldsymbol{x} \notin \mathcal{C}_i(\epsilon)$. From Assumption 4.4.2, applying $\|\tilde{\boldsymbol{g}}_i(\boldsymbol{x})\| \leq L$, we have

$$\cos \hat{\theta}_i(\boldsymbol{x}) \geq \frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}^*)}{L \, \|\boldsymbol{x}_i^* - \boldsymbol{x}\|}. \tag{B.7}$$

Let $\tilde{\boldsymbol{x}}_i \in \mathbb{R}^d$ be the point on the line connecting $\boldsymbol{x}_i^*$ and $\boldsymbol{x}$ such that $f_i(\tilde{\boldsymbol{x}}_i) = f_i(\boldsymbol{x}_i^*) + \epsilon$. We can rewrite the point $\boldsymbol{x}$ as

$$\boldsymbol{x} = \boldsymbol{x}_i^* + t(\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*) \quad \text{where} \quad t = \frac{\|\boldsymbol{x} - \boldsymbol{x}_i^*\|}{\|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*\|} \geq 1.$$

Consider the term on the RHS of (B.7). Since $\tilde{\boldsymbol{x}}_i \in \mathcal{C}_i(\epsilon)$, and (4.12) holds, we have

$$\begin{aligned} \frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*)}{\|\boldsymbol{x} - \boldsymbol{x}_i^*\|} &= \frac{f_i(\boldsymbol{x}_i^* + t(\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*)) - f_i(\boldsymbol{x}_i^*)}{t\|\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*\|} \\ &\geq \frac{f_i(\boldsymbol{x}_i^* + t(\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*)) - f_i(\boldsymbol{x}_i^*)}{t \cdot \max_{\boldsymbol{y} \in \mathcal{C}_i(\epsilon)} \|\boldsymbol{y} - \boldsymbol{x}_i^*\|} \\ &\geq \frac{f_i(\boldsymbol{x}_i^* + t(\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*)) - f_i(\boldsymbol{x}_i^*)}{t \; \delta_i(\epsilon)}. \end{aligned} \tag{B.8}$$

Since the quantity $\frac{f_i(\boldsymbol{x}_i^* + t(\tilde{\boldsymbol{x}}_i - \boldsymbol{x}_i^*)) - f_i(\boldsymbol{x}_i^*)}{t}$ is non-decreasing in $t \in [1, \infty)$ [97, Lemma 2.80], the inequality (B.8) becomes

$$\frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*)}{\|\boldsymbol{x} - \boldsymbol{x}_i^*\|} \geq \frac{f_i(\tilde{\boldsymbol{x}}_i) - f_i(\boldsymbol{x}_i^*)}{\delta_i(\epsilon)} = \frac{\epsilon}{\delta_i(\epsilon)}. \tag{B.9}$$

Therefore, combining (B.7) and (B.9), we obtain

$$\cos \hat{\theta}_i(\boldsymbol{x}) \geq \frac{\epsilon}{L\delta_i(\epsilon)}. \tag{B.10}$$

However, from Assumption 4.4.1, we have

$$f_i(\boldsymbol{x}_i^*) \geq f_i(\tilde{\boldsymbol{x}}_i) + \langle \tilde{\boldsymbol{g}}_i(\tilde{\boldsymbol{x}}_i), \; \boldsymbol{x}_i^* - \tilde{\boldsymbol{x}}_i \rangle$$

where $\tilde{\boldsymbol{g}}_i(\tilde{\boldsymbol{x}}_i) \in \partial f_i(\tilde{\boldsymbol{x}}_i)$. Since $\|\tilde{\boldsymbol{g}}_i(\tilde{\boldsymbol{x}}_i)\| \leq L$ by Assumption 4.4.2 and $\|\boldsymbol{x}_i^* - \tilde{\boldsymbol{x}}_i\| \leq \delta_i(\epsilon)$, we get

$$\epsilon = f_i(\tilde{\boldsymbol{x}}_i) - f_i(\boldsymbol{x}_i^*) \leq -\langle \tilde{\boldsymbol{g}}_i(\tilde{\boldsymbol{x}}_i),\ \boldsymbol{x}_i^* - \tilde{\boldsymbol{x}}_i \rangle \leq L\delta_i(\epsilon).$$

From $\epsilon \in \mathbb{R}_{>0}$ and the above inequality, the inequality (B.10) becomes

$$\hat{\theta}_i(\boldsymbol{x}) \leq \arccos\left(\frac{\epsilon}{L\delta_i(\epsilon)}\right) := \theta_i(\epsilon) < \frac{\boldsymbol{\pi}}{2},$$

which completes the proof. □

B.4 Proof of Lemmas 4.4.11 - 4.4.14

Proof of Lemma 4.4.11. From Proposition 4.4.1, the limit point $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ exists. Consider a time-step $k \in \mathbb{N}$ such that $k \geq k_1^*$. Using the gradient step (4.6), we can write

$$\begin{aligned}\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| &= \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty] - \eta[k]\ \boldsymbol{g}_i[k]\| \\ &\leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| + \eta[k]\ \|\boldsymbol{g}_i[k]\|.\end{aligned}$$

Using Assumptions 4.4.2 and 4.4.3, and $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq \max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\}$, we obtain

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \leq \max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\} + \eta[k_1^*]L.$$

By the definition of k_1^* and s^* in (4.15), the above inequality becomes

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \leq \max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\} + \xi \leq s^*,$$

which completes the proof. □

Proof of Lemma 4.4.12. From Proposition 4.4.1, the limit point $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ exists. Consider an agent $v_i \in \mathcal{R}$ and a time-step $k \in \mathbb{N}$ for which the condition in the lemma holds. Since $\boldsymbol{x}_i^* \in \mathcal{B}(\boldsymbol{y}[\infty], \tilde{R}_i)$ from (4.14), we have

$$\begin{aligned}\angle(\boldsymbol{x}_i^* - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]) \quad &\leq \max_{\boldsymbol{u} \in \mathcal{B}(\boldsymbol{y}[\infty], \tilde{R}_i)} \angle(\boldsymbol{u} - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]) \\ &= \arcsin \frac{\tilde{R}_i}{\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|}, \end{aligned} \tag{B.11}$$

where the last step is from using Lemma B.1.1. Using the gradient step (4.6), we can write $\angle(\boldsymbol{x}_i[k+1] - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]) = \angle(-\eta[k]\boldsymbol{g}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k])$. Since for all $v_i \in \mathcal{R}$ and $k \in \mathbb{N}$,

$$\angle(-\eta[k]\boldsymbol{g}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]) \leq \angle(-\eta[k]\boldsymbol{g}_i[k],\ \boldsymbol{x}_i^* - \boldsymbol{z}_i[k]) + \angle(\boldsymbol{x}_i^* - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k])$$

by [57, Corollary 12], applying Proposition 4.4.2 and inequality (B.11), we have

$$\angle(\boldsymbol{x}_i[k+1] - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]) \leq \theta_i + \arcsin \frac{\tilde{R}_i}{\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|} := \psi_i[k]. \tag{B.12}$$

Note that $\psi_i[k] \in [0, \boldsymbol{\pi})$ since $\theta_i \in \left[0, \frac{\pi}{2}\right)$ and $\arcsin \frac{\tilde{R}_i}{\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|} \in \left[0, \frac{\pi}{2}\right]$. Then, consider the triangle which has the vertices at $\boldsymbol{x}_i[k+1]$, $\boldsymbol{z}_i[k]$, and $\boldsymbol{y}[\infty]$. We can calculate the square of the distance by using the law of cosines:

$$\begin{aligned}\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 = \ &\|\boldsymbol{x}_i[k+1] - \boldsymbol{z}_i[k]\|^2 + \|\boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]\|^2 \\ &- 2\|\boldsymbol{x}_i[k+1] - \boldsymbol{z}_i[k]\| \cdot \|\boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]\| \cos \angle(\boldsymbol{x}_i[k+1] - \boldsymbol{z}_i[k],\ \boldsymbol{y}[\infty] - \boldsymbol{z}_i[k]).\end{aligned}$$

Using the gradient step (4.6) and the inequality (B.12), we get

$$\begin{aligned}\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 \leq \eta^2[k]\ \|\boldsymbol{g}_i[k]\|^2 + \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|^2 \\ - 2\ \eta[k]\ \|\boldsymbol{g}_i[k]\| \cdot \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \cos \psi_i[k].\end{aligned} \tag{B.13}$$

In addition, we can simplify the term $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \cos\psi_i[k]$ in the above inequality using the definition of $\psi_i[k]$ in (B.12). Regarding this, we can write

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \cos\psi_i[k] = \sqrt{\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|^2 - \tilde{R}_i^2} \cdot \cos\theta_i - \tilde{R}_i \sin\theta_i.$$

Substituting the above equation into (B.13), we obtain the result. □

Proof of Lemma 4.4.13. First, note that by the definition of s^* in (4.15), we have $a_i^+ > 0$ and $a_i^- < 0$ since $s^* \geq \tilde{R}_i + \xi$, and $b_i > 0$ since $s^* \geq \tilde{R}_i \sec\theta_i + \xi$.

For $v_i \in \mathcal{R}$, let $\Gamma_i : [\tilde{R}_i, \infty) \times \mathbb{R}_+ \to \mathbb{R}$ be the function

$$\Gamma_i(p, l) := p^2 - \Delta_i(p, l), \tag{B.14}$$

where function Δ_i is defined in (4.16). Consider an agent $v_i \in \mathcal{R}$ and a time-step $k \in \mathbb{N}$ such that $k \geq k_2^*$. We can compute the second derivative of $\Gamma_i(p, l)$ with respect to p as follows:

$$\frac{\partial^2 \Gamma_i}{\partial p^2} = 2 + 2l\tilde{R}_i^2 (p^2 - \tilde{R}_i^2)^{-\frac{3}{2}} \cos\theta_i.$$

Note that $\frac{\partial^2 \Gamma_i}{\partial p^2} > 0$ for all $p \in (\tilde{R}_i, \infty)$. This implies that

$$\sup_{p \in (\max_{v_j \in \mathcal{R}}\{\tilde{R}_j + \delta_j\},\, s^*]} \Gamma_i(p, l) \leq \max_{p \in [\tilde{R}_i,\, s^*]} \Gamma_i(p, l) = \max\Big\{\Gamma_i(\tilde{R}_i, l),\ \Gamma_i(s^*, l)\Big\}. \tag{B.15}$$

First, let consider $\Gamma_i(\tilde{R}_i,\ l)$. From the definition of Γ_i in (B.14), we have that

$$\Gamma_i(\tilde{R}_i, l) \leq (s^*)^2 \quad \Longleftrightarrow \quad l \in [a_i^-,\ a_i^+], \tag{B.16}$$

where a_i^+ and a_i^- are defined in (4.18). Using Assumption 4.4.2 and 4.4.3, and the definition of k_2^*, we have that

$$\eta[k]\ \|\boldsymbol{g}_i[k]\| \leq \eta[k] L \leq \min_{v_j \in \mathcal{R}} \Big\{ \min\{a_j^+,\ b_j\} \Big\} \leq a_i^+.$$

By the above inequality and statement (B.16), we obtain that

$$\Gamma_i(\tilde{R}_i,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|) \leq (s^*)^2. \tag{B.17}$$

Now, let consider $\Gamma_i(s^*,\ l)$. From the definition of Γ_i in (B.14), we have that

$$\Gamma_i(s^*, l) \leq (s^*)^2 \quad \Longleftrightarrow \quad l \in [0,\ b_i], \tag{B.18}$$

where b_i is defined in (4.18). Using Assumption 4.4.2 and 4.4.3, and the definition of k_2^*, we have that

$$\eta[k]\ \|\boldsymbol{g}_i[k]\| \leq \eta[k]L \leq \min_{v_j \in \mathcal{R}} \Big\{ \min\{a_j^+,\ b_j\} \Big\} \leq b_i.$$

By the above inequality and statement (B.18), we obtain that

$$\Gamma_i(s^*,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|) \leq (s^*)^2. \tag{B.19}$$

Combine (B.17) and (B.19) to get that

$$\max\Big\{\Gamma_i(\tilde{R}_i,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|),\ \Gamma_i(s^*(\xi),\ \eta[k]\ \|\boldsymbol{g}_i[k]\|)\Big\} \leq (s^*)^2. \tag{B.20}$$

From Lemma 4.4.12, we can write

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 \leq \Gamma_i(\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|).$$

Applying (B.15) and (B.20), respectively to the above inequality yields the result. □

Proof of Lemma 4.4.14. Consider any time-step $k \geq k_3^*$ and agent $v_i \in \mathcal{I}_z[k]$. By the definition of the function Δ_i in (4.16), it is clear that if $p_1 > p_2 \geq \tilde{R}_i$ then $\Delta_i(p_1, l) > \Delta_i(p_2, l)$. Then, we get

$$\Delta_i(\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|, \eta[k]\ \|\boldsymbol{g}_i[k]\|) > \Delta_i(s^*, \eta[k]\ \|\boldsymbol{g}_i[k]\|). \tag{B.21}$$

Furthermore, the function Δ_i satisfies

$$\Delta_i(p,l) \geq \left(\sqrt{p^2-\tilde{R}_i^2}\cos\theta_i - \tilde{R}_i\sin\theta_i\right)l \quad \Longleftrightarrow \quad l \in \left[0,\ \sqrt{p^2-\tilde{R}_i^2}\cos\theta_i - \tilde{R}_i\sin\theta_i\right]. \tag{B.22}$$

We restate inequality (B.9) obtained in the proof of Proposition 4.4.2 here:

$$\frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*)}{\|\boldsymbol{x}-\boldsymbol{x}_i^*\|} \geq \frac{\epsilon}{\delta_i(\epsilon)}. \tag{B.23}$$

Recall the definition of $\mathcal{C}_i(\epsilon)$ in (4.11). For $\boldsymbol{x} \notin \mathcal{C}_i(\epsilon)$, from the definition of convex functions, we have $-\langle g_i(\boldsymbol{x}),\ \boldsymbol{x}_i^* - \boldsymbol{x}\rangle \geq f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*)$. Using the inequality (B.23), we obtain

$$\|\boldsymbol{g}_i(\boldsymbol{x})\| \geq \frac{f_i(\boldsymbol{x}) - f_i(\boldsymbol{x}_i^*)}{\|\boldsymbol{x}-\boldsymbol{x}_i^*\|} \geq \frac{\epsilon}{\delta_i(\epsilon)} = \kappa_i.$$

Using the above inequality, Assumption 4.4.2, and the definition of k_3^*, we have that

$$\eta[k]\kappa_i \leq \eta[k]\ \|\boldsymbol{g}_i[k]\| \leq \eta[k]L \leq \frac{b_i}{2}. \tag{B.24}$$

Since $\eta[k]\ \|\boldsymbol{g}_i[k]\| \in [0,\ \frac{b_i}{2}]$, we can apply (B.22) to get

$$\Delta_i\Big(s^*,\ \eta[k]\ \|\boldsymbol{g}_i[k]\|\Big) \geq \frac{b_i}{2}\eta[k]\ \|\boldsymbol{g}_i[k]\|. \tag{B.25}$$

Combine (B.21), (B.25), and the first inequality of (B.24) to obtain the result. □

B.5 Proof of Proposition 4.4.3 and Lemma 4.4.15

Proof of Proposition 4.4.3. From Proposition 4.4.1, the limit point $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ exists. For all $v_i,\ v_j \in \mathcal{R}$, we have

$$\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + \|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|.$$

Apply Lemma 4.4.8 to obtain that for all $v_i \in \mathcal{R}$, we have

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k]\| \leq \max_{v_j \in \mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + \|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|.$$

Substituting the above inequality into

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}_i[k]\| + \|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|,$$

we obtain the result. □

Proof of Lemma 4.4.15. Suppose $\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq \phi[k]$ for a time-step $k \geq k_0$. From Proposition 4.4.1 and 4.4.3, we have that for all $v_i \in \mathcal{R}$,

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq \phi[k] + 2\beta e^{-\alpha k}, \tag{B.26}$$

where α and β are defined in Proposition 4.4.1.

Recall the definition of $\mathcal{I}_z[k]$ from (4.19). For all $v_i \in \mathcal{I}_z[k]$, from Lemma 4.4.12 and 4.4.14, we have

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 \leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\|^2 - \frac{1}{2} b_i \kappa_i \eta[k].$$

Applying (B.26) to the above inequality, we obtain that for all $v_i \in \mathcal{I}_z[k]$,

$$\begin{aligned}\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 &\leq \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \frac{1}{2} b_i \kappa_i \eta[k] \\ &\leq \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \frac{1}{2} \eta[k] \min_{v_j \in \mathcal{R}} b_j \kappa_j.\end{aligned}$$

On the other hand, for all $v_i \in \mathcal{R} \setminus \mathcal{I}_z[k]$, we have $\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq s^*$ by the definition of $\mathcal{I}_z[k]$. From Lemma 4.4.11 and 4.4.13, we get $\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \leq s^*$ for all $v_i \in \mathcal{R} \setminus \mathcal{I}_z[k]$. Therefore, we conclude that for all $v_i \in \mathcal{R}$,

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\|^2 \leq \max\left\{(s^*)^2, \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \frac{1}{2} \eta[k] \min_{v_j \in \mathcal{R}} b_j \kappa_j\right\}.$$

Using the update rule (4.21), the above inequality implies that $\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \leq \phi[k+1]$.

Next, consider a time-step $k \in \mathbb{N}$. From the gradient update step (4.6), for all $v_i \in \mathcal{R}$, we have

$$\begin{aligned}\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| &= \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty] - \eta[k]\boldsymbol{g}_i[k]\| \\ &\leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| + \eta[k]\ \|\boldsymbol{g}_i[k]\|.\end{aligned}$$

Since $\|\boldsymbol{g}_i[k]\| \leq L$ from Assumption 4.4.2, we can rewrite the above inequality as

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{y}[\infty]\| \leq \|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| + \eta[k]\ L. \tag{B.27}$$

On the other hand, from Proposition 4.4.1 and 4.4.3, for all $v_i \in \mathcal{R}$, we have

$$\|\boldsymbol{z}_i[k] - \boldsymbol{y}[\infty]\| \leq \max_{v_j \in \mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + 2\beta e^{-\alpha k}. \tag{B.28}$$

Combine the inequalities (B.27) and (B.28) together and apply the result recursively to obtain

$$\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k_0] - \boldsymbol{y}[\infty]\| \leq \max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[0] - \boldsymbol{y}[\infty]\| + 2\beta \sum_{k=0}^{k_0-1} e^{-\alpha k} + L \sum_{k=0}^{k_0-1} \eta[k].$$

Since the RHS of the above inequality is $\phi[k_0]$, this completes the first part of the proof.

Consider a time-step $k \in \mathbb{N}$ such that $k \geq k_0$. From the update equation (4.21), using the fact that $\frac{1}{2}\eta[k] \min_{v_i \in \mathcal{R}} b_i \kappa_i > 0$ for all $k \in \mathbb{N}$, we can write

$$\phi[k+1] < \max\left\{s^*,\ \phi[k]\right\} + 2\beta e^{-\alpha k}.$$

Applying the above inequality recursively, we can write that for all $k \geq k_0$,

$$\phi[k] < \max\left\{s^*,\ \phi[k_0]\right\} + 2\beta \sum_{k'=k_0}^{k-1} e^{-\alpha k'}.$$

Substituting equation (4.20) into the above inequality and using the fact that $\sum_{k'=0}^{k-1} e^{-\alpha k'} < \sum_{k'=0}^{\infty} e^{-\alpha k'} = \frac{1}{1-e^{-\alpha}}$ for all $k \in \mathbb{N}$, we obtain the uniform bound as follows:

$$\phi[k] < \max\left\{s^*,\ \max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[0] - \boldsymbol{y}[\infty]\| + L \sum_{k'=0}^{k_0-1} \eta[k']\right\} + \frac{2\beta}{1-e^{-\alpha}}.$$

Setting the RHS of the above inequality to $\bar{\phi}$, we obtain the result. □

B.6 Proof of Lemma 4.4.16, Proposition 4.4.4 and Theorem 4.4.9

Lemma 4.4.16 is used to establish the proof of Proposition 4.4.4.

Proof of Lemma 4.4.16. Suppose that there exists a sequence $\{u[k]\}_{k=0}^{\infty} \subset \mathbb{R}_{\geq 0}$ that satisfies the given update rule. Since $\hat{\eta}[k] \in \mathbb{R}_{\geq 0}$ for all $k \in \mathbb{N}$, we have $u^2[k+1] \leq (u[k] + \gamma_1 \lambda^k)^2$. Since $u[k] \geq 0$ for all $k \in \mathbb{N}$, it follows that

$$0 \leq u[k+1] \leq |u[k] + \gamma_1 \lambda^k| \leq u[k] + \gamma_1 |\lambda|^k.$$

Apply the above inequality recursively to obtain that for all $k \in \mathbb{N}$,

$$u[k] \leq u[0] + \gamma_1 \sum_{\ell=0}^{k} |\lambda|^\ell \leq u[0] + \frac{\gamma_1}{1-|\lambda|} := \bar{u}.$$

From the update rule, we can write

$$\begin{aligned} u^2[k+1] &= u^2[k] + 2\gamma_1 \lambda^k u[k] + \gamma_1^2 \lambda^{2k} - \gamma_2 \hat{\eta}[k] \\ &\leq u^2[k] + 2\gamma_1 \lambda^k \bar{u} + \gamma_1^2 \lambda^{2k} - \gamma_2 \hat{\eta}[k]. \end{aligned}$$

Applying the above inequality recursively, we obtain

$$u^2[k] \leq u^2[0] + 2\gamma_1 \bar{u} \sum_{\ell=0}^{k} \lambda^\ell + \gamma_1^2 \sum_{\ell=0}^{k} \lambda^{2\ell} - \gamma_2 \sum_{\ell=0}^{k} \hat{\eta}[\ell].$$

However, the first three terms on the RHS are bounded in k while the last term is unbounded. This implies that there exists a time-step $\tilde{k} \in \mathbb{N}$ such that $u^2[\tilde{k}] < 0$ which contradicts the fact that $u[\tilde{k}] \in \mathbb{R}_{\geq 0}$. □

Proof of Proposition 4.4.4. Let $\nu = \frac{1}{2}\min_{v_i \in \mathcal{R}} b_i \kappa_i$ to simplify the notations. Let $k_4^* \in \mathbb{N}$ be a time-step such that $\eta[k_4^*] \geq \frac{4\beta}{\nu}\left(\bar{\phi} e^{-\alpha k_4^*} + \beta e^{-2\alpha k_4^*}\right)$. Note that $k_4^* \in \mathbb{N}$ exists since $\eta[k]$ decreases slower than the exponential decay due to its form given in Assumption 4.4.3.

First, we will show that if the time-step $k \in \mathbb{N}$ satisfies $k \geq \max\{k_0, k_4^*\}$ and $\left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k] > (s^*)^2$, then

$$\phi[k+1] < \phi[k]. \tag{B.29}$$

Consider a time-step $k \geq \max\{k_0, k_4^*\}$. Since $\left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k] > (s^*)^2$, the update equation (4.21) reduces to

$$\phi^2[k+1] = \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k]. \tag{B.30}$$

Using the definition of k_4^*, from $k \geq k_4^*$, we can write $\eta[k] \geq \frac{4\beta}{\nu}\left(\bar{\phi} e^{-\alpha k} + \beta e^{-2\alpha k}\right)$. Since $\phi[k] < \bar{\phi}$, we have that

$$\eta[k] > \frac{4\beta}{\nu}\left(\phi[k] e^{-\alpha k} + \beta e^{-2\alpha k}\right).$$

By multiplying ν and adding $\phi^2[k]$ to both sides, and then rearranging, we can write $\phi^2[k] > \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k]$, which is equivalent to $\phi[k+1] < \phi[k]$ by (B.30). This completes our claim.

Next, we will show that there exists a time-step $\tilde{K} \in \mathbb{N}$ such that $\phi\left[\max\{k_0, k_4^*\} + \tilde{K}\right] = s^*$.

Suppose that $\left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k] > (s^*)^2$ for all $k \geq \max\{k_0, k_4^*\}$. Then, the update equation (4.21) reduces to

$$\phi^2[k+1] = \left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k].$$

However, since $\phi[k]$ is non-negative for all $k \in \mathbb{N}$ by its definition, from Lemma 4.4.16, there is no sequence $\{\phi[k]\}_{k=k_0}^{\infty}$ that can satisfy the above update rule. Hence, there exists a constant $\tilde{K} \in \mathbb{N}$ such that

$$\left(\phi[k'] + 2\beta e^{-\alpha k'}\right)^2 - \nu\eta[k'] \leq (s^*)^2,$$

where $k' = \max\{k_0, k_4^*\} + \tilde{K} - 1$, which yields $\phi\Big[\max\{k_0, k_4^*\} + \tilde{K}\Big] = s^*$ by the equation (4.21). This completes the second claim.

Consider any time-step $k \in \mathbb{N}$ such that $k \geq \max\{k_0, k_4^*\} + \tilde{K}$ and $\phi[k] = s^*$. Such a time-step exists due to the argument above. Then, suppose $\left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k] > (s^*)^2$. From (B.29), we have that $\phi[k+1] < s^*$ which is not possible due to the fact that $\phi[k'] \geq s^*$ for all $k' \geq k_0$ from the update equation (4.21). Hence, we conclude that $\left(\phi[k] + 2\beta e^{-\alpha k}\right)^2 - \nu\eta[k] \leq (s^*)^2$, and $\phi[k+1] = s^*$ by (4.21). This means that $\phi[k] = s^*$ for all $k \geq \max\{k_0, k_4^*\} + \tilde{K}$. Then, by the definition of $\phi[k]$, we can rewrite the equation as $\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq s^*$ for all $k \geq \max\{k_0, k_4^*\} + \tilde{K} := K$ which completes the proof. □

Finally, we utilize the finite-time convergence result from Proposition 4.4.4 to give a proof for Theorem 4.4.9 presented below.

Proof of Theorem 4.4.9. From Proposition 4.4.4, for a fixed $\xi \in \mathbb{R}_{>0}$ and $\epsilon \in \mathbb{R}_{>0}$, we have that for all $k \geq K$,

$$\max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq s^*(\xi, \epsilon).$$

Note that K is a function of ξ and ϵ. However, the above inequality implies that $s^*(\xi, \epsilon) \geq \limsup_k \max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\|$. Since the inequality is valid for all $\xi > 0$ and $\epsilon > 0$, and $\inf_{\xi > 0,\, \epsilon > 0} s^*(\xi, \epsilon) = \inf_{\epsilon > 0} s^*(0, \epsilon)$ by the definition of $s^*(\xi, \epsilon)$ in (4.15), we have

$$\limsup_k \max_{v_i \in \mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{y}[\infty]\| \leq \inf_{\epsilon > 0} s^*(0, \epsilon),$$

which completes the proof. □

C. SUPPLEMENTARY MATERIALS FOR CHAPTER 5

C.1 Additional Lemmas

C.1.1 Graph Robustness

We now state a lemma regarding transmissibility of information after dropping some edges from the graph (from Lemma 2.3 in [30]).

Lemma C.1.1. *Suppose $r \in \mathbb{Z}_+$ and $\mathcal{G}$ is r-robust. Let $\mathcal{G}'$ be a graph obtained by removing $r-1$ or fewer incoming edges from each node in $\mathcal{G}$. Then $\mathcal{G}'$ is rooted.*

This means that if we have enough redundancy in the network, information from at least one node can still flow to the other nodes in the network even after each regular node discards some neighboring states.

C.1.2 Series of Products

Lemma C.1.2. *Suppose $\{b[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is a sequence such that $\lim_{k\to\infty} \sum_{s=0}^{k} b[s] = b$.*

- *If $\{a[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is a sequence such that $\limsup_k a[k] = a^*$, then it holds that*

$$\limsup_k \sum_{s=0}^{k} a[s]b[k-s] \leq a^* b. \tag{C.1}$$

- *If $\{a[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is a sequence such that $\liminf_k a[k] = a_*$, then it holds that*

$$\liminf_k \sum_{s=0}^{k} a[s]b[k-s] \geq a_* b. \tag{C.2}$$

Proof. Consider the first part of the lemma. Since $\limsup_k a[k] = a^*$, we have that for a given $\epsilon \in \mathbb{R}_{>0}$, there exists $k'(\epsilon) \in \mathbb{N}$ such that $a[k] \leq a^* + \epsilon$ for all $k \geq k'(\epsilon)$. Suppose $\tilde{k} \in \mathbb{N}$ and $\tilde{k} \geq k'(\epsilon)$, and we can write

$$\begin{aligned}\sum_{s=0}^{\tilde{k}} a[s]b[\tilde{k}-s] &= \sum_{s=0}^{k'(\epsilon)-1} a[s]b[\tilde{k}-s] + \sum_{s=k'(\epsilon)}^{\tilde{k}} a[s]b[\tilde{k}-s] \\ &\leq \sum_{s=0}^{k'(\epsilon)-1} a[s]b[\tilde{k}-s] + (a^*+\epsilon)\sum_{s=0}^{\tilde{k}-k'(\epsilon)} b[s]. \end{aligned} \tag{C.3}$$

Since $\lim_{k\to\infty}\sum_{s=0}^{k} b[s] = b$, we have that $\lim_{k\to\infty} b[k] = 0$. Taking $\limsup_{\tilde{k}}$ to both sides of (C.3), we obtain that

$$\begin{aligned}\limsup_{\tilde{k}} \sum_{s=0}^{\tilde{k}} a[s]b[\tilde{k}-s] &\leq \sum_{s=0}^{k'(\epsilon)-1} a[s] \lim_{\tilde{k}\to\infty} b[\tilde{k}-s] + (a^*+\epsilon)\lim_{\tilde{k}\to\infty}\sum_{s=0}^{\tilde{k}-k'(\epsilon)} b[s] \\ &= (a^*+\epsilon)b.\end{aligned}$$

Since $\epsilon \in \mathbb{R}_{>0}$ can be chosen to be arbitrary small, we obtain (C.1).

For the second part of the lemma, since $\liminf_k a[k] = a_*$, we have that for a given $\epsilon \in \mathbb{R}_{>0}$, there exists $k'(\epsilon) \in \mathbb{N}$ such that $a[k] \geq a_* + \epsilon$ for all $k \geq k'(\epsilon)$. By using this fact and taking the steps same as the proof above, we obtain (C.2). □

Corollary C.1.3. *Suppose* $\{a[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$ *is a sequence such that* $\lim_{k\to\infty} a[k] = 0$ *and* $\{b[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$ *is a sequence such that* $\lim_{k\to\infty}\sum_{s=0}^{k} b[s]$ *is finite. Then, it holds that*

$$\lim_{k\to\infty}\sum_{s=0}^{k} a[s]b[k-s] = 0. \tag{C.4}$$

Proof. Since $\lim_{k\to\infty} a[k] = 0$, we can write

$$\limsup_k a[k] = \liminf_k a[k] = 0.$$

Using Lemma C.1.2, we have that

$$0 \leq \liminf_k \sum_{s=0}^{k} a[s]b[k-s] \leq \limsup_k \sum_{s=0}^{k} a[s]b[k-s] \leq 0.$$

The inequalities above implies that

$$\liminf_k \sum_{s=0}^{k} a[s]b[k-s] = \limsup_k \sum_{s=0}^{k} a[s]b[k-s] = 0,$$

and the result (C.4) follows. □

C.1.3 Function Analysis

Lemma C.1.4. *Given a constant $\gamma \in \mathbb{R}_{\geq 0}$, suppose $h : \left(\max\left\{1-\frac{1}{\gamma}, 0\right\}, 1\right] \to \mathbb{R}_{\geq 0}$ such that*

$$h(s) = \frac{\sqrt{s}}{1-\sqrt{\gamma}\cdot\sqrt{1-s}}.$$

Then, the following statements hold.

- *If $\gamma \in [1, \infty)$, then h is a strictly decreasing function.*
- *If $\gamma \in [0, 1)$, then h is strictly increasing on the interval $(0, 1-\gamma]$ and strictly decreasing on the interval $(1-\gamma, 1]$.*

Proof. Compute the derivative of h with respect to s yields

$$\begin{aligned} h'(s) &= \left(\frac{1}{1-\sqrt{\gamma}\sqrt{1-s}}\right)^2 \left(\frac{1-\sqrt{\gamma}\sqrt{1-s}}{2\sqrt{s}} - \frac{\sqrt{\gamma}\sqrt{s}}{2\sqrt{1-s}}\right) \\ &= \left(\frac{1}{1-\sqrt{\gamma}\sqrt{1-s}}\right)^2 \left(\frac{\sqrt{1-s}-\sqrt{\gamma}}{2\sqrt{s}\sqrt{1-s}}\right). \end{aligned}$$

From the expression $h'(s)$ above, we need to consider only the sign of $\sqrt{1-s}-\sqrt{\gamma}$. First, consider the case that $\gamma \geq 1$. We have $s \in \left(1-\frac{1}{\gamma}, 1\right]$ which implies that $\sqrt{1-s}-\sqrt{\gamma} < 0$. Next, consider the case that $\gamma \in [0, 1)$. We have that $s \in (0, 1-\gamma]$ implies that $\sqrt{1-s}-\sqrt{\gamma} \geq 0$ with equality only if $s = 1-\gamma$, and $s \in (1-\gamma, 1]$ implies that $\sqrt{1-s}-\sqrt{\gamma} < 0$. □

C.2 Proof of Convergence Results in Subsection 5.6.2

C.2.1 Convex Functions

From [98], an equivalent definition of a μ-strongly convex differentiable function f is as follows: for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$,

$$\langle \nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2),\ \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \geq \mu \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{C.5}$$

We will use the following useful result from [98] regarding the convexity and Lipschitz gradient of a function.

Lemma C.2.1. *If f is convex and has L-Lipschitz gradient then for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$,*

$$f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2) + \langle \nabla f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle + \frac{1}{2L} \|\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2)\|^2, \tag{C.6}$$

and

$$f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2) + \langle \nabla f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle + \frac{L}{2} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{C.7}$$

C.2.2 The Reduction Property Implication

We first introduce the following lemma which is useful for deriving the convergence result (Theorem 5.6.4).

Lemma C.2.2. *Suppose Assumption 5.6.1 holds. If an algorithm A in RedGraf satisfies the (γ, α)-reduction property, then $\beta\sqrt{\gamma} < 1$.*

Proof. In the first case where $\gamma \in [0, 1)$ and $\alpha_k = \alpha \in \left(0, \frac{1}{L}\right]$, we have $\beta\sqrt{\gamma} = \sqrt{\gamma} \cdot \sqrt{1 - \alpha\tilde{\mu}} < 1$. In the second case, since $\gamma \in \left[1, \frac{1}{1 - \frac{\tilde{\mu}}{\tilde{L}}}\right)$, we have that $0 \leq \frac{1}{\tilde{\mu}}\left(1 - \frac{1}{\gamma}\right) < \frac{1}{\tilde{L}}$ which indicates that setting the step-size $\alpha_k = \alpha \in \left(\frac{1}{\tilde{\mu}}\left(1 - \frac{1}{\gamma}\right), \frac{1}{\tilde{L}}\right]$ is valid. Since $\alpha > \frac{1}{\tilde{\mu}}\left(1 - \frac{1}{\gamma}\right)$, we also obtain that $\beta\sqrt{\gamma} = \sqrt{\gamma} \cdot \sqrt{1 - \alpha\tilde{\mu}} < 1$. For both cases, we have that $\beta\sqrt{\gamma} < 1$. □

C.2.3 Proof of Theorem 5.6.4

We refactor Theorem 5.6.4 into Proposition C.2.1 and Corollary C.2.3 where Proposition C.2.1 mainly captures the final convergence radius and Corollary C.2.3 captures the convergence rate.

Proposition C.2.1. *Suppose Assumption 5.6.1 holds. If algorithm A in* RedGraf *satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property (for some $\boldsymbol{x}_c \in \mathbb{R}^d$, $\gamma \in \mathbb{R}_{\geq 0}$ and $\{c[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$) and $\alpha_k = \alpha \in \left(0, \frac{1}{\tilde{L}}\right]$ then for all $k \in \mathbb{N}$ and $v_i \in \mathcal{V}_\mathcal{R}$,*

$$\|\boldsymbol{x}_i[k]-\boldsymbol{x}_c\| \leq (\beta\sqrt{\gamma})^k \max_{v_s\in\mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_s[0]-\boldsymbol{x}_c\|+\beta\sum_{s=0}^{k-1}(\beta\sqrt{\gamma})^s c[k-s-1]+r_c\sqrt{\alpha\tilde{L}}\sum_{s=0}^{k-1}(\beta\sqrt{\gamma})^s, \quad \text{(C.8)}$$

where r_c and β are defined in (5.9) and (5.11), respectively. Furthermore, if A satisfies the (γ, α)-reduction property, then for all $v_i \in \mathcal{V}_\mathcal{R}$, it holds that

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| \leq \frac{r_c\sqrt{\alpha\tilde{L}}}{1-\beta\sqrt{\gamma}}. \quad \text{(C.9)}$$

Proof. Consider a regular agent $v_i \in \mathcal{V}_\mathcal{R}$. Since $\boldsymbol{x}_i[k+1] = \tilde{\boldsymbol{x}}_i[k] - \alpha_k \boldsymbol{g}_i[k]$ from (5.4), we can write

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 = \|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2 - 2\langle \tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c,\ \alpha_k \boldsymbol{g}_i[k]\rangle + \alpha_k^2\|\boldsymbol{g}_i[k]\|^2. \quad \text{(C.10)}$$

Since f_i is μ_i-strongly convex (from Assumption 5.6.1), from (5.1) we have that $-\langle \tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c, \boldsymbol{g}_i[k]\rangle \leq (f_i(\boldsymbol{x}_c) - f_i(\tilde{\boldsymbol{x}}_i[k])) - \frac{\mu_i}{2}\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2$. Substituting this inequality into (C.10) to get

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 \leq (1-\alpha_k\mu_i)\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2 + \alpha_k^2\|\boldsymbol{g}_i[k]\|^2 + 2\alpha_k(f_i(\boldsymbol{x}_c) - f_i(\tilde{\boldsymbol{x}}_i[k])). \quad \text{(C.11)}$$

Since f_i has L_i-Lipschitz gradient (from Assumption 5.6.1), from (C.6) we have that $\|\boldsymbol{g}_i[k]\|^2 \leq 2L_i(f_i(\tilde{\boldsymbol{x}}_i[k]) - f_i(\boldsymbol{x}_i^*))$. Substituting this inequality into (C.11) yields

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 \leq (1-\alpha_k\mu_i)\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2 - 2\alpha_k(1-\alpha_k L_i)f_i(\tilde{\boldsymbol{x}}_i[k]) \\ - 2\alpha_k^2 L_i f_i(\boldsymbol{x}_i^*) + 2\alpha_k f_i(\boldsymbol{x}_c). \quad \text{(C.12)}$$

Since f_i has L_i-Lipschitz gradient (from Assumption 5.6.1), from (C.7) we have that $f_i(\boldsymbol{x}_c) \leq f_i(\boldsymbol{x}_i^*) + \frac{L_i}{2}\|\boldsymbol{x}_c - \boldsymbol{x}_i^*\|^2$. Substituting this inequality into (C.12), we obtain

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 \leq (1-\alpha_k\mu_i)\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2 \\ - 2\alpha_k(1-\alpha_k L_i)(f_i(\tilde{\boldsymbol{x}}_i[k]) - f_i(\boldsymbol{x}_i^*)) + \alpha_k L_i\|\boldsymbol{x}_c - \boldsymbol{x}_i^*\|^2.$$

Since $\alpha_k = \alpha \in \left(0, \frac{1}{\tilde{L}}\right]$, $L_i \leq \tilde{L}$ and $\mu_i \geq \tilde{\mu}$, the above inequality implies that

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 \leq (1-\alpha\tilde{\mu})\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_c\|^2 + \alpha\tilde{L}\|\boldsymbol{x}_c - \boldsymbol{x}_i^*\|^2. \quad \text{(C.13)}$$

Since the algorithm A satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property given in (5.8), (C.13) becomes

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\|^2 \leq (1-\alpha\tilde{\mu})\left(\sqrt{\gamma}\max_{v_j\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| + c[k]\right)^2 + \alpha\tilde{L}\max_{v_j\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_c - \boldsymbol{x}_j^*\|^2,$$

which implies that

$$\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\| \leq \sqrt{\gamma}\cdot\sqrt{1-\alpha\tilde{\mu}}\max_{v_j\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| \\ + \sqrt{1-\alpha\tilde{\mu}}\,c[k] + \sqrt{\alpha\tilde{L}}\max_{v_j\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_c - \boldsymbol{x}_j^*\|.$$

Recall the definition of r_c and β from (5.9) and (5.11), respectively. Since the above inequality holds for all $v_i \in \mathcal{V}_\mathcal{R}$, taking maximum over $v_i \in \mathcal{V}_\mathcal{R}$ yields

$$\max_{v_i\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_i[k+1] - \boldsymbol{x}_c\| \leq \beta\sqrt{\gamma}\max_{v_i\in\mathcal{V}_\mathcal{R}}\|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| + \beta c[k] + r_c\sqrt{\alpha\tilde{L}}.$$

Unfolding the recursive inequality above, we obtain that

$$\max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| \leq (\beta\sqrt{\gamma})^k \max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i[0] - \boldsymbol{x}_c\| + \beta \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s c[k-s-1] + r_c\sqrt{\alpha\tilde{L}} \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s,$$

which completes the first part of the proof.

Consider the second part of the theorem. Since the algorithm A satisfies the (γ, α)-reduction property, from Lemma C.2.2, we have that $\beta\sqrt{\gamma} < 1$. Considering the RHS of (C.8), since $\lim_{k\to\infty} \sum_{s=0}^{k} (\beta\sqrt{\gamma})^s$ is finite, using Corollary C.1.3, we have

$$\lim_{k\to\infty} \left[r_c\sqrt{\alpha\tilde{L}} \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s + \beta \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s c[k-s-1] + (\beta\sqrt{\gamma})^k \max_{v_s \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_s[0] - \boldsymbol{x}_c\| \right] = \frac{r_c\sqrt{\alpha\tilde{L}}}{1 - \beta\sqrt{\gamma}}.$$

The result (C.9) follows from taking $\limsup_k$ on both sides of (C.8) and then applying the above equation. □

Corollary C.2.3. *Suppose Assumption 5.6.1 holds. If there exist constants $\boldsymbol{x}_c \in \mathbb{R}^d$, $\gamma \in \mathbb{R}_{\geq 0}$ and $\xi \in (0,1) \setminus \{\beta\sqrt{\gamma}\}$ such that an algorithm A in RedGraf satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction and (γ, α)-reduction properties with $c[k] = \mathcal{O}(\xi^k)$, then for all $v_i \in \mathcal{V}_\mathcal{R}$,*

$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_c\| \leq R^* + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big). \tag{C.14}$$

Proof. Consider the second term on the RHS of (C.8). Since $c[k] = \mathcal{O}(\xi^k)$, we obtain that

$$\beta \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s c[k-s-1] = \mathcal{O}\left(\xi^{k-1} \sum_{s=0} \left(\frac{\beta\sqrt{\gamma}}{\xi}\right)^s\right) = \mathcal{O}\left(\frac{(\beta\sqrt{\gamma})^k - \xi^k}{\beta\sqrt{\gamma} - \xi}\right) = \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big).$$

Using the above equation and the fact that $r_c\sqrt{\alpha\tilde{L}} \cdot \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s \leq R^*$, we have that (C.8) implies (C.14). □

C.2.4 Proof of Theorem 5.6.5

Proof of Theorem 5.6.5. Suppose $\boldsymbol{x}$ is a point in $\mathbb{R}^d$ such that $\|\boldsymbol{x}-\boldsymbol{x}_c\| > \frac{\tilde{L}}{\tilde{\mu}} r_c$. In order to conclude that $\boldsymbol{x}^* \in \mathcal{B}\Big(\boldsymbol{x}_c, \frac{\tilde{L}}{\tilde{\mu}} r_c\Big)$, we will show that $\sum_{v_i \in \mathcal{V}_\mathcal{R}} \nabla f_i(\boldsymbol{x}) \neq \mathbf{0}$.

In the first step, we will show that $\cos \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_c) > 0$ for all $v_i \in \mathcal{V}_\mathcal{R}$. For a regular agent $v_i \in \mathcal{V}_\mathcal{R}$, consider the angle between the vectors $\boldsymbol{x}-\boldsymbol{x}_c$ and $\boldsymbol{x}-\boldsymbol{x}_i^*$. Suppose $r_c > 0$; otherwise, we have that $\angle(\boldsymbol{x}-\boldsymbol{x}_c, \boldsymbol{x}-\boldsymbol{x}_i^*) = 0$. Using Lemma 9 in [93], we can bound the angle as follows:

$$\angle(\boldsymbol{x}-\boldsymbol{x}_c, \boldsymbol{x}-\boldsymbol{x}_i^*) \leq \max_{\boldsymbol{x}_0 \in \mathcal{B}(\boldsymbol{x}_c, r_c)} \angle(\boldsymbol{x}-\boldsymbol{x}_c, \boldsymbol{x}-\boldsymbol{x}_0) = \arcsin\left(\frac{r_c}{\|\boldsymbol{x}-\boldsymbol{x}_c\|}\right) < \arcsin\left(\frac{\tilde{\mu}}{\tilde{L}}\right). \quad \text{(C.15)}$$

On the other hand, for a regular agent $v_i \in \mathcal{V}_\mathcal{R}$, since f_i is μ_i-strongly convex, from (C.5), we have $\langle \nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_i^* \rangle \geq \mu_i \|\boldsymbol{x}-\boldsymbol{x}_i^*\|^2$ which is equivalent to

$$\|\nabla f_i(\boldsymbol{x})\| \cos \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_i^*) \geq \mu_i \|\boldsymbol{x}-\boldsymbol{x}_i^*\|. \quad \text{(C.16)}$$

Since f_i has L_i-Lipschitz gradient, from (5.2), we have $\|\nabla f_i(\boldsymbol{x})\| \leq L_i \|\boldsymbol{x}-\boldsymbol{x}_i^*\|$. Substitute this inequality into (C.16) to obtain $\cos \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_i^*) \geq \frac{\mu_i}{L_i} \geq \frac{\tilde{\mu}}{\tilde{L}}$ which implies that

$$\angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_i^*) \leq \arccos\left(\frac{\tilde{\mu}}{\tilde{L}}\right). \quad \text{(C.17)}$$

Using (C.15) and (C.17), we can bound the angle between the vectors $\nabla f_i(\boldsymbol{x})$ and $\boldsymbol{x}-\boldsymbol{x}_c$ as follows:

$$\begin{aligned} \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_c) &\leq \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_i^*) + \angle(\boldsymbol{x}-\boldsymbol{x}_c, \boldsymbol{x}-\boldsymbol{x}_i^*) \\ &< \arccos\left(\frac{\tilde{\mu}}{\tilde{L}}\right) + \arcsin\left(\frac{\tilde{\mu}}{\tilde{L}}\right) = \frac{\boldsymbol{\pi}}{2}, \end{aligned}$$

where the first inequality is obtained from Corollary 12 in [57]. This means that $0 < \cos \angle(\nabla f_i(\boldsymbol{x}), \boldsymbol{x}-\boldsymbol{x}_c)$ as desired.

In the second step, we will show that $\|\nabla f_i(\boldsymbol{x})\| > 0$ for all $v_i \in \mathcal{V}_{\mathcal{R}}$. For a regular agent $v_i \in \mathcal{V}_{\mathcal{R}}$, consider the lower bound of the gradient's norm $\|\nabla f_i(\boldsymbol{x})\|$ in (C.16) which implies that

$$\|\nabla f_i(\boldsymbol{x})\| \geq \mu_i \|\boldsymbol{x} - \boldsymbol{x}_i^*\| \geq \mu_i \Big(\|\boldsymbol{x} - \boldsymbol{x}_c\| - \|\boldsymbol{x}_i^* - \boldsymbol{x}_c\| \Big).$$

Since $\|\boldsymbol{x} - \boldsymbol{x}_c\| > \frac{\tilde{L}}{\tilde{\mu}} r_c$, the above inequality becomes

$$\|\nabla f_i(\boldsymbol{x})\| > \left(\frac{\tilde{L}}{\tilde{\mu}} \max_{v_j \in \mathcal{V}_{\mathcal{R}}} \|\boldsymbol{x}_j^* - \boldsymbol{x}_c\| - \|\boldsymbol{x}_i^* - \boldsymbol{x}_c\| \right) \geq 0,$$

where the second inequality is obtained by using $\tilde{L} \geq \tilde{\mu}$.

In the last step, we will show that $\sum_{v_i \in \mathcal{V}_{\mathcal{R}}} \nabla f_i(\boldsymbol{x}) \neq \boldsymbol{0}$. Consider the following inner product

$$\left\langle \sum_{v_i \in \mathcal{V}_{\mathcal{R}}} \nabla f_i(\boldsymbol{x}), \boldsymbol{x} - \boldsymbol{x}_c \right\rangle = \|\boldsymbol{x} - \boldsymbol{x}_c\| \sum_{v_i \in \mathcal{V}_{\mathcal{R}}} \|\nabla f_i(\boldsymbol{x})\| \cos \angle (\nabla f_i(\boldsymbol{x}), \boldsymbol{x} - \boldsymbol{x}_c).$$

Since $\|\nabla f_i(\boldsymbol{x})\| > 0$ and $\cos \angle (\nabla f_i(\boldsymbol{x}), \boldsymbol{x} - \boldsymbol{x}_c) > 0$ for all $v_i \in \mathcal{V}_{\mathcal{R}}$, and $\|\boldsymbol{x} - \boldsymbol{x}_c\| > 0$, we have that $\left\langle \sum_{v_i \in \mathcal{V}_{\mathcal{R}}} \nabla f_i(\boldsymbol{x}), \boldsymbol{x} - \boldsymbol{x}_c \right\rangle > 0$. This implies that $\sum_{v_i \in \mathcal{V}_{\mathcal{R}}} \nabla f_i(\boldsymbol{x}) \neq \boldsymbol{0}$ which completes the first part of the proof.

For the second part of the theorem, in order to conclude that $\boldsymbol{x}^* \in \mathcal{B}(\boldsymbol{x}_c, R^*)$, we will show that $\frac{\tilde{L}}{\tilde{\mu}} r_c \leq R^*$ where R^* is defined in (5.12). Since $\gamma \geq 1$, we have that

$$R^* \geq \frac{r_c \sqrt{\alpha \tilde{L}}}{1 - \sqrt{1 - \alpha \tilde{\mu}}}.$$

Multiplying $1 + \sqrt{1 - \alpha \tilde{\mu}}$ to both the numerator and denominator of the RHS of the above inequality, we obtain that

$$R^* \geq r_c \sqrt{\frac{\tilde{L}}{\tilde{\mu}}} \left(\frac{1}{\sqrt{\alpha \tilde{\mu}}} + \sqrt{\frac{1}{\alpha \tilde{\mu}} - 1} \right).$$

Since $\alpha \leq \frac{1}{\tilde{L}}$ implies that $\frac{1}{\alpha\tilde{\mu}} \geq \frac{\tilde{L}}{\tilde{\mu}}$, we can bound R^* as follows:

$$R^* \geq r_c\sqrt{\frac{\tilde{L}}{\tilde{\mu}}}\left(\sqrt{\frac{\tilde{L}}{\tilde{\mu}}}+\sqrt{\frac{\tilde{L}}{\tilde{\mu}}-1}\right) = r_c\left(\frac{\tilde{L}}{\tilde{\mu}}+\sqrt{\frac{\tilde{L}}{\tilde{\mu}}\left(\frac{\tilde{L}}{\tilde{\mu}}-1\right)}\right).$$

Since $\tilde{L} \geq \tilde{\mu}$, we obtain that $R^* \geq \frac{\tilde{L}}{\tilde{\mu}} r_c$ which completes the second part of the proof. □

C.3 Proof of Consensus Results in Subsection 5.6.3

C.3.1 Bound on Gradients

As we have claimed in the main text, the states contraction property (Definition 5.6.3) implies a bound on the gradient $\|\boldsymbol{g}_i[k]\|_\infty$. The following lemma formally illustrates this fact.

Lemma C.3.1. *Suppose Assumption 5.6.1 holds. If an algorithm $\mathcal{A}$ in* RedGraf *satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property (for some $\boldsymbol{x}_c \in \mathbb{R}^d$, $\gamma \in \mathbb{R}_{\geq 0}$, and $\{c[k]\}_{k\in\mathbb{N}} \subset \mathbb{R}$) and $\alpha_k = \alpha \in \left(0, \frac{1}{\tilde{L}}\right]$ then for all $k \in \mathbb{N}$ and $v_i \in \mathcal{V}_\mathcal{R}$,*

$$\|\boldsymbol{g}_i[k]\|_\infty \leq \tilde{L}\sqrt{\gamma}\Bigg[(\beta\sqrt{\gamma})^k \max_{v_s\in\mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_s[0]-\boldsymbol{x}_c\| + \beta\sum_{s=0}^{k-1}(\beta\sqrt{\gamma})^s c[k-s-1] + r_c\sqrt{\alpha\tilde{L}}\sum_{s=0}^{k-1}(\beta\sqrt{\gamma})^s\Bigg] + \tilde{L}c[k] + r_c\tilde{L}, \quad \text{(C.18)}$$

where r_c and β are defined in (5.9) and (5.11), respectively. Furthermore, if $\mathcal{A}$ satisfies the (γ, α)-reduction property, then it holds that for all $v_i \in \mathcal{V}_\mathcal{R}$,

$$\limsup_k \|\boldsymbol{g}_i[k]\|_\infty \leq r_c\tilde{L}\left(1+\frac{\sqrt{\alpha\gamma\tilde{L}}}{1-\beta\sqrt{\gamma}}\right). \quad \text{(C.19)}$$

Proof. Consider a time-step $k \in \mathbb{N}$, and a regular agent $v_i \in \mathcal{V}_\mathcal{R}$. We can write

$$\|\tilde{\boldsymbol{x}}_i[k]-\boldsymbol{x}_i^*\| \leq \|\tilde{\boldsymbol{x}}_i[k]-\boldsymbol{x}_c\| + \|\boldsymbol{x}_i^*-\boldsymbol{x}_c\|.$$

Since the algorithm A satisfies the $(\boldsymbol{x}_c, \gamma, \{c[k]\})$-states contraction property, applying (5.8) into the above inequality yields

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_i^*\| \leq \sqrt{\gamma} \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| + c[k] + r_c, \tag{C.20}$$

where r_c is defined in (5.9). Since $\boldsymbol{g}_i[k] = \nabla f_i(\tilde{\boldsymbol{x}}_i[k])$, using Assumption 5.6.1, we can write

$$\|\boldsymbol{g}_i[k]\| = \|\nabla f_i(\tilde{\boldsymbol{x}}_i[k]) - \nabla f_i(\boldsymbol{x}_i^*)\| \leq \tilde{L}\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{x}_i^*\|.$$

Substituting (C.20) into the above inequality and using the fact that $\|\boldsymbol{g}_i[k]\|_\infty \leq \|\boldsymbol{g}_i[k]\|$, we obtain that

$$\|\boldsymbol{g}_i[k]\|_\infty \leq \tilde{L}\sqrt{\gamma} \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| + \tilde{L}c[k] + r_c\tilde{L}. \tag{C.21}$$

Substituting (C.8) from Theorem 5.6.4 into the above inequality yields (C.18).

To show the second part of the lemma, taking $\limsup_k$ to both sides of (C.21), we have that

$$\limsup_k \|\boldsymbol{g}_i[k]\|_\infty \leq \tilde{L}\sqrt{\gamma}\ \limsup_k \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{x}_c\| + \tilde{L} \lim_{k\to\infty} c[k] + r_c\tilde{L}.$$

Using (5.12) from Theorem 5.6.4 and $\lim_{k\to\infty} c[k] = 0$ yields the result (C.19). □

C.3.2 Proof of Theorem 5.6.6

Here, we present a more general version of Theorem 5.6.6 which will be used to prove Corollary 5.6.7.

Theorem C.3.2 (Consensus)**.** *If an algorithm A in* RedGraf *satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property (for some $\{\boldsymbol{W}^{(\ell)}[k]\}_{k\in\mathbb{N},\, \ell\in[d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ and $G \in \mathbb{R}_{\geq 0}$) and $\alpha_k = \alpha$ for all $k \in \mathbb{N}$, then there exist $\rho \in \mathbb{R}_{\geq 0}$ and $\lambda \in (0,1)$ such that for all $k \in \mathbb{N}$ and $v_i, v_j \in \mathcal{V}_\mathcal{R}$, it holds that*

$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \rho\sqrt{d}\left(\lambda^k \max_{v_r \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_r[0]\|_\infty + \alpha \sum_{s=0}^{k-1} \lambda^{k-s-1} \max_{v_r \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{g}_r[s]\|_\infty\right). \tag{C.22}$$

Furthermore, there exist $\rho \in \mathbb{R}_{\geq 0}$ *and* $\lambda \in (0,1)$ *such that for all* $v_i, v_j \in \mathcal{V}_{\mathcal{R}}$*, it holds that*

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \frac{\alpha \rho G \sqrt{d}}{1-\lambda}. \tag{C.23}$$

Proof. Consider a time-step $k \in \mathbb{Z}_+$ and a dimension $\ell \in [d]$. Since the algorithm A satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property in (5.10), we have that

$$\boldsymbol{x}^{(\ell)}[k] = \boldsymbol{W}^{(\ell)}[k-1]\boldsymbol{x}^{(\ell)}[k-1] - \alpha_{k-1}\boldsymbol{g}^{(\ell)}[k-1]. \tag{C.24}$$

Let $\boldsymbol{\Phi}^{(\ell)}[t,s] = \begin{cases} \boldsymbol{W}^{(\ell)}[t]\boldsymbol{W}^{(\ell)}[t-1]\cdots\boldsymbol{W}^{(\ell)}[s] & \text{if } t \geq s, \\ \boldsymbol{I} & \text{if } t < s, \end{cases}$ for $s, t \in \mathbb{N}$. We can expand (C.24) as follows:

$$\boldsymbol{x}^{(\ell)}[k] = \boldsymbol{\Phi}^{(\ell)}[k-1,0]\boldsymbol{x}^{(\ell)}[0] - \sum_{s=0}^{k-1} \alpha_s \boldsymbol{\Phi}^{(\ell)}[k-1,s+1]\boldsymbol{g}^{(\ell)}[s]. \tag{C.25}$$

Let $\boldsymbol{q}^{(\ell)}(s) \in \mathbb{R}^{|\mathcal{V}_{\mathcal{R}}|}$ be such that $\lim_{t\to\infty} \boldsymbol{\Phi}^{(\ell)}[t,s] = \mathbf{1}\boldsymbol{q}^{(\ell)T}[s]$, and let $\bar{\boldsymbol{x}}^{(\ell)}[k] = \mathbf{1}\boldsymbol{q}^{(\ell)T}[k]\boldsymbol{x}^{(\ell)}[k]$. We can write

$$\|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty = \left\| \left(\boldsymbol{I} - \mathbf{1}\boldsymbol{q}^{(\ell)T}[k]\right)\boldsymbol{x}^{(\ell)}[k] \right\|_\infty.$$

Applying (C.25) to the above equation, we obtain that

$$\begin{aligned} \|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty \leq{}& \left\|\boldsymbol{\Phi}^{(\ell)}[k-1,0] - \mathbf{1}\boldsymbol{q}^{(\ell)T}[0]\right\|_\infty \|\boldsymbol{x}^{(\ell)}[0]\|_\infty \\ &+ \sum_{s=0}^{k-1} \left(\alpha_s \left\|\boldsymbol{\Phi}^{(\ell)}[k-1,s+1] - \mathbf{1}\boldsymbol{q}^{(\ell)T}[s+1]\right\|_\infty \|\boldsymbol{g}^{(\ell)}[s]\|_\infty \right). \end{aligned} \tag{C.26}$$

From Proposition 1 in [89], we have that there exist constants $\rho' \in \mathbb{R}_{\geq 0}$ and $\lambda \in (0,1)$ such that for all $k > s \geq 0$,

$$\left\|\boldsymbol{\Phi}^{(\ell)}[k-1,s] - \mathbf{1}\boldsymbol{q}^{(\ell)T}[s]\right\|_\infty \leq \rho' \lambda^{k-s}.$$

Thus, applying the above inequality, (C.26) can be bounded as

$$\|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty \leq \rho'\lambda^k \|\boldsymbol{x}^{(\ell)}[0]\|_\infty + \sum_{s=0}^{k-1} \alpha_s \rho' \lambda^{k-s-1} \|\boldsymbol{g}^{(\ell)}[s]\|_\infty. \tag{C.27}$$

Since $\alpha_s = \alpha$ for all $s \in \mathbb{N}$, and for $s \in \mathbb{N}$, $\ell \in [d]$ and $\boldsymbol{z}^{(\ell)}[s] = \boldsymbol{x}^{(\ell)}[s]$ or $\boldsymbol{g}^{(\ell)}[s]$,

$$\|\boldsymbol{z}^{(\ell)}[s]\|_\infty \leq \max_{\ell \in [d]} \max_{v_i \in \mathcal{V}_\mathcal{R}} |z_i^{(\ell)}[s]| = \max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{z}_i[s]\|_\infty,$$

the inequality (C.27) becomes

$$\|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty \leq \rho' \lambda^k \max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_i[0]\|_\infty + \alpha \rho' \sum_{s=0}^{k-1} \lambda^{k-s-1} \max_{v_i \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{g}_i[s]\|_\infty. \tag{C.28}$$

Let $\bar{x}^{(\ell)}[k] = \boldsymbol{q}^{(\ell)T}[k] \boldsymbol{x}^{(\ell)}[k]$. For $v_i \in \mathcal{V}_\mathcal{R}$, we can write

$$\|\boldsymbol{x}_i[k] - \bar{\boldsymbol{x}}[k]\| = \sqrt{\sum_{\ell \in [d]} |x_i^{(\ell)}[k] - \bar{x}^{(\ell)}[k]|^2} \leq \sqrt{\sum_{\ell \in [d]} \|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty^2}.$$

Using the above inequality, we have that for all $v_i, v_j \in \mathcal{V}_\mathcal{R}$,

$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \|\boldsymbol{x}_i[k] - \bar{\boldsymbol{x}}[k]\| + \|\boldsymbol{x}_j[k] - \bar{\boldsymbol{x}}[k]\| \leq 2 \sqrt{\sum_{\ell \in [d]} \|\boldsymbol{x}^{(\ell)}[k] - \bar{\boldsymbol{x}}^{(\ell)}[k]\|_\infty^2}.$$

Substituting (C.28) into the above inequality and letting $\rho = 2\rho'$, we obtain the result (C.22). Taking $\limsup_k$ to both sides of (C.22), we have

$$\limsup_k \|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \alpha \rho \sqrt{d} \limsup_k \sum_{s=0}^{k-1} \lambda^{k-s-1} \max_{v_r \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{g}_r[s]\|_\infty.$$

Since for all $v_r \in \mathcal{V}_\mathcal{R}$, we have $\limsup_k \|\boldsymbol{g}_r[k]\|_\infty \leq G$ from Definition 5.6.3 and $\frac{1}{1-\lambda} = \lim_{k \to \infty} \sum_{s=0}^k \lambda^k$, using Lemma C.1.2, the above inequality becomes (C.23). □

C.3.3 Proof of Corollary 5.6.7

Proof of Corollary 5.6.7. From the inequality (C.19) in Lemma C.3.1, we have that the algorithm A satisfies the $(\{\boldsymbol{W}^{(\ell)}[k]\}, G)$-mixing dynamics property with

$$G = r_c \tilde{L} \left(1 + \frac{\sqrt{\alpha \gamma \tilde{L}}}{1 - \beta \sqrt{\gamma}}\right).$$

Substituting G into (C.23) in Theorem 5.6.6 yields the result (5.15).

To show the second part of the theorem, consider the expression in the square bracket of (C.18). Since $c[k] = \mathcal{O}(\xi^k)$ and $\xi \in (0,1) \setminus \{\beta\sqrt{\gamma}\}$, we have that

$$(\beta\sqrt{\gamma})^k \max_{v_s \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_s[0] - \boldsymbol{x}_c\| + \beta \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s c[k-s-1] + r_c \sqrt{\alpha \tilde{L}} \sum_{s=0}^{k-1} (\beta\sqrt{\gamma})^s \\ \leq R^* + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big),$$

where R^* is defined in (5.12). Substituting the above inequality into (C.18), we obtain that for all $v_i \in \mathcal{V}_\mathcal{R}$,

$$\begin{aligned} \|\boldsymbol{g}_i[k]\|_\infty &\leq \tilde{L}\sqrt{\gamma}\Big[R^* + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big)\Big] + \mathcal{O}(\xi^k) + r_c\tilde{L} \\ &= R^*\tilde{L}\sqrt{\gamma} + r_c\tilde{L} + \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^k\Big). \end{aligned}$$

Substituting the above inequality into (C.22) in Theorem 5.6.6, we have that there exist $\rho \in \mathbb{R}_{\geq 0}$ and $\lambda \in (0,1)$ such that for all $v_i, v_j \in \mathcal{V}_\mathcal{R}$,

$$\|\boldsymbol{x}_i[k] - \boldsymbol{x}_j[k]\| \leq \rho\sqrt{d}\Big[\mathcal{O}(\lambda^k) + \frac{\alpha\tilde{L}}{1-\lambda}(r_c + R^*\sqrt{\gamma}) \\ + \alpha \sum_{s=0}^{k-1} \lambda^{k-s-1} \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^s\Big)\Big]. \quad \text{(C.29)}$$

If $\lambda = \max\{\beta\sqrt{\gamma}, \xi\}$, we can replace λ in (C.29) with $\lambda' = \lambda + \epsilon$ where $\epsilon \in (0, 1-\lambda)$. Thus, without loss of generality, there exist $\rho \in \mathbb{R}_{\geq 0}$ and $\lambda \in (0,1) \setminus \{\max\{\beta\sqrt{\gamma}, \xi\}\}$ such that for all $v_i, v_j \in \mathcal{V}_\mathcal{R}$, the inequality (C.29) holds. Consider the last term of (C.29). Since $\lambda \neq \max\{\beta\sqrt{\gamma}, \xi\}$, we have that

$$\sum_{s=0}^{k-1} \lambda^{k-s-1} \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi\})^s\Big) = \mathcal{O}\Big((\max\{\beta\sqrt{\gamma}, \xi, \lambda\})^k\Big).$$

Substituting the above equality into (C.29) yields the result (5.16). □

C.4 Proof of Algorithms Results in Subsection 5.6.4

C.4.1 Proof of Theorem 5.6.8

Before proving Theorem 5.6.8, we consider a property of SDMMFD (Algorithm 2) and SDFD (Algorithm 3). Specifically, for SDMMFD and SDFD, since the dynamics of the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{V}_\mathcal{R}} \subset \mathbb{R}^d$ are independent of the dynamics of the estimated solutions $\{\boldsymbol{x}_i[k]\}_{\mathcal{V}_\mathcal{R}} \subset \mathbb{R}^d$, we restate the convergence results of the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{V}_\mathcal{R}}$ from Proposition 4.4.1.

Lemma C.4.1. *Suppose the set of estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{V}_\mathcal{R}}$ follow SDMMFD or SDFD [93]. Suppose Assumption 5.6.2 hold, the graph $\mathcal{G}$ is $(2F+1)$-robust, and the weights $w_{ij}^{(\ell)}[k]$ satisfy Assumption 5.6.3. Then, there exists $c_1 \in \mathbb{R}_{>0}$, $c_2 \in \mathbb{R}_{\geq 0}$, and $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ with*

$$y^{(\ell)}[\infty] \in \left[\min_{v_i \in \mathcal{V}_\mathcal{R}} y_i^{(\ell)}[k], \max_{v_i \in \mathcal{V}_\mathcal{R}} y_i^{(\ell)}[k]\right] \tag{C.30}$$

for all $k \in \mathbb{N}$ and $\ell \in [d]$ such that for all $v_i \in \mathcal{V}_\mathcal{R}$, we have

$$\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\| < c_1 e^{-c_2 k}. \tag{C.31}$$

Essentially, the lemma above shows that the estimated auxiliary points $\{\boldsymbol{y}_i[k]\}_{\mathcal{V}_\mathcal{R}}$ converge exponentially fast to a single point called $\boldsymbol{y}[\infty] \in \mathbb{R}^d$.

We now provide a proof of Theorem 5.6.8

Proof of Theorem 5.6.8. We first show that each algorithm satisfies the states contraction property with some particular quantities.

Consider a regular agent $v_i \in \mathcal{V}_\mathcal{R}$ following SDMMFD or SDFD. From Lemma 2 in [93], we have that

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{y}_i[k]\| \leq \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\|. \tag{C.32}$$

From Lemma C.4.1, the limit point $\boldsymbol{y}[\infty] \in \mathbb{R}^d$ exists, and we can write

$$\|\boldsymbol{x}_j[k] - \boldsymbol{y}_i[k]\| \leq \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + \|\boldsymbol{y}[\infty] - \boldsymbol{y}_i[k]\|.$$

Substitute the above inequality into (C.32) to get

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{y}_i[k]\| \leq \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + \|\boldsymbol{y}[\infty] - \boldsymbol{y}_i[k]\|. \tag{C.33}$$

Thus, we can write

$$\begin{aligned}\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{y}[\infty]\| &\leq \|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{y}_i[k]\| + \|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\| \\ &\leq \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{y}[\infty]\| + 2\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|,\end{aligned}$$

where the last inequality comes from substituting (C.33). Let $c_1 \in \mathbb{R}_{>0}$ and $c_2 \in \mathbb{R}_{\geq 0}$ be the constants given in Lemma C.4.1. By directly applying (C.31) to $\|\boldsymbol{y}_i[k] - \boldsymbol{y}[\infty]\|$ in the above inequality, we conclude that the algorithms satisfy $(\boldsymbol{y}[\infty], 1, \{2c_1 e^{-c_2 k}\})$-states contraction property.

Consider a regular agent $v_i \in \mathcal{V}_\mathcal{R}$ following CWTM. Since v_i has at least $2F+1$ in-neighbors, from Proposition 5.1 in [30], we have

$$\tilde{x}_i^{(\ell)}[k] = \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} \tilde{w}_{ij}^{(\ell)}[k] x_j^{(\ell)}[k],$$

where $\sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} \tilde{w}_{ij}^{(\ell)}[k] = 1$. Thus, we can write

$$|\tilde{x}_i^{(\ell)}[k] - c^{*(\ell)}| \leq \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} \tilde{w}_{ij}^{(\ell)}[k] \, |x_j^{(\ell)}[k] - c^{*(\ell)}| \leq \max_{v_j \in \mathcal{V}_\mathcal{R}} |x_j^{(\ell)}[k] - c^{*(\ell)}|.$$

Using the above inequality, we obtain that

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{c}^*\|^2 = \sum_{\ell \in [d]} |\tilde{x}_i^{(\ell)}[k] - c^{*(\ell)}|^2 \leq \sum_{\ell \in [d]} \max_{v_j \in \mathcal{V}_\mathcal{R}} |x_j^{(\ell)}[k] - c^{*(\ell)}|^2 \leq \sum_{\ell \in [d]} \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{c}^*\|^2.$$

Thus, we have that $\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{c}^*\| \leq \sqrt{d} \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{c}^*\|$ which corresponds to the $(\boldsymbol{c}^*, d, \{0[k]\})$-states contraction property in Definition 5.6.1.

Finally, consider a regular agent $v_i \in \mathcal{V}_\mathcal{R}$ following RVO. From (5.7), we can write

$$\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{c}^* = \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} w_{ij}[k] \; (\boldsymbol{x}_j[k] - \boldsymbol{c}^*).$$

Since $\sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} w_{ij}[k] = 1$, we have

$$\|\tilde{\boldsymbol{x}}_i[k] - \boldsymbol{c}^*\| \leq \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} w_{ij}[k] \; \|\boldsymbol{x}_j[k] - \boldsymbol{c}^*\| \leq \max_{v_j \in \mathcal{V}_\mathcal{R}} \|\boldsymbol{x}_j[k] - \boldsymbol{c}^*\|,$$

which corresponds to the $(\boldsymbol{c}^*, 1, \{0[k]\})$-states contraction property in Definition 5.6.1.

Next, we show that SDMMFD, CWTM and RVO satisfy the mixing dynamics property with some particular quantities. To this end, we need to show that there exists $\{\boldsymbol{W}^{(\ell)}[k]\}_{k \in \mathbb{N}, \ell \in [d]} \subset \mathbb{S}^{|\mathcal{V}_\mathcal{R}|}$ such that the state dynamics of each algorithm can be written as (5.10), the sequences of graphs $\{\mathbb{G}(\boldsymbol{W}^{(\ell)}[k])\}_{k \in \mathbb{N}}$ are repeatedly jointly rooted for all $\ell \in [d]$, and there exists $G \in \mathbb{R}_{\geq 0}$ such that

$$\limsup_k \|\boldsymbol{g}_i[k]\|_\infty \leq G \quad \text{for all} \quad v_i \in \mathcal{V}_\mathcal{R}.$$

For SDMMFD, since in the `dist_filter` step, each regular agent removes at most F states and in the `full_mm_filter` step, each regular agent removes at most $2dF$ states, from Proposition 5.1 in [30], for all $k \in \mathbb{N}$, $\ell \in [d]$ and $v_i \in \mathcal{V}_\mathcal{R}$, the dynamics can be rewritten as

$$x_i^{(\ell)}[k+1] = \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}} \tilde{w}_{ij}^{(\ell)}[k] \; x_j^{(\ell)}[k] - \alpha_k \; g_i^{(\ell)}[k],$$

where $\tilde{w}_{ii}^{(\ell)}[k] + \sum_{v_j \in \mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}} \tilde{w}_{x,ij}^{(\ell)}[k] = 1$, and $\tilde{w}_{ii}^{(\ell)}[k] > \omega$ and at least $|\mathcal{N}_i^{\text{in}}| - (2d+1)F$ of the other weights are lower bounded by $\frac{\omega}{2}$ (where ω is defined in Assumption 5.6.3). For $k \in \mathbb{N}$, $\ell \in [d]$ and $v_i, v_j \in \mathcal{V}_\mathcal{R}$, let

$$(\widetilde{\boldsymbol{W}}^{(\ell)}[k])_{ij} = \begin{cases} \tilde{w}_{ij}^{(\ell)}[k] & \text{if } \; v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_\mathcal{R}) \cup \{v_i\}, \\ 0 & \text{otherwise,} \end{cases}$$

and we can write the dynamics of all regular agents as

$$\boldsymbol{x}^{(\ell)}[k+1] = \widetilde{\boldsymbol{W}}^{(\ell)}[k]\boldsymbol{x}^{(\ell)}[k] - \alpha_k \boldsymbol{g}^{(\ell)}[k]. \tag{C.34}$$

Following the proof of Theorem 6.1 in [30], we have that each regular agent $v_i \in \mathcal{V}_{\mathcal{R}}$ removes at most $(2d+1)F$ incoming edges (including all incoming edges from Byzantine agents). Since the graph $\mathcal{G}$ is $((2d+1)F+1)$-robust, applying Lemma C.1.1, we can conclude that the subgraph $\mathbb{G}(\widetilde{\boldsymbol{W}}^{(\ell)}[k])$ is rooted for all $k \in \mathbb{N}$ and $\ell \in [d]$ (which implies that the sequence $\left\{\mathbb{G}(\widetilde{\boldsymbol{W}}^{(\ell)}[k])\right\}_{k\in\mathbb{N}}$ is repeatedly jointly rooted for all $\ell \in [d]$). Since SDMMFD satisfies the states contraction property with $\gamma = 1$, substituting $\gamma = 1$ into the inequality (C.19) in Lemma C.3.1, we obtain the mixing dynamics result.

For CWTM, by noting that the graph $\mathcal{G}$ is $(2F+1)$-robust and in the `cw_mm_filter` step, each regular agent removes at most $2F$ states, we can use the same steps as in SDMMFD case to show that there exists $\{\widetilde{\boldsymbol{W}}^{(\ell)}[k]\}_{k\in\mathbb{N},\ell\in[d]}$ such that the dynamics of all regular agents can be written as (C.34) and $\mathbb{G}(\widetilde{\boldsymbol{W}}^{(\ell)}[k])$ is rooted for all $k \in \mathbb{N}$ and $\ell \in [d]$. Since CWTM satisfies the states contraction property with $\gamma = d$, substituting $\gamma = d$ into the inequality (C.19) in Lemma C.3.1, we obtain the mixing dynamics result.

For RVO, from the `safe_point` step, for all $k \in \mathbb{N}$ and $v_i \in \mathcal{V}_{\mathcal{R}}$, we can write

$$\boldsymbol{x}_i[k+1] = \sum_{v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_{\mathcal{R}}) \cup \{v_i\}} w_{ij}[k]\boldsymbol{x}_j[k] - \alpha_k \boldsymbol{g}_i[k],$$

where $w_{ij}[k]$ is defined as in (5.7). For $k \in \mathbb{N}$ and $v_i, v_j \in \mathcal{V}_{\mathcal{R}}$, let

$$(\boldsymbol{W}[k])_{ij} = \begin{cases} w_{ij}[k] & \text{if } \ v_j \in (\mathcal{N}_i^{\text{in}} \cap \mathcal{V}_{\mathcal{R}}) \cup \{v_i\}, \\ 0 & \text{otherwise,} \end{cases}$$

and we can write the dynamics of all regular agents as (5.10) using the above $\boldsymbol{W}[k]$ for all $\ell \in [d]$. Since $p(d,F) > F$ from the definition of p in Section 5.6.4, the subgraph $\mathbb{G}(\boldsymbol{W}[k])$ is rooted for all $k \in \mathbb{N}$ (by Lemma C.1.1). Since RVO satisfies the states contraction property

with $\gamma = 1$, substituting $\gamma = 1$ into the inequality (C.19) in Lemma C.3.1, we obtain the mixing dynamics result. □

C.4.2 Proof of Lemma 5.6.9

Proof of Lemma 5.6.9. First, consider the case where the regular agents follow SDMMFD or SDFD. From (C.30), we have that for all $\ell \in [d]$,

$$y^{(\ell)}[\infty] \in \left[\min_{v_i \in \mathcal{V}_\mathcal{R}} y_i^{(\ell)}[0], \max_{v_i \in \mathcal{V}_\mathcal{R}} y_i^{(\ell)}[0]\right]. \tag{C.35}$$

Since in the initialization step, we set $\boldsymbol{y}_i[0] = \hat{\boldsymbol{x}}_i^*$ for all $v_i \in \mathcal{V}_\mathcal{R}$, we can rewrite (C.35) as

$$y^{(\ell)}[\infty] \in \left[\min_{v_i \in \mathcal{V}_\mathcal{R}} \hat{x}_i^{*(\ell)}, \max_{v_i \in \mathcal{V}_\mathcal{R}} \hat{x}_i^{*(\ell)}\right].$$

Using the above expression, we can write

$$y^{(\ell)}[\infty] - c^{*(\ell)} \leq \max_{v_i \in \mathcal{V}_\mathcal{R}} \hat{x}_i^{*(\ell)} - c^{*(\ell)}.$$

Let $v_{i'} \in \mathcal{V}_\mathcal{R}$ be an agent such that $\hat{x}_{i'}^{*(\ell)} = \max_{v_i \in \mathcal{V}_\mathcal{R}} \hat{x}_i^{*(\ell)}$. The above inequality becomes

$$y^{(\ell)}[\infty] - c^{*(\ell)} \leq \left(\hat{x}_{i'}^{*(\ell)} - x_{i'}^{*(\ell)}\right) + \left(x_{i'}^{*(\ell)} - c^{*(\ell)}\right).$$

Since $\|\hat{\boldsymbol{x}}_{i'}^* - \boldsymbol{x}_{i'}^*\|_\infty \leq \epsilon^*$ and $\|\boldsymbol{x}_{i'}^* - \boldsymbol{c}^*\| \leq r^*$ from the definition of $\boldsymbol{c}^*$ and r^*, the above inequality becomes

$$|y^{(\ell)}[\infty] - c^{*(\ell)}| \leq \left|\hat{x}_{i'}^{*(\ell)} - x_{i'}^{*(\ell)}\right| + \left|x_{i'}^{*(\ell)} - c^{*(\ell)}\right| \leq \epsilon^* + r^*.$$

Applying the above inequality, we have that

$$\|\boldsymbol{y}[\infty] - \boldsymbol{c}^*\|^2 = \sum_{\ell \in [d]} |y^{(\ell)}[\infty] - c^{*(\ell)}|^2 \leq d(r^* + \epsilon^*)^2.$$

Consider a regular agent $v_i \in \mathcal{V}_{\mathcal{R}}$. Using the above inequality and the definition of $\boldsymbol{c}^*$ and r^*, we obtain that

$$\|\boldsymbol{x}_i^* - \boldsymbol{y}[\infty]\| \leq \|\boldsymbol{x}_i^* - \boldsymbol{c}^*\| + \|\boldsymbol{c}^* - \boldsymbol{y}[\infty]\| \leq \sqrt{d}(r^* + \epsilon^*) + r^*.$$

The result follows from noting that $\boldsymbol{x}_c = \boldsymbol{y}[\infty]$ for SDMMFD and SDFD.

Now, consider the case where the regular agents follow CWTM or RVO. Since in this case, $\boldsymbol{x}_c = \boldsymbol{c}^*$, the result directly follows from the definition of $\boldsymbol{c}^*$ and r^*. □

VITA

Kananart Kuwarananchaeroen was born on September 16, 1994, in Bangkok, Thailand. He received his primary education from Attamit School and continued his secondary education at Yothinburana School (English program). In 2016, Kananart graduated with a B.Eng. degree in Electrical Engineering from Chulalongkorn University in Bangkok, Thailand, with a concentration in control and communication systems. He was awarded the University Gold Medal for his outstanding academic achievements, ranking first in his class.

During his undergraduate studies, Kananart gained experience as a research intern at Nara Institute of Science and Technology (NAIST) in Japan. There, he focused on integrated circuit (IC) design, specifically designing digital and analog components for CMOS image sensors.

In August 2017, Kananart began his PhD program at the Elmore Family School of Electrical and Computer Engineering at Purdue University, where he focused his research on distributed optimization and reinforcement learning. During his time at Purdue, he received the third-place award in the Research Poster Competition at the 21st Annual Information Security Symposium hosted by the Center for Education and Research in Information Assurance and Security (CERIAS). He was also selected as a finalist for the Center for Resilient Infrastructures, Systems, and Processes (CRISP) Student Research Competition 2021-22.

From September 2021 to April 2022, Kananart worked as a Research Intern at Strategic CAD Labs in Intel Labs. During this time, he gained valuable experience in developing and utilizing deep reinforcement learning (RL) techniques to improve resource allocation in microservices architecture.

Kananart earned his Ph.D. degree in August 2023, specializing in resilient distributed optimization. Alongside his research focus, he has a keen interest in machine learning, particularly advancements related to artificial general intelligence (AGI). Driven by a strong desire to contribute to intelligent systems technology, he actively strives to propel humanity forward through his research and innovations.